Title

Bias-Reduced Log-Periodogram and Whittle Estimation of the Long-Memory Parameter Without Variance Inflation

Authors

Guggenberger, Patrik
Sun, Yixiao

Publication Date

2004-11-01

Bias-Reduced Log-Periodogram and Whittle Estimation of the Long-Memory Parameter Without Variance Inflation

Patrik Guggenberger
Department of Economics
U.C.L.A.

Yixiao Sun
Department of Economics
U.C.S.D.

August 2003, this revison November 2004

Abstract

In Andrews and Guggenberger (2003) a bias-reduced log-periodogram estimator $\widehat{d}_{LP}(r)$ for the long-memory parameter d in a stationary long-memory time series has been introduced. Compared to the Geweke and Porter-Hudak (1983) estimator $\widehat{d}_{GPH} = \widehat{d}_{LP}(0)$, the estimator $\widehat{d}_{LP}(r)$ for $r \geq 1$ generally reduces the asymptotic bias by an order of magnitude but inflates the asymptotic variance by a multiplicative constant c_r, e.g. $c_1 = 2.25$ and $c_2 = 3.52$. In this paper, we introduce a new, computationally attractive estimator $\widehat{d}_{WLP}(r)$ by taking a weighted average of GPH estimators over different bandwidths. We show that, for each fixed $r \geq 0$, the new estimator can be designed to have the same asymptotic bias properties as $\widehat{d}_{LP}(r)$ but its asymptotic variance is changed by a constant c_r^* that can be chosen to be as small as desired, in particular smaller than c_r. The same idea is also applied to the local-polynomial Whittle estimator $\widehat{d}_{LW}(r)$ in Andrews and Sun (2004) leading to the weighted estimator $\widehat{d}_{WLW}(r)$. We establish the asymptotic bias, variance, and mean-squared error of the weighted estimators, and show their asymptotic normality. Furthermore, we introduce a data-dependent adaptive procedure for selecting r and the bandwidth m and show that up to a logarithmic factor, the resulting adaptive weighted estimator achieves the optimal rate of convergence.

A Monte-Carlo study shows that the adaptive weighted estimator compares very favorably to several other adaptive estimators.

JEL Classification: C13, C14, C22.

Keywords: Adaptive estimation, asymptotic bias, asymptotic normality, bias reduction, frequency domain, long-range dependence, rate of convergence, strongly dependent time series.

We thank a Co-Editor and three referees for very helpful suggestions. We are grateful for the constructive comments offered by Javier Hidalgo, Marc Henry, and especially Katsumi Shimotsu.

1 Introduction

We consider estimation of the long-memory parameter d for a stationary process $\{Y_t : t = 1, ..., n\}$. The spectral density of the time series is given by

$$f(\lambda) = |\lambda|^{-2d} g(\lambda), \tag{1.1}$$

where $d \in (-0.5, 0.5)$ is the long-memory parameter and $g(\cdot)$ is an even function on $[-\pi, \pi]$ satisfying $0 < g(0) < \infty$. The parameter d determines the long-memory properties of the process $\{Y_t\}$ and $g(\lambda)$ determines its short-run dynamics. To maintain generality of the short-run dynamics of $\{Y_t\}$, we do not impose a specific functional form on $g(\lambda)$. Consistent with the literature on narrow-band semiparametric estimation of the long-memory parameter, we only assume certain regularity conditions for $g(\lambda)$ near frequency zero.

Examples in the literature of the narrow-band approach include the widely-used log-periodogram (LP, also known as the GPH estimator) regression estimator, introduced by Geweke and Porter-Hudak (1983) and further analyzed by Robinson (1995b) and Hurvich et al. (1998) and the local Whittle estimator (LW, also known as the Gaussian semiparametric estimator), suggested by Künsch (1987) and further studied by Robinson (1995a). These narrow-band approaches approximate $\ln g(\lambda)$ by a constant in a shrinking neighborhood of the origin. As a consequence, these estimators achieve a rate of mean-squared error (MSE) convergence of only $n^{-4/5}$, no matter how regular $g(\lambda)$ is. In addition, these estimators can be quite biased in finite samples due to contamination from high frequencies (see Agiakloglou et al. (1993)). Given these problems, Andrews and Guggenberger (2003) and Andrews and Sun (2004) (hereafter AG and AS, respectively) propose approximating $\ln g(\lambda)$ locally by an even polynomial of degree $2r$. This leads to the bias-reduced log-periodogram estimator $\widehat{d}_{LP}(r)$ of AG and the local polynomial Whittle estimator $\widehat{d}_{LW}(r)$ of AS. The main feature of these estimators is that for $g(\cdot)$ smooth in a neighborhood of zero, the bias generally converges to zero at a faster rate than that achieved by the GPH and local Whittle estimators. The bias improvement however comes at the price of an increase in the asymptotic variance by a multiplicative constant c_r, e.g. $c_1 = 2.25$ and $c_2 = 3.52$.

In this paper, we introduce modified versions of the GPH and local Whittle estimators that we denote by $\widehat{d}_{WLP}(r, m)$ and $\widehat{d}_{WLW}(r, m)$, respectively. These estimators are given as a weighted average over finitely many GPH (or a k-step version of the local Whittle) estimators $\widehat{d}_{GPH}(m_i)$ calculated using different bandwidths m_i. The bandwidth set is of the form $\{[l_i m]\}_{i=1}^{K}$, for $K \in \mathbb{N}$, $m \in \mathbb{N}$, and $l_i \in \mathbb{Q}^+$. The idea of the weighted estimator is to choose weights w_i in such a way that the r dominant bias terms of the underlying GPH (or k-step local Whittle) estimator are eliminated. Building on results in Hurvich et al. (1998), Robinson (1995a, 1995b), AG, and AS, we derive the asymptotic bias, variance, and MSE of the weighted estimators, and establish their asymptotic normality. Our asymptotic results imply that, if the weights w_i are chosen appropriately, the weighted estimator shares or even improves on the asymptotic bias advantage of the estimators of AG and AS. The main advantage of the weighted estimator over the estimators of AG and AS is that the multiplicative constant c_r by which the asymptotic variance is inflated can be chosen to be arbitrarily small. In other words, the new estimator can be designed to have the asymptotic bias properties of the estimators of AG and AS but its asymptotic variance properties are better. Under local smoothness assumptions on g, the weighted estimator can be designed such that the order of the asymptotic bias gets arbitrarily close to the parametric one, while at the same time its asymptotic variance can be made arbitrarily small.

By choosing the weights appropriately, we can obtain an estimator that has the same asymptotic bias as the Gaussian semiparametric estimator but a smaller asymptotic variance. This result is unexpected as Robinson (1995a, p. 1640) conjectures that $1/4$ is the semiparametric efficiency bound for the estimation of the long-memory parameter for a given bandwidth, say m, and that the Gaussian semiparametric estimator achieves this bound. See also Hidalgo (2005) for further discussion of this issue. Our results do not contradict Robinson's conjecture, as the weighted estimator uses information beyond bandwidth m. However, our results do imply that it is possible to design a more efficient estimator than $\widehat{d}_{LW}(0)$ while maintaining or even reducing its asymptotic bias. In selecting and using only one bandwidth m, existing narrow-band estimators $\widehat{d}(m)$ discard potentially useful information that is stored in estimators using other bandwidths, e.g. $\widehat{d}(.5m)$ and $\widehat{d}(2m)$. The weighted estimator uses such additional information by averaging over the estimators $\widehat{d}([l_i m])$ over a set of l_is.

Apart from the theoretical properties, the weighted estimators are also computationally attractive. Both $\widehat{d}_{WLP}(r,m)$ and $\widehat{d}_{WLW}(r,m)$ (the latter estimator being based on a k-step version of the local Whittle estimator) require only K OLS regressions and one GLS regression. This is an advantage over the procedure in AS that requires a nonlinear optimization routine to calculate the estimator.

The weighted estimator depends on the choice of the parameters r, m, and l. From results in AG it follows that for a fixed l, the optimal choice of r and m in terms of the rate of convergence of the estimator depends on the unknown smoothness of g at zero. We provide an adaptive estimator of d (denoted "AWLW" for "adaptive weighted local Whittle"), based on Lepskii (1990), that uses the data to select r and m. We show that this estimator obtains the optimal rate of convergence up to a logarithmic factor. We suggest rule of thumb choices for the sequence l based on the relative performance of the weighted estimator in a Monte Carlo study.

In Monte Carlo simulations, using various models including one with nonsmooth spectrum, we compare the root mean-square error (RMSE) performance of the AWLW estimator with several other adaptive estimators, namely, with the ones in Giraitis et al. (2000), Hurvich (2001), Iouditsky et al. (2001), and AS. The results are very encouraging and show that if the contamination from the short-run component of the spectrum is not too large, the AWLW estimator typically outperforms the other adaptive estimators in RMSE performance.

Besides AG and AS, another bias reduction approach in the narrow-band literature is Robinson and Henry (2003), who consider a general class of semiparametric

M-estimators of d that utilize higher-order kernels to obtain bias-reduction. As they state, their results are heuristic in nature, whereas the results in our paper are established rigorously under specific regularity conditions. Also, their estimators suffer from a variance inflation that can be very large.

An alternative to the narrow-band approach to bias reduction is the so called broad-band approach where one imposes regularity conditions on g over the whole interval $[0, \pi]$ and one utilizes a nonparametric estimator of $g(\lambda)$ for $\lambda \in [0, \pi]$. Examples include Moulines and Soulier (1999, 2000), Hurvich (2001), Hurvich and Brodsky (2001), Iouditsky et al. (2001), and Hurvich et al. (2002). In these papers, $\ln g(\lambda)$ is approximated by a truncated Fourier series. It is established that under the given regularity conditions the estimators exhibit an asymptotic MSE of order $(\ln n)/n$ provided the number of parameters in the model goes to infinity at a suitable rate.

From a broad perspective, the weighted estimator bears some similarity to the generalized jackknife estimator in the kernel smoothing literature. In that literature, it has been suggested to average kernel estimators over different smoothing parameters to achieve bias reduction, see e.g. Gray and Schucany (1972) and Härdle (1990, Section 4.6).

We now quickly mention possible extensions of the paper. While we focus on scalar time series satisfying (1.1), it should be possible to extend the idea to the multivariate case along the lines of Robinson (1995b) and Lobato (1999). Although our results are obtained for stationary processes, they can also be utilized when the underlying process is nonstationary (see the discussion in AS, p. 571). In this case, a preliminary consistent estimator, such as the estimator of Kim and Phillips (1999a, 1999b), Velasco (1999), Velasco and Robinson (2000), Shimotsu and Phillips (2002), or Phillips and Shimotsu (2004) is required. While the asymptotic results for the weighted Whittle estimator are derived without assuming Gaussianity, we require such an assumption for the weighted log-periodogram estimator. Using the insight of Velasco (2000), the assumption of Gaussianity could possibly be relaxed for the latter estimator as well. Finally, rather than averaging over different $\widehat{d}_{GPH}(0)$ or (a k-step version of) $\widehat{d}_{LW}(0)$ estimators as in this paper, we could apply the weighted approach to $\widehat{d}_{LP}(r)$ or (a k-step version of) $\widehat{d}_{LW}(r)$ for $r \geq 1$ to eliminate additional bias terms. These extensions however are beyond the scope of this paper.

The remainder of this paper is organized as follows. In the next section, we review the estimators of AG and AS and lay out the assumptions. Section 3 introduces the new weighted estimators and establishes their asymptotic properties. Section 4 discusses adaptive estimation. Section 5 describes the simulation results and an appendix provides proofs of the theorems.

By $C^s(U)$ we denote the space of functions that are s-times continuously differentiable. By $[s]$ we denote the integer part of $s \in \mathbb{R}$. For a matrix A we denote by $[A]_{i,\cdot}$ and $[A]_{i,j}$ the i-th row and the (i, j)-th element of the matrix A, respectively. $1(\cdot)$ is the indicator function.

2 The Log-periodogram and Whittle Estimator

In this section we review the bias-reduced log-periodogram regression estimator $\widehat{d}_{LP}(r,m)$ of AG and the local polynomial Whittle estimator $\widehat{d}_{LW}(r,m)$ of AS. These estimators are used as benchmarks for later comparisons. We also lay out the assumptions.

For $i := \sqrt{-1}$, $m \in \mathbb{N}$, and $j = 1, ..., m$ define

$$\begin{aligned} \lambda_j &: = 2\pi j/n,\ w_j := \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Y_t \exp(it\lambda_j),\ I_j := |w_j|^2, \\ f_j &: = f(\lambda_j),\ \text{and}\ g_j := g(\lambda_j). \end{aligned} \tag{2.1}$$

The bias-reduced log-periodogram estimator $\widehat{d}_{LP}(r,m)$ is the least squares estimator of d from the pseudo-regression of $\ln I_j$ on a constant, $-2\ln\lambda_j, \lambda_j^2, ..., \lambda_j^{2r-2}$, and λ_j^{2r} for $j = 1, ..., m$, where m is the number of fundamental frequencies employed in the regression. The estimator is based on the identity derived from (1.1)

$$\ln I_j = (\ln g_0 - C) - 2d \ln \lambda_j + \sum_{k=1}^{r} \frac{b_{2k}}{(2k)!} \lambda_j^{2k} + e_j, \tag{2.2}$$

where, if defined, $\sum_{k=1}^{r} (b_{2k}/(2k)!)\lambda_j^{2k}$ are the first terms of the Taylor series expansion of $\ln(g_j/g_0)$, $C := 0.57721...$ is the Euler constant, and

$$e_j := \ln(g_j/g_0) - \sum_{k=1}^{r} \frac{b_{2k}}{(2k)!} \lambda_j^{2k} + \ln I_j/f_j + C$$

are regarded as regression errors. The GPH estimator is obtained by setting $r = 0$ and is based on approximating $\ln(g_j/g_0)$ by a constant around the origin[1]. As can be shown, the dominant bias term of the GPH estimator is caused by the term $\ln(g_j/g_0)$ rather than $E \ln I_j/f_j + C$. For $r = 1$, λ_j^2 is added as an additional regressor to the pseudo-regression model which leads to the elimination of the dominant bias term of the GPH estimator. For $r \geq 2$, additional bias terms are eliminated.

For notational convenience we typically suppress the dependence of the estimator on r and/or m and write $\widehat{d}_{LP}$, $\widehat{d}_{LP}(r)$, or $\widehat{d}_{LP}(m)$ for $\widehat{d}_{LP}(r,m)$.

The local-polynomial Whittle estimator $\widehat{d}_{LW} = \widehat{d}_{LW}(r,m)$ of AS is an M-estimator that minimizes the (negative) local-polynomial Whittle log-likelihood. The latter is given by

$$Q_r(d,G,b) := m^{-1} \sum_{j=1}^{m} \left\{ \ln\left[G\lambda_j^{-2d} \exp(-p_r(\lambda_j, b)) \right] + \frac{I_j}{G\lambda_j^{-2d} \exp(-p_r(\lambda_j, b))} \right\}, \tag{2.3}$$

where the notation $G := g_0$ is taken from Robinson (1995a) and AS and

$$p_r(\lambda_j, b) := \sum_{k=1}^{r} -\frac{b_{2k}}{(2k)!} \lambda_j^{2k} \text{ and } b := (b_2, ..., b_{2r})'.$$

Concentrating $Q_r(d,G,b)$ with respect to $G \in (-\infty,\infty)$ yields the concentrated LPW log-likelihood $\mathcal{R}_r(d,b)$:

$$\mathcal{R}_r(d,b) = \ln \widehat{G}_r(d,b) - m^{-1}\sum_{j=1}^{m} p_r(\lambda_j,b) - 2dm^{-1}\sum_{j=1}^{m}\ln\lambda_j + 1, \text{ where}$$

$$\widehat{G}_r(d,b) \ : \ = m^{-1}\sum_{j=1}^{m} I_j \exp(p_r(\lambda_j,b))\lambda_j^{2d}. \tag{2.4}$$

The local polynomial Whittle estimator $(\widehat{d}_{LW},\widehat{b}_{LW})$ of (d,b) solves the following minimization problem:

$$(\widehat{d}_{LW}(r,m),\widehat{b}_{LW}(r,m)) := \underset{d\in[d_1,d_2],b\in\Theta}{\arg\min}\ \mathcal{R}_r(d,b), \tag{2.5}$$

where Θ is a compact and convex set in $\mathbb{R}^r$. Existence and uniqueness of $(\widehat{d}_{LW},\widehat{b}_{LW})$ hinge on strict convexity of $\mathcal{R}_r(d,b)$ (see AS, p. 576) and convexity and compactness of the parameter space. The local Whittle estimator (Künsch (1987) and Robinson (1995a)) is obtained by setting $r=0$.

Before we state the assumptions we need the following definition. We say that a real function h defined on a neighborhood of zero is "smooth of order $0<s<\infty$ at zero" if $h \in C^{[s]}(U)$ for some neighborhood U of zero and its derivative of order $[s]$, denoted $h^{([s])}$, satisfies a Hölder condition of order $s-[s]$ at zero, i.e., $|h^{([s])}(\lambda)-h^{([s])}(0)| \leq C|\lambda|^{s-[s]}$ for some constant $C<\infty$ and all $\lambda \in U$. We say that h is "smooth of order ∞ at zero", if h is smooth of order s for every finite s.

AG make the following assumptions to derive the asymptotic properties of the log-periodogram estimator $\widehat{d}_{LP}(r,m)$.

Assumption AG1. $f(\lambda) = |\lambda|^{-2d}g(\lambda)$, where $d \in (-1/2, 1/2)$.

Assumption AG2. **(i)** g is smooth of order $s \geq 2+2r$ at $\lambda = 0$ for some nonnegative integer r. **(ii)** g is an even function on $[-\pi,\pi]$, $0 < g(0) < \infty$, and $\int_{-\pi}^{\pi}|\lambda|^{-2d}g(\lambda)d\lambda < \infty$.

Assumption AG3. The time series $\{Y_t : t = 1,...,n\}$ is Gaussian.

Assumption AG4. $m = m(n) \to \infty$ and $m/n \to 0$ as $n \to \infty$.

For the local polynomial Whittle estimator AS make the following assumptions.[2]

Assumption AS1. Assumption AG1 holds for $d \in [d_1,d_2]$ with $-1/2 < d_1 < d_2 < 1/2$.

Assumption AS2. Assumption AG2 holds.

Assumption AS3. (a) The time series $\{Y_t : t = 1,...,n\}$ satisfies

$$Y_t - EY_0 = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j},$$

where

$$\sum_{j=0}^{\infty}\alpha_j^2 < \infty,\ E(\varepsilon_t|F_{t-1}) = 0 \text{ a.s.},\ E(\varepsilon_t^2|F_{t-1}) = 1 \text{ a.s.},$$
$$E(\varepsilon_t^3|F_{t-1}) = \sigma_3 \text{ a.s.},\ E(\varepsilon_t^4|F_{t-1}) = \sigma_4 \text{ a.s. for } t = ..., -1, 0, 1, ...,$$

and F_{t-1} is the σ-field generated by $\{\varepsilon_s : s < t\}$.
(b) In some neighborhood of the origin, $(d/d\lambda)\alpha(\lambda) = O(|\alpha(\lambda)|/\lambda)$ as $\lambda \to 0+$, where $\alpha(\lambda) = \sum_{j=1}^{\infty}\alpha_j e^{-ij\lambda}$.

Assumption AS4. AG4 holds and $m^{2.5+2r}/n^{2+2r} \to \chi \in (0, \infty)$ as $n \to \infty$.

Under Assumptions AG1–AG4, Theorem 1 in AG gives the following formulae for the asymptotic bias and variance of the $\widehat{d}_{LP}(r, m)$ estimator[3]. For certain real numbers τ_r, c_r, and for s given in AG2(i), we have

$$\begin{aligned}(a)\ E\widehat{d}_{LP}(r,m) - d &= \tau_r b_{2+2r}\frac{m^{2+2r}}{n^{2+2r}}(1 + o(1)) + O(\frac{m^q}{n^q}) + O(\frac{\ln^3 m}{m}), \quad (2.6)\\ (b)\ Var(\widehat{d}_{LP}(r,m)) &= \frac{\pi^2}{24m}c_r + o(\frac{1}{m}),\ \text{where}\\ q &: = \min\{s, 4+2r\}. \quad (2.7)\end{aligned}$$

If $s = 2 + 2r$, then the $O(m^q/n^q)$ term can be replaced by $o(m^q/n^q)$. The numbers τ_r and c_r are defined in AG, for example, $\tau_0 = -2.19$, $\tau_1 = 2.23$, $\tau_2 = -.793$ and $c_0 = 1$, $c_1 = 2.25$, and $c_2 = 3.52$.

If in addition AS4 is assumed, which asymptotically is the MSE-optimal choice of m, AG (p. 687, comment 2) show that[4]

$$\sqrt{m}(\widehat{d}_{LP}(r,m) - d - \tau_r b_{2+2r}\frac{m^{2+2r}}{n^{2+2r}}) \to_d N(0, \frac{\pi^2}{24}c_r) \quad (2.8)$$

or equivalently that

$$\sqrt{m}(\widehat{d}_{LP}(r,m) - d) \to_d N(\chi\tau_r b_{2+2r}, \frac{\pi^2}{24}c_r). \quad (2.9)$$

Similarly, under AS1–AS4, AS show that for the $\widehat{d}_{LW}(r, m)$ estimator

$$\sqrt{m}(\widehat{d}_{LW}(r,m) - d - \tau_r b_{2+2r}\frac{m^{2+2r}}{n^{2+2r}}) \to_d N(0, \frac{1}{4}c_r). \quad (2.10)$$

Comparing the bias of $\widehat{d}_{LP}(r,m)$ ($\widehat{d}_{LW}(r,m)$) with the one of $\widehat{d}_{LP}(0,m)$ ($\widehat{d}_{LW}(0,m)$), one sees that the rate of convergence of the bias to zero is generally faster for the former, whereas its variance is increased by the multiplicative constant c_r. When $\ln g(\lambda)$ is approximated reasonably well by $\ln g_0$ in a neighborhood of zero, the bias reduction for $\widehat{d}_{LP}(r,m)$ ($\widehat{d}_{LW}(r,m)$) over $\widehat{d}_{LP}(0,m)$ ($\widehat{d}_{LW}(0,m)$) may be very small and the increase in variance may dominate the reduction of bias in finite samples. It would be desirable to have an estimator that shares the asymptotic bias properties of the estimators of AG and AS without inflating the asymptotic variance or at least inflating it by a constant smaller than the constant c_r. In the next section, we describe a new estimator for d that has these properties.

3 Weighted Semiparametric Estimators

In this section, we introduce the new weighted estimator $\widehat{d}_{We}$ and investigate its asymptotic properties. We first describe its general form as a weighted average of an estimator $\widehat{d}_e(0,m)$ at different bandwidths m and then explicitly look at the cases $e =$ LP and $e =$ LW. More precisely, rather than the local Whittle estimator $\widehat{d}_{LW}(0,m)$, we work with a k-step local Whittle estimator $\widehat{d}_{LW,k}(m)$ defined in (3.12). The k-step estimator is computationally more attractive than the local Whittle estimator (because it is given in closed form) but shares its relevant asymptotic properties.

To motivate the weighted estimator, consider the $\widehat{d}_{LP}(0)$ estimator at two different bandwidths m_1 and m_2 that are given as multiples of m for fixed numbers l_i, $m_i := [l_i m]$, $i = 1, 2$. Recall that under certain regularity conditions and rate conditions on m, the asymptotic bias of the LP estimator can be written as

$$\text{ABIAS}(\widehat{d}_{LP}(0,m)) = \tau_0 b_2 \frac{m^2}{n^2}(1 + o(1)). \tag{3.1}$$

Define a weighted estimator $\widehat{d}_{WLP}$ by

$$\widehat{d}_{WLP} := \frac{m_1^2}{m_1^2 - m_2^2}\widehat{d}_{LP}(0,m_2) - \frac{m_2^2}{m_1^2 - m_2^2}\widehat{d}_{LP}(0,m_1).$$

It follows from (3.1) that the asymptotic bias of $\widehat{d}_{WLP}$ is of order $o(m^2/n^2)$ because by choosing weights $w_2 := m_1^2/(m_1^2 - m_2^2)$ and $w_1 := -m_2^2/(m_1^2 - m_2^2)$ we have eliminated the dominant bias terms of $\widehat{d}_{LP}(0,m_1)$ and $\widehat{d}_{LP}(0,m_2)$. By choosing l_1 and l_2 appropriately, we can in addition control the asymptotic variance of $\widehat{d}_{WLP}$. In the following we describe how to design weights and pick constants l_i such that the weighted estimator has the same asymptotic bias as $\widehat{d}_{LP}(r,m)$ but smaller asymptotic variance.

The example illustrates the idea of the weighted estimator. We now consider the general case in which the average is taken over more than two different bandwidths and higher order biases may be eliminated. To describe it, let r (as in $\widehat{d}_{LP}(r)$ and $\widehat{d}_{LW}(r)$) be a nonnegative integer that denotes the number of dominant bias terms to be eliminated, $l = (l_1, l_2, ..., l_K)'$ be a K-vector that pins down the different bandwidths[5] $m_i := [l_i m]$ for $m \in \mathbb{N}$, and finally $w = (w_1, w_2, ..., w_K)'$ be a K-vector of weights that satisfies

$$\sum_{i=1}^{K} w_i = 1, \tag{3.2}$$

$$\sum_{i=1}^{K} w_i l_i^{2k} = 0, \text{ for } k = 1, ..., r. \tag{3.3}$$

Occasionally, we impose an additional condition on the weights, namely

$$\sum_{i=1}^{K} w_i l_i^{2+2r} = \delta \sum_{i=1}^{K} l_i^{2+2r} \text{ for a } \delta \in \mathbb{R}\backslash\{0\}. \tag{3.4}$$

The weighted estimator is then defined by[6]

$$\widehat{d}_{We} := \widehat{d}_{We}(r,m,l,\delta) := \sum_{i=1}^{K} w_i \widehat{d}_e(0,m_i), \tag{3.5}$$

where in the next subsections we will explicitly focus on the cases $e = \text{LP}$ and $e = \text{LW}$. If $\widehat{d}_{We}$ is only subject to conditions (3.2)-(3.3) then of course $\widehat{d}_{We} := \widehat{d}_{We}(r,m,l)$.

We now discuss conditions (3.2)-(3.4). Condition (3.2) guarantees that $\widehat{d}_{We}$ is asymptotically unbiased. Condition (3.3) is motivated from Lemma 3.1(a) below which states that the asymptotic bias of the LP estimator up to lower order terms is given by $\sum_{k=1}^{r+1} \tau_{k-1}^{*} b_{2k} m^{2k}/n^{2k}$ for some constants τ_k^{*}. Therefore, the condition implies that the highest order bias terms of $\widehat{d}_e$ up to m^{2r}/n^{2r} are eliminated for $\widehat{d}_{We}$. Under (3.2)-(3.3), we occasionally want to be able to control the multiplicative constant for the resulting dominant bias term m^{2+2r}/n^{2+2r} of the weighted estimator. For example, in this section we match the dominant bias term of $\widehat{d}_{WLP}(m)$ with that of $\widehat{d}_{LP}(r,m)$ and then compare their asymptotic variances to deduce the superiority of the weighted procedure in terms of MSE. Condition (3.4) pins down the multiplicative constant of the dominant bias term through appropriate choice of δ. Condition (3.4) is imposed mainly to make the bias/variance comparison of the weighted procedure with other estimation methods. In Section 4 we will implement the adaptive weighted estimator based solely on conditions (3.2)-(3.3).

Note that conditions (3.2)-(3.4) are invariant to renormalization of the vector l. More precisely if r, m, l, δ satisfy conditions (3.2)-(3.4) then so does $r, Lm, l/L, \delta$ for any positive constant L and we have $\widehat{d}_{We}(r,m,l,\delta) = \widehat{d}_{We}(r,Lm,l/L,\delta)$. Therefore, w.l.o.g. we can normalize $l_1 = 1$ and we will do so for the remainder of the paper unless otherwise stated.

Below we give some theory and Monte Carlo-based discussion and rules of thumb of how to choose the parameters r, m, l, and δ. For now, we take these parameters as given and describe an easy procedure to implement $\widehat{d}_{We}$, i.e. a procedure to design weights w satisfying conditions (3.2)-(3.4). The implementation without condition (3.4) is analogous. Consider the following pseudo-regression of $\widehat{d}_e(0,m_i)$ on a constant, $l_i^2, ..., l_i^{2r}$, and $l_i^{2+2r} - \delta \sum_{i=1}^{K} l_i^{2+2r}$

$$\widehat{d}_e(0,m_i) = d + \sum_{j=1}^{r} \beta_{2j} l_i^{2j} + \beta_{2+2r}(l_i^{2+2r} - \delta \sum_{i=1}^{K} l_i^{2+2r}) + u_i, \;\; i = 1, ..., K, \tag{3.6}$$

where u_i is the error term. Note that the regression coefficients β stand for different quantities in the cases $e = \text{LP}$ and $e = \text{LW}$. For notational convenience however, we use β in both cases. Let $Z_i := (1, l_i^2, l_i^4, ..., l_i^{2r}, (l_i^{2+2r} - \delta \sum_{i=1}^{K} l_i^{2+2r}))$ and $Z := (Z_1', ..., Z_K')' \in \mathbb{R}^{K \times (2+r)}$. The weighted estimator $\widehat{d}_{We}$ of d is defined as the first component of the GLS estimator of $(d, \beta) = (d, \beta_2, \beta_4, ..., \beta_{2+2r})$, i.e.

$$(\widehat{d}_{We}, \widehat{\beta})' := \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}\widehat{\mathbf{d}}_e, \tag{3.7}$$

where $\Omega := (\Omega_{ij}) \in \mathbb{R}^{K \times K}$, $\Omega_{ij} := 1/\max(l_i, l_j)$, and $\widehat{\mathbf{d}}_e$ is the K-vector with i-th element $\widehat{d}_e(0,m_i)$. In Lemma 3.1(b) below we establish that up to a multiplicative constant Ω is the asymptotic variance matrix of $u = (u_1, ..., u_K)$. It is easy to see that $\widehat{d}_{We}$ is a weighted average of the $\widehat{d}_e(0, m_i)$ with weights given by

$$w' := \left[\left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}\right]_{1,\cdot}. \tag{3.8}$$

These weights satisfy conditions (3.2)-(3.4) and give rise to our weighted estimator $\widehat{d}_{We}$.

In the next two subsections, we analyze the properties of the weighted estimator $\widehat{d}_{We}$ when the components of $\widehat{\mathbf{d}}_e$ are the log-periodogram regression estimator $\widehat{d}_{LP}(0, m)$ or a k-step version of the local Whittle estimator $\widehat{d}_{LW}(0, m)$. Both these estimators are easy to compute. For example, for the estimator $\widehat{d}_{WLP}$ we only need to first run K OLS regressions to get $\widehat{d}_{LP}(0, m_i)$ and then one GLS regression (3.7) to obtain $\widehat{d}_{WLP}$.

3.1 The Weighted Log-Periodogram (WLP) Estimator

We now consider the weighted estimator $\widehat{d}_{We}$ in (3.7) when $e =$ LP. The asymptotic properties of $\widehat{d}_{WLP}$ depend crucially on those of $\widehat{d}_{LP}$ and the covariance structure of $\{\widehat{d}_{LP}(0, m_i)\}$. We first state the following refinement of the results of Theorem 1 in Hurvich et al. (1998).

Lemma 3.1 *Suppose Assumptions AG1-AG4 hold. Then,*
(*a*) *The bias of $\widehat{d}_{LP}(0, m)$ satisfies*

$$E\widehat{d}_{LP}(0, m) - d = \textstyle\sum_{k=1}^{r+1} \tau_{k-1}^* b_{2k} \dfrac{m^{2k}}{n^{2k}} + O(\dfrac{m^q}{n^q}) + O(\dfrac{\ln^3 m}{m}),$$

where

$$\tau_{k-1}^* := \frac{-(2\pi)^{2k} k}{(2k)!(2k+1)^2}$$

and q is defined in (2.7). If $s = 2+2r$ the $O(m^q/n^q)$ term can be replaced by $o(m^q/n^q)$.
(*b*) *The covariance between $\widehat{d}_{LP}(0, m_i)$ and $\widehat{d}_{LP}(0, m_j)$ is given by*

$$Cov(\widehat{d}_{LP}(0, m_i), \widehat{d}_{LP}(0, m_j)) = \frac{\pi^2}{24m} \frac{1}{\max(l_i, l_j)}(1 + o(1)).$$

The lemma shows that we can actually ignore the $o(\cdot)$ term in the asymptotic bias formula (2.6). This is crucial when proving the next theorem which is a corollary of the lemma.

Theorem 3.2 *Let r and l be fixed constants and $\delta := \tau/(\tau_r^* \sum_{i=1}^K l_i^{2+2r})$ for some $\tau \neq 0$. Then, under Assumptions AG1-AG4, we have*

$$\begin{aligned} E\widehat{d}_{WLP} - d &= \tau b_{2+2r} \frac{m^{2+2r}}{n^{2+2r}} + O\left(\frac{m^q}{n^q}\right) + O(\frac{\ln^3 m}{m}), \\ \text{var}(\widehat{d}_{WLP}) &= \frac{\pi^2}{24m} c_r^* (1+o(1)), \textit{ where} \\ c_r^* &:= \left[\left(Z'\Omega^{-1}Z\right)^{-1}\right]_{1,1} = \sum_{i,j=1}^{K} \frac{w_i w_j}{\max(l_i, l_j)}, \end{aligned}$$

where the $O(m^q/n^q)$ term can be replaced by $o(m^q/n^q)$ if $s = 2 + 2r$.

Remarks (a) If $\lim(m/n^\kappa) \in (0,\infty)$ for $1 > \kappa > (2+2r)/(3+2r)$, then the dominant bias term of $\widehat{d}_{WLP}(r,m)$ is $\tau b_{2+2r} m^{2+2r}/n^{2+2r}$. From (2.6), the dominant bias term of $\widehat{d}_{LP}(r,m)$ is $\tau_r b_{2+2r} m^{2+2r}/n^{2+2r}$. Therefore, the asymptotic biases of the two estimators converge to zero at the same rate. In particular, the rate is quicker than that of the GPH estimator if $r > 0$ and $b_2 \neq 0$.

(b) Suppose $\lim(m/n^\kappa) \in (0,\infty)$ for some $0 < \kappa < 1$ and the assumptions of Theorem 3.2 hold. Then the asymptotic MSE of $\widehat{d}_{WLP}$ satisfies

$$\text{AMSE}(\widehat{d}_{WLP}) = \tau^2 b_{2+2r}^2 (\frac{m}{n})^{4+4r}(1+o(1)) + O(\frac{m^{1+2r}\ln^3 m}{n^{2+2r}}) + \frac{\pi^2}{24}\frac{c_r^*}{m}(1+o(1)). \quad (3.9)$$

If $\kappa > (2+2r)/(3+2r)$, then the $O(\cdot)$ term in (3.9) is of smaller order than the other two terms. Ignoring the $O(\cdot)$ term and assuming that $\tau^2 b_{2+2r}^2 \neq 0$, minimization of $\text{AMSE}(\widehat{d}_{WLP})$ with respect to m gives the AMSE-optimal choice of m :

$$m_{opt} = \left[\left(\frac{\pi^2 c_r^*}{24(4+4r)\tau^2 b_{2+2r}^2} \right)^{1/(5+4r)} n^{(4+4r)/(5+4r)} \right]. \quad (3.10)$$

Note that the AMSE-optimal growth rate of $n^{(4+4r)/(5+4r)}$ allows one to ignore the $O(\cdot)$ term in (3.9). The bandwidth m_{opt} depends on the unknown b_{2+2r}. Guggenberger and Sun (2003) propose a plug-in method to estimate m_{opt}.

(c) To compare the asymptotic properties of $\widehat{d}_{WLP}(r,m)$ and $\widehat{d}_{LP}(r,m)$, we choose $\tau = \tau_r$ which makes the dominant bias terms of the two estimators equal. Table I demonstrates that l can be chosen such that the constant c_r^* in the asymptotic variance of $\widehat{d}_{WLP}(r)$ is smaller than c_r, that is, while having the same asymptotic bias properties, the asymptotic variance properties of the new estimator are better than those of $\widehat{d}_{LP}(r,m)$. We have $c_r^* = c_r$ for $r = 0$ and $r = 1$ if $l_K = 1.2$ and $l_K = 1.85$, respectively, and for larger choices of l_K, the weighted estimator has smaller asymptotic variance than the corresponding $\widehat{d}_{LP}(r)$ estimator in AG with same asymptotic bias. Comparing with the GPH estimator $\widehat{d}_{LP}(0,m)$, the estimator $\widehat{d}_{WLP}(r,m)$ can be designed such that it improves (if $r > 0$ and $b_2 \neq 0$) or retains (if $r = 0$) the asymptotic bias and at the same time reduces the asymptotic variance.

(d) From numerical results as the ones reported in Table I, we conclude that for each r and $\varepsilon > 0$, l can be chosen such that $c_r^* \leq \varepsilon$. Therefore, *asymptotic theory* suggests that larger values of l_K in the definition of $\widehat{d}_{WLP}(r,m)$ imply better variance properties while the asymptotic bias can remain unchanged. This is not surprising because by Lemma 3.1(b) larger l values deliver LP estimates with smaller variance. However, in *finite samples* larger values for l_K have a negative effect on the bias properties of $\widehat{d}_{WLP}$, that is, just as for the bandwidth m, there is a bias/variance trade-off when choosing l in finite samples. The reason for the negative effect is that the approximation of $\ln(g_j/g_0)$ by its Taylor polynomial $b_2\lambda_j^2/2 + ... + b_{2+2r}\lambda_j^{2+2r}/(2+2r)!$ can be very poor for large values of j. Large values for l imply large values for the m_i and the bias reduction does not work as well as for smaller values of l.

Table I. Values of c_r^* for Different l Sequences
(where $l_1 = 1, \tau = \tau_r$, and $l_{j+1} - l_j = .05$ for $j = 1, ..., K-1$)
Recall that $c_0 = 1$, $c_1 = 2.25$, and $c_2 = 3.52$

l_K	1.5	1.85	2.0	2.5	3.0	3.5	4.0	4.5	5.0
c_0^*	.96	.90	.86	.76	.66	.59	.52	.47	.43
c_1^*	3.86	2.25	1.97	1.46	1.19	1.01	.88	.78	.70
c_2^*	31.87	6.48	4.57	2.42	1.77	1.44	1.23	1.08	.97

Using the results of Hurvich et al. (1998), Robinson (1995b), and AG, we now establish the asymptotic normality of $\widehat{d}_{WLP}$:

Theorem 3.3 *Under the assumptions of Theorem 3.2 and AS4, we have*

$$\sqrt{m}(\widehat{d}_{WLP} - d - \tau b_{2+2r} m^{2+2r} n^{-(2+2r)}) \to_d N(0, \frac{\pi^2}{24} c_r^*).$$

For $r \geq 1$, the assumptions of the theorem allow one to take m much larger than for the asymptotic normality result for the GPH estimator (compare to Theorem 2 in Hurvich et al. (1998)). Therefore, by choosing m appropriately, one has asymptotic normality of $\widehat{d}_{WLP}(r)$ with a faster rate of convergence than is possible with $\widehat{d}_{GPH}$. This feature is also obtained for the estimator $\widehat{d}_{LP}(r)$. The difference between $\widehat{d}_{LP}(r)$ and $\widehat{d}_{WLP}(r)$ is that the asymptotic variance of the latter estimator is smaller than that of the former estimator when l is chosen such that $c_r^* < c_r$.

3.2 The Weighted Local Whittle (WLW) Estimator

In this subsection, we investigate the properties of the weighted estimator $\widehat{d}_{We}$ defined in (3.7) when e stands for a k-step local Whittle estimator $\widehat{d}_{LW,k}$. We show that for given $r \geq 1$, $\widehat{d}_{WLW}$ can be constructed to share the higher-order bias advantages of $\widehat{d}_{LW}(r)$ but offers the additional advantage of a smaller variance than $\widehat{d}_{LW}(r)$ and computational simplicity.

We first introduce the k-step estimator $\widehat{d}_{LW,k}$ and motivate its choice. As shown in the Appendix (see (A.6) and Lemma A.2 for $r = 0$), for each $k \geq 1$, $\widehat{d}_{LW,k}$ has the same asymptotic $N(0, 1/4)$ distribution as the LW estimator $\widehat{d}_{LW}(0)$ defined in (2.5) but has a closed form expression. Hence, while sharing its relevant theoretical properties, $\widehat{d}_{LW,k}$ has computational advantages over $\widehat{d}_{LW}(0)$. While in our simulations below we take $k = 1$, we give the theory for general $k \geq 1$.

For notational convenience, we write $\mathcal{R}(d) := \mathcal{R}_0(d, b)$; see equation (2.4). Let $S_m(d) := m\mathcal{R}'(d)$ and $H_m(d) := m\mathcal{R}''(d)$ be the normalized first (score) and second (Hessian) derivatives of $\mathcal{R}(d)$, respectively. Define

$$J_m := X^*(m)'X^*(m), \qquad (3.11)$$

where $X^*(m)$ is an m-vector with j-th element $X_j^*(m)$ given by

$$X_j^*(m) := -2\ln\lambda_j - m^{-1}\sum_{k=1}^{m}(-2\ln\lambda_k).$$

For notational simplicity, we usually write X_j^* for $X_j^*(m)$. The k-step (Gauss-Newton-type) local Whittle estimator is defined recursively by[7]

$$\widehat{d}_{LW,j}(m) := \widehat{d}_{LW,j-1}(m) - J_m^{-1} S_m(\widehat{d}_{LW,j-1}(m)), \; j = 1, 2, ..., k, \tag{3.12}$$

where $\widehat{d}_{LW,0}(m)$ is a $\sqrt{m}$-consistent estimator of d. For example, we can define $\widehat{d}_{LW,0}(r, m) := \widehat{d}_{WLP}(r, m)$.

Let $\widehat{\mathbf{d}}_{LW,k}$ be the K-vector with i-th element $\widehat{d}_{LW,k}(m_i)$. The weighted LW estimator $\widehat{d}_{WLW}$ is obtained by replacing $\widehat{\mathbf{d}}_e$ in equation (3.7) by $\widehat{\mathbf{d}}_{LW,k}$, i.e.

$$(\widehat{d}_{WLW}, \widehat{\beta})' := \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}\widehat{\mathbf{d}}_{LW,k}. \tag{3.13}$$

Note that in the notation we suppressed the dependence on r, m, l, and δ. Using Lemma A.2 in the Appendix, the following theorem can be proved:

Theorem 3.4 *Let r and l be fixed constants and $\delta := \tau/(\tau_r^* \sum_{i=1}^K l_i^{2+2r})$ for some $\tau \neq 0$. Then under Assumptions AS1-AS4, we have*

$$\sqrt{m}(\widehat{d}_{WLW} - d - (\tau/\tau_r^*) D_r m^{2+2r} n^{-(2+2r)}) \to_d N(0, \frac{1}{4} c_r^*)$$

for $D_r = D_r(b) \in \mathbb{R}$ defined in (A.21) in the proof of the theorem, e.g. $D_0 = \tau_0^* b_2$ and $D_1 = \tau_1^* \left(b_4 + (2/9) b_2^2\right)$.

Remarks (a) The asymptotic bias of $\widehat{d}_{WLW}(r)$ equals $(\tau/\tau_r^*) D_r m^{2+2r} n^{-(2+2r)}$ and is thus of order $m^{2+2r} n^{-(2+2r)}$. Therefore, for $r > 1$, the asymptotic bias is smaller than that of $\widehat{d}_{LW}(0)$ by an order of magnitude; recall that for $\widehat{d}_{LW}(0)$ the asymptotic bias is of order m^2/n^2. Furthermore, using the weighted version $\widehat{d}_{WLW}(r)$ of the local Whittle estimator $\widehat{d}_{LW}(0)$ changes the asymptotic variance by the multiplicative constant c_r^*. If c_r^* is smaller than 1, then the asymptotic variance of the weighted estimator is smaller than that of $\widehat{d}_{LW}(0)$.

(b) The estimator $\widehat{d}_{WLW}(r)$ can be modified to have the same dominant asymptotic bias term as $\widehat{d}_{LW}(r)$. More specifically, define

$$\widetilde{d}_{WLW}(r) := \widehat{d}_{WLW}(r) - \left(\tau/\tau_r^* \widehat{D}_r - \tau_r \widehat{b}_{2+2r}\right) m^{2+2r} n^{-(2+2r)} \tag{3.14}$$

for consistent estimators $\widehat{b}_{2+2r}$ of b_{2+2r} and $\widehat{D}_r$ of D_r. Guggenberger and Sun (2003, Proposition 11) give examples of such estimators. Then, under the assumptions of Theorem 3.4 we have

$$\sqrt{m}\left[\widetilde{d}_{WLW}(r) - d - \tau_r b_{2+2r} m^{2+2r} n^{-(2+2r)}\right] \to_d N(0, \frac{1}{4} c_r^*).$$

Therefore, the estimator $\widetilde{d}_{WLW}(r)$ has the same dominant asymptotic bias term $\tau_r b_{2+2r} m^{2+2r} n^{-(2+2r)}$ as $\widehat{d}_{LW}(r)$ but asymptotic variance depending on c_r^* rather than c_r. If l is chosen appropriately, we can have $c_r^* < c_r$ (see Table I) implying that $\widetilde{d}_{WLW}(r)$ has better asymptotic variance properties than $\widehat{d}_{LW}(r)$. In consequence, $\widetilde{d}_{WLW}(r)$ has the same asymptotic advantage over $\widehat{d}_{LW}(r)$, i.e. same asymptotic bias but smaller asymptotic variance, as $\widehat{d}_{WLP}(r)$ has over $\widehat{d}_{LP}(r)$. Note that there is no clear advantage of $\widetilde{d}_{WLW}$ over $\widehat{d}_{WLW}$. The adjustment in (3.14) may increase or decrease the asymptotic bias of $\widehat{d}_{WLW}$, depending on the values of $\widehat{D}_r$ and $\widehat{b}_{2+2r}$.

(c) It follows from Table I that we can choose l such that $c_0^* < 1$. Therefore, the weighted estimator $\widetilde{d}_{WLW}(0)$ has the same asymptotic bias as the Gaussian semiparametric estimator $\widehat{d}_{LW}(0)$ but smaller asymptotic variance. This result is surprising as it is widely believed that the Gaussian semiparametric estimator attains the semiparametric efficiency bound (Robinson (1995a), p. 1640). However, our result does not constitute a contradiction to this belief, because the efficiency of the Gaussian semiparametric estimator is conjectured for a fixed bandwidth while $\widetilde{d}_{WLW}(0)$ is based on a number of different bandwidths.

(d) Results in AG and AS show that $\widehat{d}_{LP}(r,m)$ and $\widehat{d}_{LW}(r,m)$ have the same dominant bias term $\tau_r b_{2+2r}(m/n)^{(2+2r)}$. In contrast, Theorems 3.2 and 3.4 imply that the dominant bias terms $\tau b_{2+2r}(m/n)^{(2+2r)}$ and $(\tau/\tau_r^*)D_r(m/n)^{(2+2r)}$ of the weighted estimators $\widehat{d}_{WLP}(r)$ and $\widehat{d}_{WLW}(r)$ are different for $r \geq 1$ with the difference depending on b.

(e) Assume $\lim(m/n^\kappa) \in (0,\infty)$ for some $0 < \kappa < 1$. It follows that under the assumptions of Theorem 3.4 the asymptotic mean-squared error of $\widehat{d}_{WLW}$ is given by

$$\text{AMSE}(\widehat{d}_{WLW}) = \frac{D_r^2\tau^2}{\tau_r^{*2}}(\frac{m}{n})^{4+4r}(1+o(1)) + \frac{c_r^*}{4m}(1+o(1)).$$

The AMSE-optimal choice of m is thus given by

$$m_{opt} = \left[\left(\frac{c_r^*\tau_r^{*2}}{(4+4r)4D_r^2\tau^2}\right)^{1/(5+4r)} n^{(4+4r)/(5+4r)}\right]. \tag{3.15}$$

A plug-in method to estimate m_{opt} has been proposed in Guggenberger and Sun (2003).

4 Adaptive Estimation

The estimator $\widehat{d}_{We}(r,m,l)$ depends on the choice of r and m, the number of dominant bias terms to be eliminated and the bandwidth, respectively.

Using the plug-in method suggested in Hurvich and Deo (1999) and also used in AG, one can choose m to minimize the asymptotic MSE of the estimator of d for given r and s large enough. The formulae for the optimal m in (3.10) and (3.15) depend on the unknown parameter b_{2+2r} or D_r The plug-in method replaces the unknown parameter by a consistent estimate and can be adapted to our situation, see Guggenberger and Sun (2003). However, this method has several drawbacks. First, it is not fully automatic because it depends on a first stage regression where one has to make an initial choice of the bandwidth. Second, because r is assumed fixed, the method does not adapt to the local smoothness of the spectrum at zero and therefore, in general, the MSE convergence of the estimator to zero is not rate optimal.

In this section, we therefore choose a different route that adapts the choice of r and m to the smoothness of the spectrum at 0. The basic method for this approach comes from Lepskii (1990) and has been used in the context of estimation of the long-memory parameter by Giraitis et al. (2000), Iouditsky et al. (2001), Hurvich et al. (2002), and AS.

According to our above definition of "smooth of order s at zero," a function that is smooth of order s_1 is also smooth of order s_2 whenever $s_1 > s_2$. The rate optimal choice of r is an increasing function of the smoothness of the spectrum at zero and we are therefore interested in determining the "maximal degree s_0 of smoothness at 0". In the following we first precisely define the quantity s_0 and then describe a procedure to consistently estimate it. More precisely, we use the weighted local Whittle estimator to design a consistent estimator $\widehat{s}$ of s_0. As described in the following this estimator can then be used to construct an adaptive weighted estimator. A similar analysis could be done for the weighted log-periodogram estimator. However, such an analysis would make the proof section much longer and we therefore focus on the weighted local Whittle estimator only.

We now define the "maximal degree of smoothness s_0 at 0" of an even function h. This definition is then applied to $\ln g$. In the following, denote by U a generic neighborhood of zero that may be different across different occurrences.

Definition 4.1 *Let $h \in C^1(U)$ be even and let*
$T := \sup\{t \in \mathbb{N}_0 : \exists e_{2i} \in \mathbb{R}$ *for* $i = 1, ..., t$ *s.t.*

$$R_t(\lambda) := h(\lambda) - (h(0) + \sum_{k=1}^{t} e_{2k}\lambda^{2k})$$

satisfies $R_t(\lambda) = o(|\lambda|^{2t})$ *as* $\lambda \to 0\}$.

If $T = \infty$, then define the "maximal degree of smoothness of h at 0" as $s_0 := \infty$. Otherwise, assume that for some positive, finite constants $\tau > 2T, C_{\min}$, and $C_{\max}$

$$C_{\min}|\lambda|^{\tau} \leq |R_T(\lambda)| \leq C_{\max}|\lambda|^{\tau} \tag{4.1}$$

for $\lambda \in U$. Then define $s_0 := \tau$.[8]

Note that if g is smooth of order s at zero (as used in Assumptions AG2 and AS2), then AG (p. 682) shows that $h = \ln g$ has an $[s]$-order Taylor expansion with coefficients $e_k := 1/(k!)\,(d^k/d\lambda^k)\ln g(\lambda)\big|_{\lambda=0}$ and remainder term in $o(|\lambda|^{[s]})$, and, because $\ln g$ is an even function, only even monomials λ^{2k} appear in the expansion. Therefore, in the above definition $2T \geq [s]$ holds for $h = \ln g$, but possibly the inequality is strict. For the rate optimal choice of r we need to know the "maximal" s for which g is smooth of order s at zero. However, seemingly easy definitions for maximal degree of smoothness at 0 for $h \in C^1(U)$ such as $s_0 := \sup\{s \geq 1;\ h$ is smooth of order s at zero$\}$ cause technical problems. For example, it is unclear whether a function that has $s_0 = 2$ according to this definition necessarily allows for a Taylor expansion up to order 2 and therefore in general, it is unclear how many bias terms can be eliminated. We therefore choose the above definition that avoids such technical problems. This definition is closely related to the definition of the sets $\mathcal{F}(s, a, \delta, K)$ in AG (p. 688) and AS (p. 584) and reflects the fact that, all that is needed in Theorems 3.2 and 3.4 to achieve the faster rate of bias convergence, is a Taylor expansion with appropriately bounded remainder term.

We now propose an estimator for the maximal degree of smoothness s_0 of $\ln g$ at 0. Assume we have fixed l with $l_1 = 1$. Let $1 \leq s_* < s^* < \infty$ be constants such that $s_* \leq s_0$. As in AS, for a positive constant ψ_1, define

$$\begin{aligned} m_s &:= \psi_1 n^{\frac{2s}{2s+1}}, \\ r(s) &:= w \text{ for } s \in (2w, 2w+2] \text{ for } w = 0, 1, ..., \\ \zeta(n) &:= (\ln n)(\ln \ln(n))^{1/2}, \text{ and} \\ h &:= 1/\ln n. \end{aligned} \tag{4.2}$$

By $\mathcal{S}_h$ we denote the h-net of the interval $[s_*, s^*]$, that is

$$\mathcal{S}_h := \{\tau : \tau = s_* + kh, \tau \leq s^* \text{ for } k = 0, 1, 2, ...\}.$$

Our proposed estimator for s_0 is given by

$$\begin{aligned} \widehat{s} &:= \max \Big\{ \tau_2 \in \mathcal{S}_h : \left| \widehat{d}_{WLW}(\tau_1) - \widehat{d}_{WLW}(\tau_2) \right| \leq \\ &\quad (m_{\tau_1})^{-1/2} \psi_2 (c^*_{r(\tau_1)}/4)^{1/2} \zeta(n) \text{ for all } \tau_1 \leq \tau_2, \tau_1 \in \mathcal{S}_h \Big\}, \end{aligned} \tag{4.3}$$

where ψ_2 is a positive constant and for notational simplicity we write $\widehat{d}_{We}(\tau)$ for $\widehat{d}_{We}(r(\tau), m_\tau, l)$ for $e = LP$ and LW. Define $m_{i,\tau} := l_i m_\tau$. We use $\widehat{d}_{LW,0}(m_{i,\tau}) := \widehat{d}_{WLP}(\tau)$ for all i as the zero-step estimator in the definition of the k-step LW estimator $\widehat{d}_{WLW}(\tau)$. Graphically, one can view the bound in the definition of $\widehat{s}$ as a function of τ_1. Then, $\widehat{s}$ is the largest value of $\tau_2 \in \mathcal{S}_h$ such that $|\widehat{d}_{WLW}(\tau_1) - \widehat{d}_{WLW}(\tau_2)|$ lies below the bound for all $\tau_1 \leq \tau_2$, $\tau_1 \in \mathcal{S}_h$. Calculation of $\widehat{s}$ is carried out by considering successively larger τ_2 values s_*, $s_* + h$, $s_* + 2h$, ... until for some τ_2 the deviation $|\widehat{d}_{WLW}(\tau_1) - \widehat{d}_{WLW}(\tau_2)|$ exceeds the bound for some $\tau_1 \leq \tau_2$, $\tau_1 \in \mathcal{S}_h$.

Definition 4.2 *The adaptive weighted local Whittle (AWLW) estimator is defined by*

$$\widehat{d}_{AWLW} := \widehat{d}_{WLW}(r(\widehat{s}), m_{\widehat{s}}, l)$$

for $\widehat{s}$ defined in (4.3) and l being user-chosen constants.

Through $\widehat{s}$, AWLW depends on several user-chosen constants, namely ψ_1, ψ_2, s_*, and s^*. Other adaptive estimators, for example, Giraitis et al. (2000), Hurvich (2001), and Iouditsky et al. (2001) all also depend on user-chosen constants, see AS (p. 587, Comment 5) or the Monte Carlo Section 5 below for further discussion. In Section 5 we make suggestions for the choice of ψ_1 and ψ_2. For the bounds s_* and s^* we suggest using 1 and ∞, respectively.

The following theorem establishes important asymptotic properties of $\widehat{s}$ and $\widehat{d}_{AWLW}$. The symbol U denotes a neighborhood of zero.

Theorem 4.3 *Let Assumptions AS1, AS2(ii), and AS3 hold and suppose* $\ln g$ *has maximal degree of smoothness* $s_0 \geq s_* \geq 1$ *at* 0. *If* $T < \infty$ *in Definition 4.1, assume in addition that*

$$(C_{\max} - C_{\min})e^{-s_0-1} < C_{\min} s_0 \tag{4.4}$$

and $R_T(\lambda) > 0$ *for all* $\lambda \in U$ *or* $R_T(\lambda) < 0$ *for all* $\lambda \in U$. *Then*

(a) $\widehat{s} = \min(s_0, s^*) + O_p\left(\frac{\ln\ln n}{\ln n}\right)$ *as* $n \to \infty$ *and*

(b) $\lim_{C\to\infty} \operatorname{limsup}_{n\to\infty} \sup_{s_0\in[s_*,s^*]} P\left(n^{\frac{s_0}{2s_0+1}}\zeta^{-1}(n)\left|\widehat{d}_{AWLW} - d\right| > C\right) = 0.$ (4.5)

Remarks (a) The theorem shows that $\widehat{s}$ consistently estimates the maximal degree of smoothness s_0 when it is finite and s_* and s^* are appropriately chosen. The additional assumptions made under finite T are weak technicalities that are needed in the proof of the theorem.

(b) Theorem 3 in AG shows that the optimal rate of convergence for estimation of d for spectral densities in their set $\mathcal{F}(s_0, a, \delta, K)$ is given by $n^{s_0/(2s_0+1)}$ when s_0 is finite and known. Theorem 4.3(b) shows that AWLW achieves this rate up to a logarithmic factor $\zeta^{-1}(n)$ when s_0 is finite and *not* known . When s_0 is not known, this extra logarithmic factor is an unavoidable price for adaptation and the adaptive estimator of AS shares this property.

(c) Like the adaptive estimator in AS, the adaptive estimator $\widehat{d}_{AWLW}$ is not necessarily asymptotically normal. However, at the cost of a slower rate of convergence, an adaptive estimator can be constructed that is asymptotically normal with zero asymptotic bias by altering the definition of m_s so that m_s diverges to infinity at a slower rate than $n^{2s/(2s+1)}$. More specifically, after obtaining $\widehat{s}$ using the above adaptive procedure, we define

$$\widehat{d}^*_{AWLW} := \widehat{d}_{WLW}(r(\widehat{s}), m^*_{\widehat{s}}, l), \text{ where } m^*_s = \psi_1 n^{4r(s)/(4r(s)+1)}.$$

If $s_0 < \infty$ and s_0 is not an even integer, Part (a) of Theorem 4.3 implies that $r(\widehat{s}) = r(s_0)$ with probability approaching one. Thus, both $r(\widehat{s})$ and $m^*_{\widehat{s}}$ are essentially nonrandom for large n. In consequence, the adaptive estimator $\widehat{d}^*_{AWLW}$ is asymptotically normal:

$$\sqrt{m^*_{\widehat{s}}}\left(\widehat{d}^*_{AWLW} - d\right) \to_d N(0, \frac{1}{4}c^*_{r(s_0)}).$$

Of course, one would expect that a given level of accuracy of approximation by the normal distribution would require a larger sample size when r and m are adaptively selected than otherwise.

5 Monte Carlo Experiment

5.1 Experimental Design

In this section, we present some simulation results that compare the RMSE performance of the AWLW estimator with several adaptive estimators in the literature. The simulation design is taken from AS. Additional simulation results for non-adaptive plug-in weighted estimators are given in Guggenberger and Sun (2003).

5.1.1 Models

As in AS, we consider three models and several parameter combinations for each model.

The first model we consider for the time series $\{Y_t : t \geq 1\}$ is a first-order autoregressive fractionally integrated (ARFIMA (1, d, 0)) model with autoregressive (AR) parameter ϕ and long-memory parameter d_0:

$$(1 - \phi L)(1 - L)^{d_0} Y_t = u_t, \tag{5.1}$$

where the innovations $\{u_t : t = ..., 0, 1, ...\}$ are iid random variables and L denotes the lag operator. We consider three distributions for u_t: standard normal, t_5, and χ^2_2. The t_5 and χ^2_2 distributions are considered because they exhibit thick tails and asymmetry, respectively. All of the estimators of d_0 that we consider are invariant with respect to the mean and variance of the time series. In consequence, the choice of location and scale of the innovations is irrelevant. Note that the spectral density of an ARFIMA(1, d, 0) process is in $C^\infty((0, \pi])$.

The second model is a stationary ARFIMA(1, d, 0)-*like* model that has a discontinuity in its spectral density at frequency $\lambda = \lambda_0$. We call this model a DARFIMA(1, d, 0) model. Its spectral density is that of an ARFIMA(1, d, 0) process for $\lambda \in (0, \lambda_0]$, but is zero for $\lambda \in (\lambda_0, \pi]$. A DARFIMA(1, d, 0) process $\{Y_t : t \geq 1\}$ is defined as in (5.1), but with innovations $\{u_t : t = ..., 0, 1, ...\}$ that are an iid Gaussian process filtered by a low pass filter. Specifically,

$$u_t = \sum_{j=-\infty}^{\infty} c_j \varepsilon_{t-j} \text{ for } t = ..., 0, 1, ..., \text{ where } c_j = \begin{cases} \frac{\lambda_0}{\pi}, & \text{for } j = 0 \\ \frac{\sin(\lambda_0 j)}{j\pi}, & \text{for } j \neq 0 \end{cases} \tag{5.2}$$

and $\{\varepsilon_t : t = ..., 0, 1, ...\}$ are iid random variables with standard normal, t_5, or χ^2_2 distribution. The spectral density $f_u(\lambda)$ of $\{u_t : t \geq 1\}$ equals $\sigma^2_u/(2\pi)$ for $0 < \lambda \leq \lambda_0$ and equals 0 for $\lambda_0 < \lambda \leq \pi$, where σ^2_u denotes the variance of u_t, see Brillinger (1975, equation (3.3.25)). The spectral density of $\{Y_t : t \geq 1\}$ is $f_u(\lambda)$ times the spectral density of the ARFIMA(1, d, 0) process that has the same AR parameter. Thus, the spectral density of a DARFIMA process is a truncated discontinuous version of that of the corresponding ARFIMA process.

The third model is called a *long-memory components* (LMC) model. It is designed to have a finite degree of smoothness s_0 at frequency zero in the short-run part, $g(\lambda)$, of its spectral density. The process $\{Y_t : t \geq 1\}$ is defined by

$$(1 - L)^{d_0} Y_t = u_t + k(1 - L)^{s_0/2} v_t, \tag{5.3}$$

where $\{u_t : t \geq 1\}$ and $\{v_t : t \geq 1\}$ are independent iid processes both with normal, t_5, or χ_2^2 distribution. The spectral density function of $\{u_t + k(1-L)^{s_0/2} v_t : t \geq 1\}$ is

$$g_{s_0,k}(\lambda) = \frac{\sigma^2}{2\pi} + \frac{k^2\sigma^2}{2\pi}|1 - e^{i\lambda}|^{s_0}, \tag{5.4}$$

where σ^2 denotes the variance of u_t and v_t. Because $|1 - e^{i\lambda}| \sim \lambda$ as $\lambda \to 0$, the smoothness of $g_{s_0,k}(\lambda)$ at $\lambda = 0$ is s_0.

For the ARFIMA and DARFIMA models, we consider seven values of ϕ, viz., 0, .3, .6, .9, $-.3$, $-.6$, and $-.9$. For the DARFIMA model, we take λ_0 to equal $\pi/2$. For the LMC model, we take $s_0 = 1.5$ and consider five values of k, viz., $1/3, 1/2, 1, 2,$ and 3. For all three models, we consider three values of d_0, viz., $d_0 = -.4, 0,$ and .4. For each model, we consider two sample sizes $n = 512$ and $n = 4,096$. In all cases, 1000 simulation repetitions are used. This produces simulation standard errors that are roughly 3% of the magnitudes of the reported RMSEs.

5.1.2 Estimators

We consider the AWLW estimator defined in the previous section, the adaptive local polynomial Whittle estimator (ALPW) of AS, the adaptive log-periodogram regression estimator of Giraitis et al. (2000), the adaptive FEXP estimator of Iouditsky et al. (2001), and the FEXP estimator of Hurvich (2001) with the number of terms in the expansion chosen by his local C_L method. Each of these estimators requires the specification of certain constants in the adaptive or local C_L procedure. In addition, the AWLW estimator requires choosing the l sequence, the estimator analyzed by Giraitis et al. (2000) requires trimming of frequencies near zero and tapering of the periodogram, and the estimator analyzed by Iouditsky et al. (2001) requires tapering of the periodogram and allows for pooling of the periodogram.

The constants in the adaptive procedures are tuned to the Gaussian ARFIMA model with $\phi = .6$ with $n = 512$. That is, they are determined by simulation to be the values (from a grid) that yield the smallest RMSE for the Gaussian ARFIMA model with $\phi = .6$ and $n = 512$. These values are then used for all of the processes considered in the experiment. For AWLW and ALPW procedures, the grids for ψ_1 and ψ_2 are $\{.1, .15, .2, ..., .5\}$ and $\{.05, .10, ..., .70\}$, respectively. For the AWLW procedure, we consider three equally-spaced l sequences with increments $l_{j+1} - l_j$ of .05, $l_1 = 1$, and $l_K = 1.5, 2,$ and 3. For the procedure in Giraitis et al. (2000), we set their constant β^* equal to 2 (as suggested on their p. 192). In addition, we introduce two constants ψ_1 and ψ_2 that are analogous to the constants that appear in the definition of AWLW. The grid for ψ_1 is $\{.1, .2, ..., 1.0\}$ and the grid for ψ_2 is $\{.05, .10, ..., .70\}$. The constants ψ_1 and ψ_2 are introduced in order to give the adaptive procedure in Giraitis et al. (2000) a degree of flexibility that is comparable to that of the AWLW procedure. For the procedure in Iouditsky et al. (2001), their constant κ is analogous to the constant ψ_2 of the AWLW estimator and is chosen from the same grid as ψ_2 and the pooling size (denoted *pool* below and m in the notation of Iouditsky et al. (2001)) is determined simultaneously with the constant κ from the grid $\{1, 2, ..., 6\}$.

We analyze several groups of adaptive estimators. The first group contains those adaptive estimators that are *closest* to being covered by the theoretical results in the literature. We refer to these as being the "theoretically-justified" estimators and they are denoted AWLW, ALPW, GRS1, and IMS1. (The AWLW estimator *is* covered by the results of this paper and the ALPW estimator is covered in AS. The GRS1 estimator uses constants ψ_1 and ψ_2 that are not covered by the results of Giraitis et al. (2000) and the IMS1 estimator uses a constant κ and an upper bound on the number p of Fourier terms that are not covered by the results of Iouditsky et al. (2001), see footnote 4 in AS.) The GRS1 estimator uses the cosine-bell taper with two out of every three frequencies dropped (as in Giraitis et al. (2000)) and three frequencies near the origin are trimmed ($trim = 3$) when $n = 512$ and six are trimmed ($trim = 6$) when $n = 4,096$. The IMS1 estimator uses the Hurvich taper (of order one) with one frequency dropped between each pool group of frequencies, as in Section 2.2 of Hurvich et al. (2002). (Note that for the case where no differencing is carried out to eliminate potential trends, the adaptive estimators in Iouditsky et al. (2001) and Hurvich et al. (2002) are essentially the same except that the latter uses a scheme that deletes fewer frequencies, which we employ here.)

The second group of adaptive estimators that we consider does not have known theoretical properties. The estimator GRS2 differs from GRS1 in that it does not trim any frequencies near zero. The estimator GRS3 differs from GRS1 in that it does not trim frequencies near zero, use a taper, or drop two out of every three frequencies. The estimator IMS2 differs from IMS1 in that it does not use a taper or drop any frequencies. Although no asymptotic theory has been established for these estimators, it is expected that these estimators have better performance than the respective first version as both trimming and tapering reduce the efficiency of the estimators.

The constants determined by simulation are: AWLW: $(\psi_1, \psi_2) = (.25, .05)$ and $l = (1.0, 1.05, ..., 2)$; ALPW: $(\psi_1, \psi_2) = (.3, .2)$; GRS1: $(\psi_1, \psi_2) = (.6, .5)$; GRS2: $(\psi_1, \psi_2) = (.3, .6)$; GRS3: $(\psi_1, \psi_2) = (.2, .25)$; IMS1: $(pool, \kappa) = (2, .65)$; and IMS2: $(pool, \kappa) = (2, .45)$.

Hurvich's (2001) procedure requires the specification of a constant α. We consider the two values $\alpha = .5$ and $\alpha = .8$ that are considered in Hurvich (2001). The corresponding estimators are denoted H1 and H2. (The value $\alpha = .8$ turns out to minimize the RMSE of the FEXP estimator for the ARFIMA process with $\phi = .6$ and $n = 512$ over α values in $\{.1, .2, ..., .8\}$.) The theoretical properties of Hurvich's (2001) procedure, such as its rate of convergence, are not given in Hurvich (2001). For this reason, we do not put the H1 and H2 estimators in with the first group of "theoretically-justified" estimators.

The final estimator that we consider is the parametric Whittle quasi-maximum likelihood (QML) estimator for a Gaussian ARFIMA(1, d, 0) model. This estimator is misspecified when the model under consideration is the DARFIMA(1, d, 0) model or the LMC model and is included in the simulations for comparative purposes.

5.2 Monte Carlo Results

Tables II-IV contain the Monte Carlo RMSE results for the three models and the various estimators. Each table has a separate panel of results for $n = 512$ and $n = 4,096$. The first four rows of each panel give results for the "theoretically-justified" adaptive estimators. The next rows give results for the adaptive estimators that do not (currently) have theoretical justification and for the parametric Whittle QML estimator.

For brevity, we only report a subset of the parameter combinations. For example, results for $\phi = -.3$, $-.6$, and $-.9$ are not given because they are close to those for $\phi = 0$ and $.3$. We only give selected results for $d = .4$ and no results for $d = -.4$, because the value of d has only a small effect on RMSE (except for the IMS1 estimator which exhibits some sensitivity to d). Similarly, we only give selected results for t_5 innovations and we give no results for χ^2_2 innovations, because the innovation distribution turns out to have only a small effect on RMSE.

We start by noting some general features of the results presented in the tables. First, for $d = 0$ the results for normal and t_5 innovations are quite similar. Second, the results with normal innovations are very similar for $d = 0$ and $d = .4$ (with the exception of the IMS1 estimator that is noticeably worse when $d = .4$ in all ARFIMA and DARFIMA models for $n = 512$). Thirdly and not surprisingly, the performance of the AWLW and all other estimators improve substantially as n increases from 512 to 4,096 without generally affecting the qualitative ranking of the estimators.

Now, we compare the estimators. Among the four theoretically-justified estimators, the AWLW and ALPW estimators perform the best in an overall sense across all three models, parameter combinations, and sample sizes. Across the 48 scenarios reported in the tables, AWLW and ALPW have lowest RMSE in 35 and 11 cases, respectively, amongst all the theoretically-justified estimators. For ARFIMA and DARFIMA processes with $\phi \leq .6$, the AWLW estimator uniformly outperforms the other semiparametric estimators while for $\phi = 0.9$ it is beaten by ALPW. Similar to the explanation in Remark (d) after Theorem 3.2 regarding the negative effect of large l_K-values on the finite-sample bias properties of AWLW, this result for $\phi = 0.9$ is caused by the poor approximation of $\ln(g_j/g_0)$ by its Taylor expansion in a neighborhood of the origin in this case. As expected, the performance of all the estimators deteriorates with increasing k in the LMC models. Compared to the ALPW estimator, the AWLW estimator has smaller RMSE when $k \leq 2$ and slightly larger RMSE when $k = 3$.

Both GRS1 and IMS1 perform poorly in an absolute sense. Relative to AWLW and ALPW, GRS1 performs poorly when $\phi \geq 0.3$ or $k \geq 1$ while IMS1 is outperformed for all values of ϕ and k in our three models. Because of the broad-band character of the IMS1 estimator, it is not robust against discontinuous spectral densities which is reflected in its worse performance in the DARFIMA model compared to the ARFIMA model.

Next, we consider the adaptive estimators GRS2, GRS3, and IMS2 without theoretical justification. GRS2 and GRS3 perform noticeably better than GRS1, especially when ϕ or k is large. Hence, trimming is found to have a negative impact.

The GRS3 estimator outperforms the GRS2 estimator across all cases and all models considered. Hence, tapering also is found to have a negative impact. In an overall sense, the GRS3 estimator performs reasonably well relative to the other semiparametric estimators. However, with very few exceptions its performance is worse than that of the AWLW estimator. The IMS2 estimator outperforms the IMS1 estimator in all but two cases. In many cases the difference is substantial. Hence, again we find the effect of tapering to be negative. Compared to other semiparametric estimators, IMS2 performs best in the cases where the short-run contamination is highest, i.e., $\phi = .9$ or $k = 3$. But in most other cases it is out-performed by the AWLW, ALPW, and GRS3 estimators. It appears that the relative performance of the IMS2 estimator compared to the narrow-band AWLW, ALPW, and GRS1-3 estimators improves as the sample size increases from 512 to $4,096$. The H1 and H2 estimators perform well when the sample size is $4,096$ and ϕ or k are large. In other cases, they do not perform well relative to the AWLW, ALPW, or GRS3 estimators. In particular, they perform poorly for DARFIMA processes except when $\phi = .9$. The relative strengths of the H1 and H2 estimators are similar to those of the IMS1 and IMS2. This is not surprising, because all of these estimators are broadband FEXP estimators.

The parametric Whittle QML estimator performs as expected. For ARFIMA processes it typically has smaller RMSE than the semiparametric estimators, often substantially smaller, for example when $\phi = .9$ or $n = 4,096$. For DARFIMA processes, for which it is misspecified, it performs very poorly. It is substantially outperformed by all semiparametric estimators. For LMC processes, for which it is misspecified, it has lower RMSE than the semiparametric estimators for small values of k (when the LMC process is close to ARFIMA processes). But, it is outperformed for larger values of k. One surprising finding that is worth mentioning is that the parametric Whittle QML is outperformed by AWLW for certain parameter combinations in the ARFIMA model, for example, for $n = 512$ and $\phi = .3$ and .6.

To conclude, among the four theoretically-justified estimators, the new AWLW estimator seems to be the best. Of all the semiparametric estimators, the two best ones in an overall sense seem to be AWLW and ALPW. When the short run contamination is not too large, AWLW outperforms ALPW and the opposite is true when the short run contamination is large. Trimming hurts the performance of the GRS1 estimator. Tapering hurts the performance of the GRS1 and IMS1 estimators. The narrow-band estimators AWLW and ALPW perform well over a broad range of parameter values, but are outperformed by the broad-band estimator IMS2 when the short run contamination is very large. The broadband estimator IMS2 performs relatively well when the sample size is large and the short run contamination is large. The parametric Whittle QML estimator performs very well when the model is correctly specified, moderately well when the amount of misspecification is small, and disastrously when the amount of misspecification is large.

TABLE II

RMSE for ARFIMA(1,d,0) Processes with AR Parameter ϕ

(a) $n = 512$

	$d=0$ Normal ϕ				$d=0$ t_5 ϕ		$d=.4$ Normal ϕ	
Estimator	0	.3	.6	.9	.6	.9	.6	.9
AWLW	.122	.121*	.134*	.473	.132*	.473	.139*	.472
ALPW	.145	.142	.145	.423	.140	.420	.151	.425
GRS1 (*taper*, *trim* = 3)	.160	.197	.397	.855	.393	.864	.394	.857
IMS1 (*pool* = 2, *taper*)	.234	.252	.291	.448	.311	.460	.379	.579
GRS2 (*taper*, *no trim*)	.164	.172	.244	.662	.254	.657	.253	.668
GRS3 (*no taper*, *no trim*)	.131	.133	.159	.502	.157	.501	.166	.499
IMS2 (*pool* = 2, *no taper*)	.203	.209	.217	.315	.215	.320	.222	.301
H1 ($\alpha = 0.5$)	.288	.309	.321	.438	.308	.420	.310	.436
H2 ($\alpha = 0.8$)	.182	.206	.233	.459	.235	.457	.241	.480
Parametric Whittle QML	.066*	.128	.134*	.112*	.141	.107*	.140	.156*

(b) $n = 4096$

	$d=0$ Normal ϕ				$d=0$ t_5 ϕ		$d=.4$ Normal ϕ	
Estimator	0	.3	.6	.9	.6	.9	.6	.9
AWLW	.046	.046	.050	.224	.048	.227	.054	.230
ALPW	.061	.061	.060	.207	.058	.201	.062	.213
GRS1 (*taper*, *trim* = 6)	.056	.081	.169	.586	.168	.581	.169	.587
IMS1 (*pool* = 2, *taper*)	.045	.133	.122	.218	.120	.220	.142	.273
GRS2 (*taper*, *no trim*)	.063	.066	.108	.396	.109	.388	.114	.401
GRS3 (*no taper*, *no trim*)	.052	.052	.065	.222	.064	.219	.070	.228
IMS2 (*pool* = 2, *no taper*)	.046	.076	.073	.155	.073	.154	.071	.145
H1 ($\alpha = 0.5$)	.067	.076	.083	.125	.082	.123	.084	.132
H2 ($\alpha = 0.8$)	.052	.062	.073	.147	.075	.144	.077	.156
Parametric Whittle QML	.013*	.028*	.046*	.025*	.045*	.025*	.045*	.032*

Notes: Asterisks denote the smallest RMSE across all the estimators for each design.

TABLE III

RMSE for DARFIMA(1,d,0) Processes with $\lambda_0 = \pi/2$ and AR Parameter ϕ

(a) $n = 512$

	$d=0$ Normal ϕ				$d=0$ t_5 ϕ		$d=.4$ Normal ϕ	
Estimator	0	.3	.6	.9	.6	.9	.6	.9
AWLW	.123*	.121*	.134*	.474	.139*	.472	.138*	.472
ALPW	.143	.142	.146	.423	.149	.430	.146	.425
GRS1 (*taper*, *trim* = 3)	.160	.200	.390	.857	.395	.868	.394	.859
IMS1 (*pool* = 2, *taper*)	.448	.448	.445	.464	.468	.490	.520	.578
GRS2 (*taper*, *no trim*)	.176	.181	.243	.656	.254	.657	.253	.668
GRS3 (*no taper*, *no trim*)	.131	.142	.159	.503	.157	.501	.165	.499
IMS2 (*pool* = 2, *no taper*)	.268	.266	.262	.303*	.252	.311*	.246	.299*
H1 ($\alpha = 0.5$)	.356	.348	.317	.404	.326	.401	.309	.427
H2 ($\alpha = 0.8$)	.396	.393	.384	.389	.403	.409	.363	.429
Parametric Whittle QML	.866	.616	.949	1.202	.951	1.209	.852	.855

(b) $n = 4096$

	$d=0$ Normal ϕ				$d=0$ t_5 ϕ		$d=.4$ Normal ϕ	
Estimator	0	.3	.6	.9	.6	.9	.6	.9
AWLW	.046*	.046*	.050*	.224	.052*	.227	.054*	.230
ALPW	.060	.060	.059	.207	.055	.204	.061	.213
GRS1 (*taper*, *trim* = 6)	.056	.081	.169	.586	.168	.581	.169	.587
IMS1 (*pool* = 2, *taper*)	.146	.146	.146	.127	.148	.130	.121	.120*
GRS2 (*taper*, *no trim*)	.063	.066	.108	.396	.109	.388	.114	.401
GRS3 (*no taper*, *no trim*)	.051	.052	.066	.222	.064	.219	.070	.229
IMS2 (*pool* = 2, *no taper*)	.096	.096	.096	.106*	.093	.103*	.100	.125
H1 ($\alpha = 0.5$)	.126	.125	.123	.117	.124	.119	.110	.127
H2 ($\alpha = 0.8$)	.120	.120	.119	.113	.122	.119	.119	.139
Parametric Whittle QML	.346	.579	.930	1.256	.933	1.257	.905	1.068

TABLE IV
RMSE for LCM Model with Smoothness Index $s_0 = 1.5$ and Weight k

(a) $n = 512$

	$d=0$ Normal k					$d=0$ t_5 k		$d=.4$ Normal k	
Estimator	1/3	1/2	1	2	3	1/2	2	1/2	2
AWLW	.121	.122	.128*	.166*	.226	.113	.156*	.119	.161*
ALPW	.139	.139	.139	.168	.216*	.145	.171	.138	.164
GRS1 (*taper*, *trim* = 3)	.172	.190	.230	.384	.486	.172	.384	.177	.375
IMS1 (*pool* = 2, *taper*)	.247	.249	.264	.298	.336	.262	.297	.307	.323
GRS2 (*taper*, *no trim*)	.165	.173	.175	.262	.337	.182	.269	.181	.265
GRS3 (*no taper*, *no trim*)	.131	.131	.136	.181	.245	.121	.184	.128	.176
IMS2 (*pool* = 2, *no taper*)	.207	.207	.209	.209	.229	.194	.215	.215	.233
H1 ($\alpha = 0.5$)	.264	.265	.276	.303	.333	.259	.307	.271	.311
H2 ($\alpha = 0.8$)	.178	.182	.199	.241	.298	.169	.229	.182	.245
Parametric Whittle QML	.064*	.069*	.129	.273	.368	.064*	.269	.069*	.274

(b) $n = 4096$

	$d=0$ Normal k					$d=0$ t_5 k		$d=.4$ Normal k	
Estimator	1/3	1/2	1	2	3	1/2	2	1/2	2
AWLW	.045	.046	.048*	.065*	.095	.046	.067*	.046	.060
ALPW	.061	.060	.060	.065*	.100	.060	.068	.061	.058*
GRS1 (*taper*, *trim* = 6)	.061	.066	.108	.213	.305	.066	.210	.066	.212
IMS1 (*pool* = 2, *taper*)	.056	.075	.123	.136	.166	.075	.138	.070	.105
GRS2 (*taper*, *no trim*)	.068	.068	.079	.137	.201	.067	.137	.066	.135
GRS3 (*no taper*, *no trim*)	.051	.052	.056	.085	.122	.053	.087	.049	.080
IMS2 (*pool* = 2, *no taper*)	.050	.058	.070	.085	.085*	.058	.086	.062	.089
H1 ($\alpha = 0.5$)	.072	.073	.078	.089	.100	.078	.089	.079	.091
H2 ($\alpha = 0.8$)	.056	.058	.065	.087	.107	.060	.089	.061	.087
Parametric Whittle QML	.023*	.033*	.104	.236	.320	.032*	.235	.031*	.232

6 Appendix of Proofs

The next Lemma is needed for the proof of Lemma 3.1. It is a refined version of statements in Lemma 2 of AG.

Lemma A.1 *Suppose Assumption AG4 holds. Then, for J_m defined in (3.11),*
(a) $J_m^{-1} = (X^{*\prime}X^*)^{-1} = \frac{1}{4}\frac{1}{m} + \frac{1}{8}\frac{\ln^2 m}{m^2} + O(\frac{\ln m}{m^2})$ *and*
(b) $\sum_{j=1}^m X_j^* \lambda_j^k = -(2\pi)^k \frac{2k}{(k+1)^2}\frac{m^{k+1}}{n^k} + O(\frac{m^k}{n^k}\ln m)$.

Proof of Lemma A.1. For the proof we need the following approximations:
(i) $\sum_{j=1}^m \ln j = m\ln m - m + \frac{1}{2}\ln m + O(1)$,
(ii) $\sum_{j=1}^m \ln^2 j = m\ln^2 m - 2m\ln m + 2m + \frac{1}{2}\ln^2 m + O(1)$,
(iii) $\sum_{j=1}^m j^k \ln j = \frac{1}{k+1}m^{k+1}\ln m - \frac{1}{(k+1)^2}m^{k+1} + O(m^k \ln m)$, and
(iv) $\sum_{j=1}^m j^k = \frac{1}{k+1}m^{k+1} - \frac{1}{2}m^k + O(m^{k-1})$.

These approximations follow from straightforward calculations using the following version of Euler's summation formula (see, e.g., Walter (1992, p. 349)). For $f \in C^3([1,m])$, we have

$$\sum_{j=1}^m f(j) = \int_1^m f(x)dx + 0.5(f(1) + f(m)) + err,$$

where $err := [2^{-1}E_2(x)f'(x)]_1^m + 6^{-1}\int_1^m f'''(x)E_3(x)dx$, $E_p(x) := B_p(x-[x])$, $B_p(x) := \sum_{j=0}^p \binom{p}{j} B_j x^{p-j}$ is the p-th Bernoulli polynomial, and B_j in this lemma denotes the j-th Bernoulli number. It can be shown that $E_2(m) = 6^{-1}$ and for all real numbers x, $|E_3(x)| < 20^{-1}$. For part (a) of Lemma A.1 then note that

$$X^{*\prime}X^* = 4\sum_{j=1}^m \ln^2 j - 4m^{-1}(\sum_{j=1}^m \ln j)^2 = 4m - 2\ln^2 m + O(\ln m),$$

where the second equality follows from (i) and (ii). Part (a) then follows. For part (b), note that using (i), (iii), and (iv), we have

$$\begin{aligned}\sum_{j=1}^m X_j^*\lambda_j^k &= -2\left(\frac{2\pi}{n}\right)^k \left(\sum_{j=1}^m j^k \ln j - \sum_{j=1}^m j^k \frac{1}{m}\sum_{i=1}^m \ln i\right) \\ &= -2\left(\frac{2\pi}{n}\right)^k \left(\frac{1}{k+1}m^{k+1} - \frac{1}{(k+1)^2}m^{k+1} + O(m^k \ln m)\right) \\ &= -(2\pi)^k \frac{2k}{(k+1)^2}\frac{m^{k+1}}{n^k} + O(\frac{m^k}{n^k}\ln m), \text{ as desired. } \blacksquare\end{aligned}$$

Proof of Lemma 3.1. Part (a) The log-periodogram can be written as

$$\ln I_j = (\ln g_0 - C) - 2d\ln\lambda_j + R_j + \epsilon_j \text{ for } j = 1, ..., m,$$

where $R_j := \ln g_j - \ln g_0$ and $\epsilon_j := \ln(I_j/f_j) + C$, see (2.2). Define the m-vectors $R(m) := (R_1, ..., R_m)'$ and $\epsilon(m) := (\epsilon_1, ..., \epsilon_m)'$. Then

$$\widehat{d}_{LP}(0,m) - d = J_m^{-1}X^*(m)'(\epsilon(m) + R(m)). \tag{A.1}$$

By Lemma 2(a) and (f) in AG,

$$EJ_m^{-1}X^*(m)'\epsilon(m) = O\left(m^{-1}\ln^3 m\right). \tag{A.2}$$

Regarding $J_m^{-1}X^*(m)'R(m)$, first note that under Assumption AG2, a Taylor expansion of $\ln g_j$ about $\lambda = 0$ as in AG (3.3) yields

$$R_j = \sum_{k=1}^{1+r} \frac{b_{2k}}{(2k)!}\lambda_j^{2k} + Rem_j, \tag{A.3}$$

where $b_k := (d^k/d\lambda^k)\ln g(\lambda)\big|_{\lambda=0}$, Rem_j, defined implicitly as the difference between R_j and the sum on the right hand side, satisfies $\max_{1\le j\le m}|Rem_j/\lambda_j^q| = O(1)$ for q defined in (2.7) and $\max_{1\le j\le m}|Rem_j/\lambda_j^q| = o(1)$ if $s = 2+2r$. Combining this with Lemma A.1(b) and using AG (7.16) yields

$$\begin{aligned} X^*(m)'R(m) &= -\textstyle\sum_{k=1}^{1+r}\left[\frac{(2\pi)^{2k}4kb_{2k}}{(2k)!(2k+1)^2}\frac{m^{2k+1}}{n^{2k}} + O(\frac{m^{2k}}{n^{2k}}\ln m)\right] + O(\frac{m^{q+1}}{n^q}), \\ J_m^{-1}X^*(m)'R(m) &= \textstyle\sum_{k=1}^{1+r}\left[\frac{-(2\pi)^{2k}kb_{2k}}{(2k)!(2k+1)^2}\frac{m^{2k}}{n^{2k}} + O(\frac{m^{2k-1}}{n^{2k}}\ln^2 m)\right] + O(\frac{m^q}{n^q}), \end{aligned} \tag{A.4}$$

where the second equation uses Lemma A.1(a). The $O(m^q/n^q)$ term in (A.4) is $o(m^q/n^q)$ if $s = 2+2r$. Note that $O((m^{2k-1}/n^{2k})\ln^2 m)$ is dominated by $O\left(m^{-1}\ln^3 m\right)$ from (A.2). Combining (A.1), (A.2), and (A.4) yields the desired result.

Part (b) Let $Cov := Cov(\widehat{d}_{LP}(0,m_i), \widehat{d}_{LP}(0,m_j))$. Without loss of generality, we assume that $l_i \le l_j$. Because $(X^{*\prime}X^*)^{-1} = 4^{-1}m^{-1}(1+o(1))$, (A.1) implies that

$$Cov = \frac{1}{16m_im_j}(1+o(1))\textstyle\sum_{k=1}^{m_i}\sum_{l=1}^{m_j} X_{ki}^*X_{lj}^*Cov(\epsilon_k,\epsilon_l),$$

where $X_{ki}^* := X_k^*(m_i)$. We decompose the double sum in the formula for Cov into $(\sum_{k,l=1}^{m_i} + \sum_{k=1}^{m_i}\sum_{l=m_i+1}^{m_j})(X_{ki}^*X_{lj}^*Cov(\epsilon_k,\epsilon_l))$.

By Lemma 2(k) in AG, $\max_{1\le k\le m_v}|X_{kv}^*| = O(\ln m_i)$ for $v = i$ and j. Similar calculations to the ones in Lemma A.1(a) yield $\sum_{k=1}^{m_i} X_{ki}^*X_{kj}^* = 4m_i(1+o(1))$. Thus, the proof of Theorem 1 in Hurvich et al. (1998) where they calculate the variance of the GPH-estimator (p. 25) can be modified along the same lines as done in the proof of Theorem 1(b) in AG (p. 702) to prove that the first part of the double sum equals $(4/6)\pi^2m_i(1+o(1))$. It is easily shown that the contribution of the second part is in $o(m_i)$ and thus negligible and thus $Cov = (4\pi^2m_i/96m_im_j)(1+o(1)) = (\pi^2/24l_jm)(1+o(1))$. ■

Proof of Theorem 3.3. The proof is analogous to the proof of Theorem 2 in AG. The latter proof in turn relies on the proof of Theorem 2 in Hurvich et al. (1998) and the proof of (5.14) in Robinson (1995b). Note that by the definition of $\widehat{d}_{WLP}$

and Lemma A.1(a)

$$\begin{aligned} m^{1/2}(\widehat{d}_{WLP}-d) &= m^{1/2}\sum_{i=1}^{K} w_i J_{m_i}^{-1} X^*(m_i)'(R(m_i)+\epsilon(m_i)) \\ &= m^{1/2}\sum_{i=1}^{K} w_i J_{m_i}^{-1} X^*(m_i)' R(m_i) \\ &\quad + \left(\sum_{i=1}^{K} w_i \frac{1}{4 l_i m^{1/2}} \sum_{j=1}^{m_i} X_{ji}^* \epsilon_j\right)(1+o_p(1)). \end{aligned}$$

It follows from equation (A.4), the fact that the weights satisfy (3.2)-(3.4), and the rate condition $m^{2.5+2r}/n^{2+2r} \to \chi$ that for the first summand we get

$$m^{1/2}\sum_{i=1}^{K} w_i J_{m_i}^{-1} X^*(m_i)' R(m_i) = m^{2.5+2r} n^{-(2+2r)} \tau b_{2+2r}(1+o(1)).$$

Ignoring the $o_p(1)$ term, the second summand can be written as $l_K^{1/2} m_K^{-1/2} \sum_{j=1}^{m_K} a_j \epsilon_j$, where $a_j := \sum_{i=1}^{K}(w_i/(4l_i))1(j \leq m_i) X_{ji}^*$ and $1(\cdot)$ is the indicator function. The same proof as in AG (see p. 705, following equation (7.24)) can be used to show that

$$2(c_r^*)^{-1/2} l_K^{1/2} m_K^{-1/2} \sum_{j=1}^{m_K} a_j \epsilon_j \to_d N(0, \pi^2/6).$$

We just have to verify conditions (i)-(iii) of the asymptotic normality result in AG (7.28) which then yields the desired result. For (i) note that by Lemma 2(k) in AG we have $\max_{1\leq j\leq m_K} |X_{ji}^*| = O(\ln m)$ and thus $\max_{1\leq j \leq m_K} |a_j| = O(\ln m)$. For (ii), $m_i^{-1}\sum_{j=1}^{m_i} X_{ji}^* X_{jk}^* = 4(1+o(1))$ for $l_i < l_k$, implies that $m_K^{-1}\sum_{j=1}^{m_K} a_j^2 = l_K^{-1} 4^{-1} c_r^* (1+o(1))$. Finally, for (iii) note that from AG (7.30) we have $\sum_{j=1+m_i^{0.5+\delta}}^{m_i} |X_{ji}^*|^p = O(m)$ for all $p \geq 1$. This implies $\sum_{j=1+m^{0.5+\delta}}^{m_K} |a_j|^p = O(m)$ for all $p \geq 1$. ■

To prove Theorem 3.4 and establish the asymptotic properties of $\widehat{d}_{WLW}$, we first investigate those of $\widehat{d}_{LW,k}(m)$. For some d^+ between $\widehat{d}_{LW,k-1}(m)$ and d we have

$$\begin{aligned} & \sqrt{m}(\widehat{d}_{LW,k}(m)-d) \\ =\ & \sqrt{m}(\widehat{d}_{LW,k}(m)-\widehat{d}_{LW,k-1}(m)) + \sqrt{m}(\widehat{d}_{LW,k-1}(m)-d) \\ =\ & -\sqrt{m}J_m^{-1} S_m(\widehat{d}_{LW,k-1}(m)) + \sqrt{m}(\widehat{d}_{LW,k-1}(m)-d) \\ =\ & -\sqrt{m}J_m^{-1} S_m(d) - \sqrt{m}J_m^{-1}[S_m(\widehat{d}_{LW,k-1}(m)) - S_m(d)] + \sqrt{m}(\widehat{d}_{LW,k-1}(m)-d) \\ =\ & -\sqrt{m}J_m^{-1} S_m(d) - \sqrt{m}J_m^{-1} H_m(d^+)(\widehat{d}_{LW,k-1}(m)-d) + \sqrt{m}(\widehat{d}_{LW,k-1}(m)-d) \\ =\ & -\sqrt{m}J_m^{-1} S_m(d) - \sqrt{m}(\widehat{d}_{LW,k-1}(m)-d)(J_m^{-1}H_m(d^+)-1). \end{aligned} \tag{A.5}$$

If $\sqrt{m}(\widehat{d}_{LW,k-1}(m)-d) = O_p(1)$, then applying Lemma 2(b) and (d) of AS for the case $r=0$ (in which case their θ and θ_0 are not present) yields $J_m^{-1}H_m(d^+)-1 = o_p(1)$. As a consequence of (A.5),

$$\sqrt{m}(\widehat{d}_{LW,k}(m)-d) = -\sqrt{m}J_m^{-1}S_m(d) + o_p(1). \tag{A.6}$$

To show that $\sqrt{m}(\widehat{d}_{LW,k-1}(m)-d)=O_p(1)$, by an induction argument it suffices to show that $\sqrt{m}(\widehat{d}_{LW,0}(m)-d)=O_p(1)$ and that $-\left(\frac{1}{m}J_m\right)^{-1}\frac{1}{\sqrt{m}}S_m(d)=O_p(1)$. The former is assumed through $\widehat{d}_{LW,0}(m)=\widehat{d}_{WLP}(r,m)$ and the latter is proved now. The term $\left(\frac{1}{m}J_m\right)^{-1}$ is $O_p(1)$ by Lemma A.1(a) and so we are left with the term $\frac{1}{\sqrt{m}}S_m(d)$. Some algebra (see AS (4.2)) gives

$$\begin{aligned}\frac{1}{\sqrt{m}}S_m(d) &= -G\widehat{G}^{-1}(d)\frac{1}{\sqrt{m}}\sum_{j=1}^{m}(\frac{I_j}{\varphi_j}-1)X_j^*(m) \\ &= -G\widehat{G}^{-1}(d)(T_1+T_2+T_3+T_4),\ \text{where}\end{aligned} \tag{A.7}$$

$$\begin{aligned}\varphi_j &:= G\lambda_j^{-2d},\ \widehat{G}(d):=\widehat{G}_0(d,b), \\ T_1 &:= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\frac{I_j}{\varphi_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{\varphi_j}-2\pi I_{\varepsilon j})\right)X_j^*, \\ T_2 &:= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\frac{EI_j}{f_j}-1\right)\frac{f_j}{\varphi_j}X_j^*, \\ T_3 &:= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}(2\pi I_{\varepsilon j}-1)X_j^*,\ T_4:=\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\frac{f_j}{\varphi_j}-1\right)X_j^*\ \text{for} \\ w_\varepsilon(\lambda) &:= (2\pi n)^{-1/2}\sum_{t=1}^{n}\varepsilon_t e^{it\lambda},\ \text{and}\ I_{\varepsilon j}:=|w_\varepsilon(\lambda_j)|^2.\end{aligned} \tag{A.8}$$

The second equality in (A.7) uses the fact that $E2\pi I_{\varepsilon j}=1$. For notational convenience, we write $T_j=T_j(m)$, $j=1,...,4$. In the proof of their Lemma 2(e), AS show that $T_1=T_2=o_p(1)$ as $1/m+m/n\to 0$. The asymptotic properties of T_3, T_4 and $\widehat{G}$ are established in the next lemma. These properties combined with $T_1=T_2=o_p(1)$ imply that $1/\sqrt{m}S_m(d)=O_p(1)$.

Lemma A.2 *Suppose Assumptions AS1-AS4 hold.*

(a) *Let $T_{3i}:=T_3(m_i)$. Then*

$$(T_{31},T_{32},...,T_{3K})'\to_d N(0,4\Sigma),\ \textit{where} \tag{A.9}$$

$$\Sigma:=(\Sigma_{ij})_{K\times K}\ \textit{and}\ \Sigma_{ij}:=\left\{\min(l_i,l_j)/\max(l_i,l_j)\right\}^{1/2}.$$

(b) *For $s=1,...,r+1$ let*

$$A_{2s}:=\frac{(2\pi)^{2s}s}{(2s+1)^2}\sum_{k=1}^{s}\frac{1}{k!}\sum_{i_1+...+i_k=s}(\frac{b_{2i_1}}{(2i_1)!}...\frac{b_{2i_k}}{(2i_k)!}), \tag{A.10}$$

where the i_j $(j=1,...,k)$ are positive integers. Then

$$T_4(m)=-4\sum_{k=1}^{1+r}A_{2k}\frac{m^{0.5+2k}}{n^{2k}}(1+O(\frac{\ln m}{m}))+O(\frac{m^{0.5+q}}{n^q}),$$

where q is defined in (2.7) and $O(\frac{m^{0.5+q}}{n^q})$ can be replaced by $o(\frac{m^{0.5+q}}{n^q})$ if $s = 2 + 2r$.

(c) *For real numbers B_{2j} $(j = 1, ..., r)$ defined in (A.15) below, we have*

$$G\widehat{G}^{-1}(d) = 1 + \sum_{k=1}^{r} B_{2k} \frac{m^{2k}}{n^{2k}} + o_p(\frac{m^{2r}}{n^{2r}}). \tag{A.11}$$

Note that $\widehat{G}$ depends on m.

Proof of Lemma A.2. Part (a) Using the Cramer-Wold device, we need to show that for $\rho = (\rho_1, ..., \rho_K)'$ we have $\sum_{i=1}^{K} \rho_i T_{3i} \to_d N(0, 4\rho\Sigma\rho')$. Note that

$$\begin{aligned}
\sum_{i=1}^{K} \rho_i T_{3i} &= -\sum_{i=1}^{K} \frac{\rho_i}{\sqrt{m_i}} \sum_{p=1}^{m_i} (2\pi I_{\varepsilon p} - 1) \left(2\ln p - m_i^{-1} \sum_{k=1}^{m_i} 2\ln k\right) \\
&= -\sum_{p=1}^{m_K} (2\pi I_{\varepsilon p} - 1) \sum_{i=1}^{K} \frac{\rho_i}{\sqrt{m_i}} \left(2\ln p - m_i^{-1} \sum_{k=1}^{m_i} 2\ln k\right) 1\left(p \le m_i\right) \\
&= -\frac{1}{\sqrt{m_K}} \sum_{p=1}^{m_K} (2\pi I_{\varepsilon p} - 1)\nu_p^*, \text{ where} \\
\nu_p^* &:= \sqrt{m_K} \sum_{i=1}^{K} \frac{\rho_i}{\sqrt{m_i}} \left(2\ln p - m_i^{-1} \sum_{k=1}^{m_i} 2\ln k\right) 1\left(p \le m_i\right).
\end{aligned}$$

Robinson (1995a, p. 1644 lines 11-13 – p. 1647 line 6↑) proves that

$$-\frac{1}{\sqrt{m_K}} \sum_{p=1}^{m_K} (2\pi I_{\varepsilon p} - 1)\nu_p \to_d N(0, \lim_{m_K \to \infty} \frac{4}{m_K} \sum_{p=1}^{m_K} \nu_p^2), \tag{A.12}$$

where $\nu_p := 2\ln p - m_K^{-1} \sum_{k=1}^{m_K} 2\ln k$. Inspection of his proof reveals that (A.12) continues to hold with ν_p replaced by ν_p^* if the following properties of the ν_p^* sequence hold: (i) $\nu_p^* = O(\log m_K)$ and (ii) $\sum_{p=1}^{m_K} \left|\nu_{p+1}^* - \nu_p^*\right| = O\left(\log m_K\right)$. Property (i) holds trivially and (ii) holds because

$$\begin{aligned}
\nu_{p+1}^* - \nu_p^* &= \sqrt{m_K} \sum_{i:m_i \ge p+1} \frac{\rho_i}{\sqrt{m_i}} \left(2\ln\left(p+1\right) - m_i^{-1} \sum_{k=1}^{m_i} 2\ln k\right) \\
&\quad -\sqrt{m_K} \sum_{i:m_i \ge p} \frac{\rho_i}{\sqrt{m_i}} \left(2\ln p - m_i^{-1} \sum_{k=1}^{m_i} 2\ln k\right) \\
&= \sqrt{m_K} \sum_{i:m_i \ge p+1} \frac{\rho_i}{\sqrt{m_i}} \left(2\ln\left(1 + \frac{1}{p}\right)\right) \\
&\quad -\sqrt{m_K} \sum_{i=1}^{K} \frac{\rho_i}{\sqrt{m_i}} \left(2\ln m_i - \frac{1}{m_i} \sum_{k=1}^{m_i} 2\ln k\right) 1\left(p = m_i\right).
\end{aligned}$$

So

$$\begin{aligned}\sum_{p=1}^{m_K}\left|\nu_{p+1}^*-\nu_p^*\right| &= O\left(\sum_{p=1}^{m_K}2\ln\left(1+\frac{1}{p}\right)\right)+O\left(\sum_{i=1}^{K}\left|2\ln m_i-\frac{1}{m_i}\sum_{k=1}^{m_i}2\ln k\right|\right)\\ &= O\left(\log m_K\right).\end{aligned}$$

Therefore Part (a) follows upon showing that $\lim_{m_K\to\infty}\frac{1}{m_K}\sum_{p=1}^{m_K}\left(\nu_p^*\right)^2 = 4\rho\Sigma\rho'$. But $m_K^{-1}\sum_{p=1}^{m_K}\left(\nu_p^*\right)^2$ equals $\sum_{i,j=1}^{K}\nu_{i,j}^*$ for

$$\begin{aligned}\nu_{i,j}^* &:= \frac{1}{m_K}\sum_{p=1}^{m_K}\left\{\sqrt{m_K}\frac{\rho_i}{\sqrt{m_i}}\left(2\ln p-m_i^{-1}\sum_{k=1}^{m_i}2\ln k\right)1\left(p\le m_i\right)\right.\\ &\quad\left.\times\sqrt{m_K}\frac{\rho_j}{\sqrt{m_j}}\left(2\ln p-m_j^{-1}\sum_{k=1}^{m_j}2\ln k\right)1\left(p\le m_j\right)\right\}\\ &= \sum_{p=1}^{\min(m_i,m_j)}\frac{\rho_i}{\sqrt{m_i}}\frac{\rho_j}{\sqrt{m_j}}\left(2\ln p-m_i^{-1}\sum_{k=1}^{m_i}2\ln k\right)\left(2\ln p-m_j^{-1}\sum_{k=1}^{m_j}2\ln k\right)\\ &= \sum_{p=1}^{\min(m_i,m_j)}\frac{\rho_i}{\sqrt{m_i}}\frac{\rho_j}{\sqrt{m_j}}\left(2\ln p-\frac{1}{\min\left(m_i,m_j\right)}\sum_{k=1}^{\min(m_i,m_j)}2\ln k\right)^2\\ &= 4\rho_i\rho_j\frac{\min\left(m_i,m_j\right)}{\sqrt{m_im_j}}(1+o(1)) = 4\rho_i\rho_j\Sigma_{ij}(1+o(1)),\end{aligned}$$

where the second to last equality follows from Lemma 2(a) in AG. Thus, (A.9) follows.

Part (b) Using (A.3), $\ln(f_j/\varphi_j)=R_j$, and applying "exp" on both sides implies that for A_{2k} defined in (A.10), we have

$$\begin{aligned}f_j/\varphi_j &= 1+\sum_{k=1}^{1+r}\frac{(2k+1)^2}{(2\pi)^{2k}\,k}A_{2k}\lambda_j^{2k}+Rem_j^*,\qquad\text{(A.13)}\\ \max_{1\le j\le m}|Rem_j^*/\lambda_j^q| &= O(1)\text{ and } =o(1)\text{ if }s=2+2r.\end{aligned}$$

Therefore,

$$\begin{aligned}T_4 &= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\sum_{k=1}^{1+r}\frac{(2k+1)^2}{(2\pi)^{2k}\,k}A_{2k}\lambda_j^{2k}+Rem_j^*\right)X_j^*\\ &= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\sum_{k=1}^{1+r}\frac{(2k+1)^2}{(2\pi)^{2k}\,k}A_{2k}\lambda_j^{2k}\right)X_j^*+O(m^{0.5+q}/n^q),\end{aligned}$$

where for the second equation we used equation (7.16) in AG and the O can be replaced by o if $s=2+2r$. Using Lemma A.1(b), we thus get the desired result:

$$\begin{aligned}T_4 &= \frac{1}{\sqrt{m}}\sum_{k=1}^{1+r}\frac{(2k+1)^2}{(2\pi)^{2k}\,k}A_{2k}\sum_{j=1}^{m}X_j^*\lambda_j^{2k}+O(m^{0.5+q}/n^q)\\ &= \frac{-4}{\sqrt{m}}\sum_{k=1}^{1+r}\left[A_{2k}\frac{m^{2k+1}}{n^{2k}}(1+O(\frac{\ln m}{m}))\right]+O(m^{0.5+q}/n^q).\end{aligned}$$

Part (c) We show that for certain constants $C_{2k} \in \mathbb{R}$ $(k = 1, ..., r)$ defined in (A.17), we have

$$G^{-1}\widehat{G}(d) = 1 + \sum_{k=1}^{r} C_{2k} \frac{m^{2k}}{n^{2k}} + o_p(\frac{m^{2r}}{n^{2r}}). \tag{A.14}$$

It is tedious but straightforward to show that this implies the desired result (A.11) for

$$(B_2, ..., B_{2r})' := -M^{-1}(C_2, ..., C_{2r})', \tag{A.15}$$

where $M = (M_{i,j})$ is a lower triangular $r \times r$-matrix with the number 1 on the main diagonal and $M_{i,j} := C_{2(i-j)}$ for $i > j$. To prove (A.14), write

$$G^{-1}\widehat{G}(d) - 1 = m^{-1} \sum_{j=1}^{m} \left(\frac{I_j}{\varphi_j} - 1 \right) = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4, \tag{A.16}$$

where

$$\mathcal{T}_1 := \frac{1}{m} \sum_{j=1}^{m} \left(\frac{I_j}{\varphi_j} - 2\pi I_{\varepsilon j} - E(\frac{I_j}{\varphi_j} - 2\pi I_{\varepsilon j}) \right), \; \mathcal{T}_2 := \frac{1}{m} \sum_{j=1}^{m} \left(\frac{EI_j}{f_j} - 1 \right) \frac{f_j}{\varphi_j},$$
$$\mathcal{T}_3 := \frac{1}{m} \sum_{j=1}^{m} (2\pi I_{\varepsilon j} - 1) \text{ and } \mathcal{T}_4 := \frac{1}{m} \sum_{j=1}^{m} \left(\frac{f_j}{\varphi_j} - 1 \right).$$

We claim that $\mathcal{T}_i = o_p(m^{2r}n^{-2r})$ for $i = 1, 2$, and 3. In fact, equation (A.21) of AS (with their g_j playing the role of our φ_j, their $\phi = 2$, and $k = m$) states $\mathcal{T}_1 = O_p(m^{-2/3} \ln^{2/3} m + m^{3/2}n^{-2} + m^{-1/2}n^{-1/4}) = o_p(m^{2r}n^{-2r})$, where for the last equality we use Assumption AS4. Equation (A.23) of AS states $E(I_j/f_j) = 1 + O(j^{-1} \ln j)$ uniformly over $j = 1, ..., m$. Therefore $\mathcal{T}_2 = O(m^{-1} \sum_{j=1}^{m} j^{-1} \ln j) = O(m^{-1} \ln^2 m) = o(m^{2r}n^{-2r})$. Equation (A.13) of AS states $\mathcal{T}_3 = O_p(m^{-1/2}) = o_p(m^{2r}n^{-2r})$. Finally (A.13) implies that

$$\mathcal{T}_4 = \frac{1}{m} \sum_{j=1}^{m} \left(\sum_{k=1}^{1+r} \frac{(2k+1)^2}{(2\pi)^{2k} k} A_{2k} \lambda_j^{2k} + Rem_j^* \right) = \sum_{k=1}^{r} \frac{2k+1}{k} A_{2k} \frac{m^{2k}}{n^{2k}} + o(\frac{m^{2r}}{n^{2r}}),$$

where the second equality uses (iv) in the proof of Lemma A.1. This proves (A.14) with

$$C_{2k} := \frac{2k+1}{k} A_{2k} \text{ for } k = 1, ..., r. \blacksquare \tag{A.17}$$

Proof of Theorem 3.4. It follows from equation (A.6), $\sum_{i=1}^{K} w_i = 1$ and Lemma A.1(a) that

$$\begin{aligned} \sqrt{m} \left(\widehat{d}_{WLW} - d \right) &= \sum_{i=1}^{K} \frac{w_i}{\sqrt{l_i}} \sqrt{m_i} \left(\widehat{d}_{LW,k}(m_i) - d \right) \\ &= -\sum_{i=1}^{K} \frac{w_i}{4\sqrt{l_i}} (1 + O(\frac{\ln^2 m_i}{m_i})) \frac{S_{m_i}(d)}{\sqrt{m_i}} + o_p(1). \end{aligned} \tag{A.18}$$

Using (A.7), $T_1 = T_2 = o_p(1)$, and Lemma A.2(b) and (c), it follows that the $O((\ln^2 m_i)/m_i)$ term in (A.18) can be ignored[9] and thus

$$\sqrt{m}\left(\widehat{d}_{WLW} - d\right) = \sum_{i=1}^{K} \frac{w_i}{4\sqrt{l_i}} G\widehat{G}^{-1}(d)\left(T_3(m_i) + T_4(m_i)\right) + o_p(1). \tag{A.19}$$

Lemma A.2(a) implies that $\sum_{i=1}^{K} w_i/(4\sqrt{l_i})T_3(m_i) \to_d N(0, c_r^*/4)$. Using Lemma A.2(b), (c), (3.3), (3.4), and AS4, we have, up to $o_p(1)$ terms, for $B_0 := 1$

$$\sum_{i=1}^{K} \frac{w_i}{4\sqrt{l_i}} G\widehat{G}^{-1}(d) T_4(m_i) = -\sum_{i=1}^{K} \frac{w_i}{\sqrt{l_i}}(1 + \sum_{k=1}^{r} B_{2k}\frac{m_i^{2k}}{n^{2k}})(\sum_{k=1}^{1+r} A_{2k}\frac{m_i^{0.5+2k}}{n^{2k}}) \tag{A.20}$$

$$= -\sum_{i=1}^{K} \frac{w_i}{\sqrt{l_i}}(\sum_{k=1}^{1+r} B_{2(k-1)}A_{2+2r-2(k-1)})\frac{m_i^{2.5+2r}}{n^{2r+2}} \to -(\sum_{k=1}^{1+r} B_{2(k-1)}A_{2+2r-2(k-1)})\chi\tau/\tau_r^*.$$

Setting

$$D_r := -\sum_{k=1}^{1+r} B_{2(k-1)}A_{2+2r-2(k-1)} \tag{A.21}$$

and combining the above statements gives

$$\sqrt{m}(\widehat{d}_{WLW} - d) \to_d N(\chi(\tau/\tau_r^*)D_r, \frac{1}{4}c_r^*),$$

i.e.

$$\sqrt{m}(\widehat{d}_{WLW} - d - (\tau/\tau_r^*)D_r m^{2+2r} n^{-(2+2r)}) \to_d N(0, \frac{1}{4}c_r^*). \blacksquare$$

To prove Theorem 4.3 we need the following lemma to be proved below. Unless otherwise stated, from now on C is a generic constant, that may be different across different lines.

Lemma A.3 *Suppose the assumptions of Theorem 4.3 hold.*

(a) For any constant $C > 0$, *we have*

$$\sup_{\tau \in [s_*, \min(s_0, s^*)]} P\left(m_\tau^{1/2}|\widehat{d}_{WLW}(\tau) - d| > C\zeta(n)\right) = O\left(\zeta^{-2}(n)\right).$$

(b) If $s_0 < \infty$, *then for any constant* $C > 0$,

$$P\left(m_{s_0}^{1/2}|\widehat{d}_{WLW}(s_0 + 8s_0 h \ln\ln n) - d| \le C\zeta(n)\right) = O\left(\zeta^{-2}(n)\right).$$

Proof of Theorem 4.3. Part (a) Set $\overline{s} := \min(s_0, s^*)$ and recall that $h = 1/\ln n$. We first bound $P\left(\widehat{s} < \overline{s} - h\right)$. We have $P\left(\widehat{s} < \overline{s} - h\right) = \sum_{\tau \in \mathcal{S}_h : \tau + h < \overline{s}} P(\widehat{s} = \tau)$. By the definition of $\widehat{s}$, if $\widehat{s} = \tau$, there exists $\tau' \le \tau$, $\tau' \in \mathcal{S}_h$ such that

$$|\widehat{d}_{WLW}(\tau + h) - \widehat{d}_{WLW}(\tau')| > m_{\tau'}^{-1/2}\psi_2(c_{r(\tau')}^*/4)^{1/2}\zeta(n).$$

In consequence, for all $\tau \in \mathcal{S}_h$ with $\tau + h < \overline{s}$,

$$
\begin{aligned}
& P(\widehat{s} = \tau) \\
\leq & \sum_{\tau' \in \mathcal{S}_h : \tau' \leq \tau} P\left(|\widehat{d}_{WLW}(\tau + h) - \widehat{d}_{WLW}(\tau')| > m_{\tau'}^{-1/2} \psi_2 (c^*_{r(\tau')}/4)^{1/2} \zeta(n) \right) \\
\leq & \sum_{\tau' \in \mathcal{S}_h : \tau' \leq \tau} P\left(m_{\tau+h}^{1/2} |\widehat{d}_{WLW}(\tau + h) - d| > \kappa \zeta(n) \right) \\
& + \sum_{\tau' \in \mathcal{S}_h : \tau' \leq \tau} P\left(m_{\tau'}^{1/2} |\widehat{d}_{WLW}(\tau') - d| > \kappa \zeta(n) \right), \\
\leq & \; 2(s^* - s_*)(\ln n) \sup_{\tau'' \in \mathcal{S}_h : \tau'' \leq \overline{s}} P\left(m_{\tau''}^{1/2} |\widehat{d}_{WLW}(\tau'') - d| > \kappa \zeta(n) \right),
\end{aligned}
$$

where $\kappa := \psi_2 (c^*_{r(s_*)}/4)^{1/2}/2$. Note that we used that c^*_r is an increasing function in r. The last inequality holds because there are at most $(s^* - s_*)(\ln n)$ elements $\tau' \in \mathcal{S}_h$ for which $\tau' \leq \tau$. It now follows from Lemma A.3(a) that

$$
P(\widehat{s} < \overline{s} - h) = O((s^* - s_*)^2 (\ln n)^2 \zeta^{-2}(n)) = O\left((\ln \ln n)^{-1} \right) = o(1) \text{ as } n \to \infty. \tag{A.22}
$$

If $s_0 = \infty$ (and thus $s_0 \geq s^*$), then (A.22) clearly implies the result in Part (a). Therefore from now on we can assume $s_0 < s^*$. We now prove that $P(\widehat{s} > s_0 + 8 s_0 h \ln \ln n) = o(1)$. Assume w.l.o.g. that $s_0 + 8 s_0 h \ln \ln n \in \mathcal{S}_h$. By the definition of $\widehat{s}$,

$$
\begin{aligned}
& P(\widehat{s} > s_0 + 8 s_0 h \ln \ln n) \\
\leq & \; P\left(m_{s_0}^{1/2} |\widehat{d}_{WLW}(s_0 + 8 s_0 h \ln \ln n) - \widehat{d}_{WLW}(s_0)| \leq \psi_2 \left(c^*_{r(s_0)}/4 \right)^{1/2} \zeta(n) \right) \\
\leq & \; P\left\{ m_{s_0}^{1/2} |\widehat{d}_{WLW}(s_0 + 8 s_0 h \ln \ln n) - d| \leq \psi_2 \left(c^*_{r(s_0)}/4 \right)^{1/2} \zeta(n) \right. \\
& \left. + m_{s_0}^{1/2} |\widehat{d}_{WLW}(s_0) - d| \right\} \\
\leq & \; P\left(m_{s_0}^{1/2} |\widehat{d}_{WLW}(s_0 + 8 s_0 h \ln \ln n) - d| \leq \psi_2 \left(c^*_{r(s_0)}/4 \right)^{1/2} \zeta(n) + \zeta(n) \right) \\
& + P\left(m_{s_0}^{1/2} |\widehat{d}_{WLW}(s_0) - d| \geq \zeta(n) \right) \\
= & \; O\left(\zeta^{-2}(n) \right) = o(1),
\end{aligned} \tag{A.23}
$$

where the last line uses both parts of Lemma A.3. In the above proof, we implicitly assume that $s_0 \in \mathcal{S}_h$. If this is not the case, we can bound $P(\widehat{s} > s_0 + 8 s_0 h \ln \ln n)$ by

$$
P\left(m_{s_0^*}^{1/2} |\widehat{d}_{WLW}(s_0 + 8 s_0 h \ln \ln n) - \widehat{d}_{WLW}(s_0^*)| \leq \psi_2 \left(c^*_{r(s_0^*)}/4 \right)^{1/2} \zeta(n) \right)
$$

where $s_0^* := \max\{s : s \in \mathcal{S}_h, s \leq s_0\}$. The rest of the proof goes through with obvious changes.

Combining (A.22) and (A.23), we get $\widehat{s} = s_0 + O_p(\ln\ln n/\ln n)$ as desired, completing the proof of Theorem 4.3(a).

Part (b) Let $\tilde{s}_0 := \max\{s \in \mathcal{S}_h : s \leq s_0 - h\}$, then $|\tilde{s}_0 - s_0| \leq 2h$. Using the definition of $\widehat{s}$ and Lemma A.3(a), we have

$$
\begin{aligned}
& P\left(n^{\frac{s_0}{2s_0+1}}\zeta^{-1}(n)\left|\widehat{d}_{WLW}(\widehat{s}) - d\right| > C\right) \\
\leq\; & P\left(n^{\frac{s_0}{2s_0+1}}\zeta^{-1}(n)\left|\widehat{d}_{WLW}(\widehat{s}) - d\right| > C, \widehat{s} \geq s_0 - h\right) + P\left(\widehat{s} < s_0 - h\right) \\
\leq\; & P\left(n^{\frac{s_0}{2s_0+1}}\zeta^{-1}(n)\left|\widehat{d}_{WLW}(\widehat{s}) - \widehat{d}_{WLW}(\tilde{s}_0)\right| > \frac{C}{2}, \widehat{s} \geq \tilde{s}_0\right) \\
& + P\left(n^{\frac{s_0}{2s_0+1}}\zeta^{-1}(n)\left|\widehat{d}_{WLW}(\tilde{s}_0) - d\right| > \frac{C}{2}, \widehat{s} \geq \tilde{s}_0\right) + o(1) \\
\leq\; & P\left(n^{\frac{s_0}{2s_0+1} - \frac{\tilde{s}_0}{2\tilde{s}_0+1}}\psi_2(c^*_{r(\tilde{s}_0)}/4)^{1/2} > \frac{C}{2}\right) \\
& + P\left(n^{\frac{s_0}{2s_0+1} - \frac{\tilde{s}_0}{2\tilde{s}_0+1}}\zeta^{-1}(n) m^{1/2}(\tilde{s}_0)\left|\widehat{d}_{WLW}(\tilde{s}_0) - d\right| > \frac{C}{2}\right) = o(1)
\end{aligned}
$$

uniformly as $C \to \infty$, where the last equality holds because

$$
\lim_{n\to\infty} n^{\frac{s_0}{2s_0+1} - \frac{\tilde{s}_0}{2\tilde{s}_0+1}} \leq \lim_{n\to\infty} n^{\frac{2s_0 h}{2s_0+1}} \leq \lim_{n\to\infty} n^h = e. \blacksquare
$$

To prove Lemma A.3, we introduce some notation and state one more lemma. Let $S_{i,\tau}(d)$, $H_{i,\tau}(d)$, $J_{i,\tau}, T_{k,i\tau}, \mathcal{T}_{k,i\tau}$, and $\widehat{G}_{i,\tau}$ be defined as the corresponding quantities $S_m(d)$, $H_m(d)$, $J_m, T_{ki}, \mathcal{T}_k$, and $\widehat{G}(d)$ (from (A.8)) with m replaced by $m_{i,\tau} = l_i m_\tau$. Unless otherwise stated, from now on "uniformly" stands for "uniformly over $\tau \in [s_*, \min(s_0, s^*)]$".

Lemma A.4 *Set* $D_{m_{i,\tau}} = \left\{d^+ : |d^+ - d| \leq \zeta^{-2}(n)\ln^{-5} m_{i,\tau}\right\}$, $\mathcal{C}_{i,\tau} = \sum_{k=1}^{r(\tau)} C_{2k} m_{i,\tau}^{2k}/n^{2k}$. *Then, under Assumptions AS1, AS2(ii), AS3, and for* $a = 1, 2,$ *and* 3 *we have uniformly*

$$
\begin{aligned}
& (a) P\left(\sup\nolimits_{d^+\in D_{m_{i,\tau}}} m_{i,\tau}^{-1}\left|H_{i,\tau}(d^+) - J_{i,\tau}\right| > C\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right), \\
& (b) P\left(\left|G^{-1}\widehat{G}_{i,\tau}(d) - 1 - \mathcal{C}_{i,\tau}\right| > C\left(m_{i,\tau}/n\right)^{2r(\tau)}\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right), \\
& (c) P\left(|T_{a,i\tau}| > C\zeta(n)\right) = O(\zeta^{-2}(n)).
\end{aligned}
\qquad (A.24)
$$

Proof of Lemma A.4. Part (a) Inspection of the proof of Lemma 7 of AS (p. 609) shows that if $m_{i,\tau} = O(n^{\Delta})$ for some $\Delta < 1$, then uniformly

$$
P\left(\sup_{d^+\in D_{m_{i,\tau}}} m_{i,\tau}^{-1}\left|H_{i,\tau}(d^+) - J_{i,\tau}\right| > C\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right). \qquad (A.25)
$$

Part (a) holds because $m_{i,\tau} = O(n^{2\tau/(2\tau+1)})$ satisfies the required rate condition.

Part (b) Recall from (A.16) that

$$G^{-1}\widehat{G}_{i,\tau}(d)-1=m_{i,\tau}^{-1}\sum_{j=1}^{m_{i,\tau}}\left(\frac{I_j}{\varphi_j}-1\right)=\mathcal{T}_{1,i\tau}+\mathcal{T}_{2,i\tau}+\mathcal{T}_{3,i\tau}+\mathcal{T}_{4,i\tau}.$$

To prove Part (b), it suffices to prove that uniformly for $a=1,2,3$

$$P\left(|\mathcal{T}_{a,i\tau}|>C\,(m_{i,\tau}/n)^{2r(\tau)}\,\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right) \text{ and} \tag{A.26}$$

$$P\left(|\mathcal{T}_{4,i\tau}-\mathcal{C}_{i,\tau}|>C\,(m_{i,\tau}/n)^{2r(\tau)}\,\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right). \tag{A.27}$$

The proofs of (A.26) and (A.27) use the following result: If a sequence $h_j(\tau)=h_j(m_\tau)$, $j=1,...,m_\tau$ uniformly satisfies $m_\tau^{-1}\sum_{J=1}^{m_\tau}h_j^2(\tau)=O(1)$ then

$$E\left(\sum_{j=1}^{m_\tau}(2\pi I_{\varepsilon j}-1)h_j(\tau)\right)^2=O(m_\tau)\text{ uniformly.} \tag{A.28}$$

To prove (A.28), note that by Theorem 5.2.4 of Brillinger (1975, p. 125) and AS (p. 604), $\mathrm{Var}(I_{\varepsilon j})=O(1)$ and $\mathrm{Cov}(I_{\varepsilon i},I_{\varepsilon j})=O(n^{-1})$ uniformly over $i,j=1,...,n$ with $i\neq j$. Thus, as desired, (A.28) equals

$$(2\pi)^2\sum_{j=1}^{m_\tau}\mathrm{Var}(I_{\varepsilon j})h_j^2(\tau)+8\pi^2\sum_{j,k=1,j\neq k}^{m_\tau}\mathrm{Cov}(I_{\varepsilon j},I_{\varepsilon k})h_j(\tau)h_k(\tau)=O(m_\tau).$$

We now proceed to establish (A.26). For $a=1$, we use a proof similar to AS' proof of their (A.21) on p. 603. Recall that

$$\mathcal{T}_{1,i\tau}=m_{i,\tau}^{-1}\sum_{j=1}^{m_{i,\tau}}[I_j/\varphi_j-2\pi I_{\varepsilon j}-E(I_j/\varphi_j-2\pi I_{\varepsilon j})].$$

Let $\ell:=(m_{i,\tau})^{1/3}\,(\ln m_{i,\tau})^{2/3}$ and

$$A_{10}:=m_{i,\tau}^{-1}\sum_{j=1}^{\ell}\left(\frac{I_j}{\varphi_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{\varphi_j}-2\pi I_{\varepsilon j})\right).$$

Using $E\left|I_j\varphi_j^{-1}\right|\leq C$, for $j=1,...,\ell$, proved in Robinson (1995a, (3.16)), we have

$$E\,|A_{10}|=O\left(m_{i,\tau}^{-2/3}\,(\ln m_{i,\tau})^{2/3}\right)\text{ uniformly.} \tag{A.29}$$

We decompose the remaining summands in $\mathcal{T}_{1,i\tau}$ into

$$\begin{aligned} & m_{i,\tau}^{-1}\sum_{j=\ell+1}^{m_{i,\tau}}\left(\frac{I_j}{\varphi_j}-2\pi I_{\varepsilon j}-E(\frac{I_j}{\varphi_j}-2\pi I_{\varepsilon j})\right) \\ = \; & m_{i,\tau}^{-1}\sum_{j=\ell+1}^{m_{i,\tau}}\left((\frac{I_j}{f_j}-2\pi I_{\varepsilon j})\frac{f_j}{\varphi_j}-E(\frac{I_j}{f_j}-2\pi I_{\varepsilon j})\frac{f_j}{\varphi_j}\right) \\ & +2\pi m_{i,\tau}^{-1}\sum_{j=\ell+1}^{m_{i,\tau}}(I_{\varepsilon j}-EI_{\varepsilon j})\left(\frac{f_j}{\varphi_j}-1\right):=A_{11}+A_{12}. \end{aligned} \tag{A.30}$$

We can uniformly bound EA_{11}^2 by

$$m_{i,\tau}^{-2}E(\sum_{j=\ell+1}^{m_{i,\tau}}(\frac{I_j}{f_j}-2\pi I_{\varepsilon j})\frac{f_j}{\varphi_j})^2=O(m_{i,\tau}^{-4/3}\ln^{4/3}m_{i,\tau}+m_{i,\tau}^{-1}n^{-1/2}), \tag{A.31}$$

where the equality holds by the same proof as in Robinson (1995a, p. 1648–51) for the quantity given in the third equation on his p. 1648. The only difference is that the factor f_j/φ_j does not appear in Robinson (1995a). It can be shown that this factor has no impact on the proof because $f_j/\varphi_j=1+o(1)$ uniformly over $j=1,...,.m_{i,\tau}$ For A_{12}, we have,

$$\begin{aligned} EA_{12}^2 = \; & m_{i,\tau}^{-2}4\pi^2\sum_{j=\ell+1}^{m_{i,\tau}}\mathrm{Var}(I_{\varepsilon j})\left(\frac{f_j}{\varphi_j}-1\right)^2 \\ & +m_{i,\tau}^{-2}8\pi^2\sum_{i=\ell+1}^{m_{i,\tau}}\sum_{j=\ell+1}^{i-1}\mathrm{Cov}(I_{\varepsilon i},I_{\varepsilon j})\left(\frac{f_i}{\varphi_i}-1\right)\left(\frac{f_j}{\varphi_j}-1\right) \\ = \; & O(1)m_{i,\tau}^{-2}\sum_{j=\ell+1}^{m_{i,\tau}}\lambda_j^{2\phi}+O(n^{-1})m_{i,\tau}^{-2}\sum_{i=\ell+1}^{m_{i,\tau}}\sum_{j=\ell+1}^{i-1}\lambda_i^{\phi}\lambda_j^{\phi} \\ = \; & O\left(m_{i,\tau}^{-1}(\frac{m_{i,\tau}}{n})^{2\phi}\right)+O\left(\frac{1}{n}(\frac{m_{i,\tau}}{n})^{2\phi}\right)=O\left(m_{i,\tau}^{-1}(\frac{m_{i,\tau}}{n})^{2\phi}\right), \end{aligned} \tag{A.32}$$

uniformly, where $\phi=\min(\tau,2)$ and we have used Theorem 5.2.4 of Brillinger (1975, p. 125) again. Combining (A.29)–(A.32) and using $2r(\tau)<\tau$ gives

$$\begin{aligned} & P\left(|\mathcal{T}_{1,i\tau}|>C\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\zeta^{-1}(n)\right) \\ = \; & P\left(|A_{10}|+|A_{11}|+|A_{12}|>C\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\zeta^{-1}(n)\right) \\ = \; & O\left(\zeta^2(n)\left(\frac{m_{i,\tau}}{n}\right)^{-4r(\tau)}\left[m_i^{-4/3}(\tau)\left(\ln m_{i,\tau}\right)^{4/3}+m_{i,\tau}^{-1}n^{-1/2})+m_{i,\tau}^{-1}(\frac{m_{i,\tau}}{n})^{2\phi}\right]\right) \\ & +O\left(\left(\frac{m_{i,\tau}}{n}\right)^{-2r(\tau)}\zeta(n)m_i^{-2/3}(\tau)\left(\ln m_{i,\tau}\right)^{2/3}\right) \end{aligned}$$

$$
\begin{aligned}
&= O\left\{\zeta^2(n)\left[n^{4r(\tau)}m_{i,\tau}^{-4r(\tau)-4/3}\left(\ln m_{i,\tau}\right)^{4/3}+n^{4r(\tau)-0.5}m_{i,\tau}^{-4r(\tau)-1}\right]\right\} \\
&\quad +O\left\{\zeta^2(n)n^{4r(\tau)-2\phi}m_{i,\tau}^{-4r(\tau)-1+2\phi}+\zeta(n)m_{i,\tau}^{-2r(\tau)-2/3}n^{2r(\tau)}\left(\ln m_{i,\tau}\right)^{2/3}\right\} \\
&= O\left[\zeta^2(n)\left(n^{\frac{4r(\tau)-8/3\tau}{2\tau+1}}\left(\ln m_{i,\tau}\right)^{4/3}+n^{\frac{4r(\tau)-3\tau-0.5}{(2\tau+1)}}+n^{\frac{4r(\tau)-2\tau-2\phi}{2\tau+1}}\right)\right] \\
&\quad +O\left[\zeta(n)n^{\frac{2r(\tau)-4/3\tau}{(2\tau+1)}}\left(\ln m_{i,\tau}\right)^{2/3}\right]=O\left(\zeta^{-2}(n)\right).
\end{aligned}
$$

For $a=2$, (A.26) holds because uniformly $\mathcal{T}_{2,i\tau}=O\left(\frac{\ln^2 m_{i,\tau}}{m_{i,\tau}}\right)=o\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\zeta^{-1}(n)$. For $a=3$, by Markov's inequality

$$
\begin{aligned}
&P\left(\left|\mathcal{T}_{3,i\tau}\right|>C\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\zeta^{-1}(n)\right) \\
&= P\left(\left|m_{i,\tau}^{-1}\sum_{j=1}^{m_{i,\tau}}\left(2\pi I_{\varepsilon j}-1\right)\right|>C\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\zeta^{-1}(n)\right) \\
&\leq C^{-2}\zeta^2(n)\left(\frac{m_{i,\tau}}{n}\right)^{-4r(\tau)}E\left(m_{i,\tau}^{-1}\sum_{j=1}^{m_{i,\tau}}\left(2\pi I_{\varepsilon j}-1\right)\right)^2 \\
&\leq C^{-2}\zeta^2(n)n^{4r(\tau)}/\left(m_{i,\tau}\right)^{4r(\tau)+1}=O\left(\zeta^2(n)n^{4r(\tau)-2\tau}\right)=O\left(\zeta^{-2}(n)\right),
\end{aligned}
$$

uniformly, where the second inequality follows from (A.28) by setting $h_j(\tau)=1$ and the last equality holds because $4r(\tau)-2\tau<0$. This proves (A.26). To prove (A.27), we only need to show that $\left|\mathcal{T}_{4,i\tau}-\mathcal{C}_{i,\tau}\right|=o((m_{i,\tau}/n)^{2r(\tau)}\zeta^{-1}(n))$, because $\mathcal{T}_{4,i\tau}$ is nonrandom. In the proof of Part (c) of Lemma A.2, we have shown that $\mathcal{T}_{4,i\tau}=\mathcal{C}_{i,\tau}+o\left(m_{i,\tau}^{2r(\tau)}/n^{2r(\tau)}\right)$ and it is easy to see that the $o\left(m_{i,\tau}^{2r(\tau)}/n^{2r(\tau)}\right)$ term is actually $O((m_{i,\tau})^{\tau}/n^{\tau})$. In view of $2r(\tau)<\tau$, we get $\left|\mathcal{T}_{4,i\tau}-\mathcal{C}_{i,\tau}\right|=o\left((m_{i,\tau}/n)^{2r(\tau)}\zeta^{-1}(n)\right)$ as required.

Part (c) The result holds for $a=2$ because

$$
\begin{aligned}
T_{2,i\tau} &:= m_{i,\tau}^{-1/2}\sum_{j=1}^{m_{i,\tau}}\left(\frac{EI_j}{f_j}-1\right)\frac{f_j}{\varphi_j}X_j^*=m_{i,\tau}^{-1/2}\sum_{j=1}^{m_{i,\tau}}O\left(\frac{\ln j}{j}\right)O(1)O(X_j^*) \\
&= O\left(m_{i,\tau}^{-1/2}\ln m_{i,\tau}\sum_{j=1}^{m_{i,\tau}}\frac{\ln j}{j}\right)=O\left(m_{i,\tau}^{-1/2}\ln^3 m_{i,\tau}\right)=o(\zeta(n))
\end{aligned}
$$

uniformly, where we have used $EI_j/f_j-1=O(j^{-1}\ln j)$ and $X_j^*=O(\ln m_{i,\tau})$ uniformly over $j=1,...,m_{i,\tau}$.

Part (c) holds for $a = 3$ because $ET^2_{3,i\tau} = O(1)$ uniformly, using (A.28) with $h_j = X^*_j$ and then

$$P\left(|T_{3,i\tau}| > C\zeta(n)\right) = C^{-2}\zeta^{-2}(n)ET^2_{3,i\tau} = O(\zeta^{-2}(n)).$$

Finally, for $a = 1$, we can prove Part (c) using the proof similar to that of (A.26) for $a = 1$. Details are omitted. ■

Proof of Lemma A.3. We prove only the case for the 1-step estimator. The extension to the general k-step estimator is straightforward. Recall $m_{i,\tau} = l_i m_\tau$.

Part (a) In the proof of Theorem 3.4, we have shown that, for some $d^+_{i,\tau}$ between $\widehat{d}_{WLP}(\tau)$ and d,

$$\begin{aligned} m_\tau^{1/2}\left(\widehat{d}_{WLW}(\tau) - d\right) &= \sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} m_{i,\tau}^{1/2}(\widehat{d}_{LW,1}(m_{i,\tau}) - d) \\ = & -\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} m_{i,\tau}^{1/2} J_{i,\tau}^{-1} S_{i,\tau}(d) + e_{m,\tau}, \end{aligned} \tag{A.33}$$

where $\{w_{i,\tau}\}_{i=1}^{K}$ are the weights when $r = r(\tau)$,

$$e_{m,\tau} = -\left(m_\tau^{1/2}(\widehat{d}_{WLP}(\tau) - d)\right)\left(\sum_{i=1}^{K} w_{i,\tau}(J_{i,\tau}^{-1} H_{i,\tau}(d^+_{i,\tau}) - 1)\right), \tag{A.34}$$

and $m_{i,\tau}^{1/2}(d^+_{i,\tau} - d) = O_p(1)$.

We now consider each of the two terms in (A.33). It follows from Lemma A.1(a) and Lemma A.4(a) that uniformly

$$P\left(\left|J_{i,\tau}^{-1} H_{i,\tau}(d^+_{i,\tau}) - 1\right| > C\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right)$$

and thus

$$P\left(\sum_{i=1}^{K} |w_{i,\tau}| \left|J_{i,\tau}^{-1} H_{i,\tau}(d^+_{i,\tau}) - 1\right| > C\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right)$$

uniformly. Using the preceding inequality, we have uniformly

$$\begin{aligned} & P\left(|e_{m,\tau}| > C\zeta(n)\right) \\ \leq\ & P\left(m_\tau^{1/2}\left|\widehat{d}_{WLP}(\tau) - d\right| > C\zeta^2(n)\right) + O\left(\zeta^{-2}(n)\right) \\ \leq\ & P\left(m_\tau^{1/2}\left|\widehat{d}_{WLP}(\tau) - d\right| > C\zeta^2(n)\right) + O\left(\zeta^{-2}(n)\right) \\ \leq\ & \frac{m_\tau E\left(\widehat{d}_{WLP}(\tau) - d\right)^2}{C^2\zeta^4(n)} + O\left(\zeta^{-2}(n)\right) = O\left(\zeta^{-2}(n)\right), \end{aligned} \tag{A.35}$$

where the last equality follows from Theorem 3.2. The uniformity holds because the MSE $m_\tau E(\widehat{d}_{WLP}(\tau)-d)^2$ is uniformly bounded for τ in a finite interval.

Next, we show that for some $C>0$,

$$P\left(\left|\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} m_{i,\tau}^{1/2} J_{i,\tau}^{-1} S_{i,\tau}(d)\right| > C\zeta(n)\right) = O\left(\zeta^{-2}(n)\right) \text{ uniformly.} \tag{A.36}$$

As in the proof of Lemma A.2, we write

$$\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} m_{i,\tau}^{1/2} J_{i,\tau}^{-1} S_{i,\tau}(d) = -\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} G\widehat{G}_{i,\tau}^{-1}(d)\left(m_{i,\tau}^{-1} J_{i,\tau}\right)^{-1} \sum_{a=1}^{4} T_{a,i\tau}. \tag{A.37}$$

Let $\mathcal{E}_i$ denote the event $\{|G^{-1}\widehat{G}_{i,\tau}(d)-1-\mathcal{C}_{i,\tau}| < C\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\zeta^{-1}(n)$ and $|m_{i,\tau}J_{i,\tau}^{-1} - 1/4| < Cm_{i,\tau}^{-1}\ln^2 m_{i,\tau}\}$. It follows from Lemma A.1(a) and Lemma A.4(b) that $P\left(\left(\cap_{i=1}^{K}\mathcal{E}_i\right)^c\right) = O\left(\zeta^{-2}(n)\right)$ uniformly if C is large enough, where $\left(\cap_{i=1}^{K}\mathcal{E}_i\right)^c$ is the complement of $\cap_{i=1}^{K}\mathcal{E}_i$. We can therefore focus on the event $\cap_{i=1}^{K}\mathcal{E}_i$.

$$\begin{aligned}
& P\left(\left|\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} m_{i,\tau}^{1/2} J_{i,\tau}^{-1} S_{i,\tau}(d)\right| > C\zeta(n), \cap_{i=1}^{K}\mathcal{E}_i\right) \\
\leq\ & P\left(\left|\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} G\widehat{G}_{i,\tau}^{-1}(d)\left(m_{i,\tau}^{-1}J_{i,\tau}\right)^{-1} T_{4,i\tau}\right| > C\zeta(n)/4, \cap_{i=1}^{K}\mathcal{E}_i\right) \\
& + \sum_{a=1}^{3} P\left(\sum_{i=1}^{K} |w_{i,\tau} l_i^{-1/2} G\widehat{G}_{i,\tau}^{-1}(d)\left(m_{i,\tau}^{-1}J_{i,\tau}\right)^{-1}|\,|T_{a,i\tau}| > C\zeta(n)/4, \cap_{i=1}^{K}\mathcal{E}_i\right) \\
\leq\ & P\left(\left|\sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} G\widehat{G}_{i,\tau}^{-1}(d)\left(m_{i,\tau}^{-1}J_{i,\tau}\right)^{-1} T_{4,i\tau}\right| > C\zeta(n)/4, \cap_{i=1}^{K}\mathcal{E}_i\right) \qquad \text{(A.38)} \\
& + \sum_{a=1}^{3} P\left(|T_{a,i\tau}| > C\zeta(n)\right)
\end{aligned}$$

and the last summand is uniformly $O\left(\zeta^{-2}(n)\right)$ by Lemma A.4(c). To bound the probability in (A.38), note that when $\mathcal{E}_i$ is true

$$G^{-1}\widehat{G}_{i,\tau}(d) = 1 + \mathcal{C}_{i,\tau} + o\left(m_{i,\tau}/n\right)^{2r(\tau)}, \tag{A.39}$$

where the small order term is $o\left(m_{i,\tau}/n\right)^{2r(\tau)}$ instead of $o_p\left(m_{i,\tau}/n\right)^{2r(\tau)}$ and

$$\begin{aligned}
T_{4,i\tau} &= -4\mathcal{A}_{i,\tau} + O\left(\frac{\ln m_{i,\tau}}{m_{i,\tau}} \frac{m_{i,\tau}^{\phi+0.5}}{n^{\phi}}\right) + O\left(\frac{m_{i,\tau}^{\tau+0.5}}{n^{\tau}}\right) \\
&= -4\mathcal{A}_{i,\tau} + O(1), \qquad \text{(A.40)}
\end{aligned}$$

where $\phi = \min(\tau, 2)$ and

$$\mathcal{A}_{i,\tau} = \sum_{k=1}^{r(\tau)} A_{2k} \frac{m_{i,\tau}^{2k+0.5}}{n^{2k}},\ \mathcal{A}_{i,\tau} = 0 \text{ if } r(\tau) = 0.$$

Using (A.39) and (A.40) and letting $\mathcal{B}_{i,\tau} = \sum_{k=1}^{r(\tau)} B_{2k} m_{i,\tau}^{2k}/n^{2k}$, we have

$$
\begin{aligned}
& \sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} G \widehat{G}_{i,\tau}^{-1}(d) \left(m_{i,\tau}^{-1} J_{i,\tau}\right)^{-1} T_{4,i\tau} \\
= & \sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} \left(1 + \mathcal{C}_{i,\tau} + o\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\right)^{-1} \left(\frac{1}{4} + O\left(\frac{\ln^2 m_{i,\tau}}{m_{i,\tau}}\right)\right) \left(-4\mathcal{A}_{i,\tau} + O(1)\right) \\
= & \sum_{i=1}^{K} w_{i,\tau} l_i^{-1/2} \left(1 + \mathcal{B}_{i,\tau} + o\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)}\right) \left(\frac{1}{4} + O\left(\frac{\ln^2 m_{i,\tau}}{m_{i,\tau}}\right)\right) \left(-4\mathcal{A}_{i,\tau} + O(1)\right) \\
= & -\sum_{i=1}^{K} w_{i,\tau} \left(1 + \mathcal{B}_{i,\tau}\right) \mathcal{A}_{i,\tau} + o\left(\left(\frac{m_{i,\tau}}{n}\right)^{2r(\tau)} \frac{m_{i,\tau}^{\phi+0.5}}{n^{\phi}}\right) + O\left(\frac{\ln^2 m_{i,\tau}}{m_{i,\tau}} \frac{m_{i,\tau}^{\phi+0.5}}{n^{\phi}}\right) + O\left(1\right) \\
= & O\left(\frac{m_{i,\tau}^{2r(\tau)+\phi+0.5}}{n^{2r(\tau)+\phi}}\right) + O(1) = O(1)
\end{aligned}
$$

uniformly, which implies that the probability in (A.38) is $O\left(\zeta^{-2}(n)\right)$. Combining this with (A.33) and (A.35) gives the desired result.

Part (b) For notational simplicity, set $\tau_0 := s_0 + 8s_0 h \ln\ln n$ so that

$$
\begin{aligned}
& P\left(m_{s_0}^{1/2} |\widehat{d}_{WLW}\left(s_0 + 8s_0 h \ln\ln n\right) - d| \le C\zeta(n)\right) \\
= & P\left(m_{\tau_0}^{1/2} |\widehat{d}_{WLW}\left(\tau_0\right) - d| \le m_{s_0}^{-1/2} m_{\tau_0}^{1/2} C\zeta(n)\right) \\
= & P\left(m_{\tau_0}^{1/2} |\widehat{d}_{WLW}\left(\tau_0\right) - d| \le \psi_1^{1/2} n^{-\frac{s_0}{2s_0+1} + \frac{\tau_0}{2\tau_0+1}} C\zeta(n)\right) \\
= & P\left(m_{\tau_0}^{1/2} |\widehat{d}_{WLW}\left(\tau_0\right) - d| \le C\left(\ln n\right)^{8s_0/(2s_0+1)(2\tau_0+1)} \zeta(n)\right). \qquad \text{(A.41)}
\end{aligned}
$$

The last equality holds because

$$
n^{-\frac{s_0}{2s_0+1} + \frac{\tau_0}{2\tau_0+1}} = n^{\frac{8s_0 h \ln\ln n}{(2s_0+1)(2\tau_0+1)}} = \left(\ln n\right)^{8s_0/(2s_0+1)(2\tau_0+1)}.
$$

As in the proof of Part (a), we write

$$
m_{\tau_0}^{1/2}\left(\widehat{d}_{WLW}\left(\tau_0\right) - d\right) = -\sum_{i=1}^{K} w_{i,\tau_0} l_i^{-1/2} m_{i,\tau_0}^{1/2} J_{i,\tau_0}^{-1} S_{i,\tau_0}(d) + e_{m,\tau_0}, \qquad \text{(A.42)}
$$

where for some d_{i,τ_0}^{+} between $\widehat{d}_{WLP}(\tau_0)$ and d,

$$
e_{m,\tau_0} := -\left(m_{\tau_0}^{1/2}(\widehat{d}_{WLP}(\tau_0) - d)\right)\left(\sum_{i=1}^{K} w_{i,\tau_0}(J_{i,\tau_0}^{-1} H_{i,\tau_0}(d_{i,\tau_0}^{+}) - 1)\right).
$$

We now consider e_{m,τ_0}, the second term in (A.42). Set

$$
V(\tau_0) := \left(m_{\tau_0} E\left(\widehat{d}_{WLP}(\tau_0) - d\right)^2\right)^{1/2}.
$$

It follows from Theorem 3.2 that

$$\begin{aligned} V(\tau_0) &= O\left(m_{\tau_0}^{0.5}\left(\frac{m_{\tau_0}}{n}\right)^{s_0}\right) = O\left(m_{\tau_0}^{0.5}\left(n^{\frac{2\tau_0}{2\tau_0+1}-1}\right)^{s_0}\right) \\ &= O\left(n^{\frac{\tau_0}{2\tau_0+1}} n^{\frac{-s_0}{2\tau_0+1}}\right) = O\left(n^{\frac{\tau_0-s_0}{2\tau_0+1}}\right) = O\left((\ln n)^{8s_0/(2\tau_0+1)}\right). \quad \text{(A.43)} \end{aligned}$$

Hence

$$\begin{aligned} P\left(d_{i,\tau_0}^{+} \notin D_{m_{i,\tau_0}}\right) &\leq P\left(\left|\widehat{d}_{WLP}(\tau_0) - d\right| > \zeta^{-2}(n) \ln^{-5} m_{i,\tau_0}\right) \\ &= O\left(m_{i,\tau_0}^{-1} \zeta^4(n) (\ln m_{i,\tau_0})^{10} (\ln n)^{16s_0/(2\tau_0+1)}\right) = O\left(\zeta^{-2}(n)\right). \end{aligned}$$

As a consequence, we can now focus on the case $d_{i,\tau_0}^{+} \in D_{m_{i,\tau_0}}$. To obtain a bound for $P\left(|e_{m,\tau_0}| > C\zeta^{-1}(n) V(\tau_0)\right)$, we use the following result

$$P\left(\sup_{d^{+} \in D_{m_{i,\tau_0}}} m_i^{-1}(\tau_0) \left|H_{i,\tau_0}(d^{+}) - J_{i,\tau_0}\right| > C\zeta^{-2}(n)\right) = O\left(\zeta^{-2}(n)\right), \quad \text{(A.44)}$$

which is slightly stronger than (A.25) with $\tau = \tau_0$. Inspection of the proof of Lemma 7 of AS (p. 609) shows that the above result holds. Using (A.44), we have,

$$\begin{aligned} & P\left(|e_{m,\tau_0}| > C\zeta^{-1}(n) V(\tau_0)\right) \\ \leq\ & P\left(m_{\tau_0}^{1/2} \left|\widehat{d}_{WLP}(\tau_0) - d\right| > CV(\tau_0)\zeta(n)\right) + C\zeta^{-2}(n) \\ \leq\ & C^{-2}\zeta^{-2}(n) + C\zeta^{-2}(n) = O\left(\zeta^{-2}(n)\right). \quad \text{(A.45)} \end{aligned}$$

Next, we consider the first term in (A.42). Using a proof similar to (A.36), we get for $a = 1, 2, 3$

$$P\left(\left|\sum_{i=1}^{K} w_{i,\tau_0} l_i^{-1/2} G \widehat{G}_{i,\tau_0}^{-1}(d) \left(m_{i,\tau_0} J_{i,\tau_0}^{-1}\right) T_{a,i\tau_0}\right| > C\zeta(n)\right) = O\left(\zeta^{-2}(n)\right). \quad \text{(A.46)}$$

Let

$$\xi(n, \tau_0) := C\,(\ln n)^{8s_0/(2s_0+1)(2\tau_0+1)}\, \zeta(n) + C\zeta^{-1}(n) V(\tau_0) + C\zeta(n),$$

then it follows from (A.41), (A.42), (A.45), and (A.46) that

$$\begin{aligned} & P\left(m_{s_0}^{1/2} |\widehat{d}_{WLW}\left(s_0 + 8s_0 h \ln\ln n\right) - d| \leq C\zeta(n)\right) \quad \text{(A.47)} \\ \leq\ & P\left(|\sum_{i=1}^{K} w_{i,\tau_0} l_i^{-1/2} G \widehat{G}_{i,\tau_0}^{-1}(d) \left(m_{i,\tau_0} J_{i,\tau_0}^{-1}\right) T_{4,i\tau_0}| < \xi(n, \tau_0)\right) + O\left(\zeta^{-2}(n)\right). \end{aligned}$$

Let $\mathcal{E}_i'$ denote the event $\{|G^{-1}\widehat{G}_{i,\tau_0}(d) - 1| < C\zeta^{-1}(n)$ and $|m_{i,\tau_0} J_{i,\tau_0}^{-1} - 1/4| < Cm_{i,\tau_0}^{-1} \ln^2 m_{i,\tau_0}\}$. A proof similar to the one for Lemma A.4(b) shows that

$$P\left(\left|G^{-1}\widehat{G}_{i,\tau_0}(d) - 1\right| > C\zeta^{-1}(n)\right) = O\left(\zeta^{-2}(n)\right).$$

It follows from the preceding equation and Lemma A.1(a) that $P((\cap_{i=1}^K \mathcal{E}_i')^c) = O\left(\zeta^{-2}(n)\right)$. As before, we now focus on the event $\cap_{i=1}^K \mathcal{E}_i'$, in which case,

$$G^{-1}\widehat{G}_{i,\tau_0}(d) = 1 + O\left(\zeta^{-1}(n)\right)$$

and

$$T_{4,i\tau_0} = -4\mathcal{A}_{i,\tau_0} + m_{i,\tau_0}^{-1/2}\sum_{j=1}^{m_{i,\tau_0}} R_T(\lambda_j)X_j^* + o(1),$$

implying that the dominating term in $\sum_{i=1}^K w_{i,\tau_0} l_i^{-1/2} G\widehat{G}_{i,\tau_0}^{-1}(d)\left(m_{i,\tau_0}J_{i,\tau_0}^{-1}\right) T_{4,i\tau_0}$ is at least $Cm_{\tau_0}^{-1/2}\sum_{j=1}^{m_{\tau_0}} R_T(\lambda_j)X_j^*$ for some nonzero constant C.

We now prove that $|m_{\tau_0}^{-1/2}\sum_{j=1}^{m_{\tau_0}} R_T(\lambda_j)X_j^*|$ can be bounded from below by $C\ln^2 n$ for some constant $C > 0$. We know that $m_{\tau_0}^{-1}\sum_{k=1}^{m_{\tau_0}} \ln k = \ln m_{\tau_0} - 1 + O(\ln m_{\tau_0}/m_{\tau_0})$ and therefore by the definition of X_j^* it is enough to bound $|B|$ from below by $C\ln^2 n$, where $B = m_{\tau_0}^{-1/2}\sum_{j=1}^{m_{\tau_0}} R_T(\lambda_j)(-\ln j + \ln m_{\tau_0} - 1)$. Set $\bar{m}_0 := m_{\tau_0}/e$. Then $-\ln j + \ln m_{\tau_0} - 1 \le 0$ iff $j \ge \bar{m}_{\tau_0}$. So when $R_T(\lambda) > 0$ for all $\lambda \in U$

$$\begin{aligned}
B &= m_{\tau_0}^{-1/2}\{\sum_{j=1}^{\bar{m}_0} R_T(\lambda_j)(-\ln j + \ln m_{\tau_0} - 1) + \sum_{j=\bar{m}_0+1}^{m_{\tau_0}} R_T(\lambda_j)(-\ln j + \ln m_{\tau_0} - 1)\} \\
&\le \frac{Cm_{\tau_0}^{-1/2}}{n^{s_0}}\{\sum_{j=1}^{\bar{m}_0} C_{\max} j^{s_0}(-\ln j + \ln m_{\tau_0} - 1) + \sum_{j=\bar{m}_0+1}^{m_{\tau_0}} C_{\min} j^{s_0}(-\ln j + \ln m_{\tau_0} - 1)\} \\
&\le \frac{Cm_{\tau_0}^{s_0+1/2}}{n^{s_0}}\{(C_{\max} - C_{\min})e^{-s_0-1} - C_{\min}s_0\}(s_0+1)^{-2} < 0
\end{aligned}$$

using (4.4) where $C > 0$ is a constant. Similarly, when $R_T(\lambda) < 0$ for all $\lambda \in U$,

$$\begin{aligned}
B &\ge -\frac{Cm_{\tau_0}^{-1/2}}{n^{s_0}}\{\sum_{j=1}^{\bar{m}_0} C_{\max} j^{s_0}(-\ln j + \ln m_{\tau_0} - 1) + \sum_{j=\bar{m}_0+1}^{m_{\tau_0}} C_{\min} j^{s_0}(-\ln j + \ln m_{\tau_0} - 1)\} \\
&\ge -\frac{Cm_{\tau_0}^{s_0+1/2}}{n^{s_0}}\{(C_{\max} - C_{\min})e^{-s_0-1} - C_{\min}s_0\}(s_0+1)^{-2} > 0
\end{aligned}$$

Therefore, under (4.4)

$$\begin{aligned}
|B| &\ge C^*\frac{m_{\tau_0}^{s_0+1/2}}{n^{s_0}} = n^{\frac{\tau_0 - s_0}{2\tau_0+1}}(C^* + o(1) = (\ln n)^{8s_0/(2\tau_0+1)}(C^* + o(1)), \\
&= (\ln n)^{8s_0/(2s_0+1)}(C^* + o(1)) \ge (\ln n)^{8/3}(C^* + o(1)),
\end{aligned}$$

where C^* is a positive constant. As a result, up to lower order terms

$$\left| Cm_{\tau_0}^{-1/2}\sum_{j=1}^{m_{\tau_0}} R_T(\lambda_j)X_j^*\right| \ge C^*(\ln n)^{8/3}.$$

However, since

$$\begin{aligned}(\ln n)^{8s_0/(2s_0+1)(2\tau_0+1)} &= C(\ln n)^{8s_0/(2s_0+1)^2}(1+o(1)) \\ &\leq C(\ln n)^{8/9}(1+o(1))\end{aligned}$$

as $8s_0/(2s_0+1)^2$ is a decreasing function of s_0 when $s_0 \geq 1$, we have

$$\xi(n,\tau_0) = O\left((\ln n)^{8/9}\zeta(n)\right) + o\left((\ln n)^{8s_0/(2\tau_0+1)}\right) + o\left(\ln^2 n\right)$$

and thus

$$P\left(|\sum_{i=1}^{K} w_{i,\tau_0} l_i^{-1/2} G\widehat{G}_{i,\tau_0}^{-1}(d)\left(m_{i,\tau_0} J_{i,\tau_0}^{-1}\right) T_{4,i\tau_0}| < \xi(n,\tau_0)\right) = 0,$$

when n is large enough. Combining this with (A.47) leads to the stated result. ■

Notes

[1]More precisely, instead of $-2\log\lambda_j$, the original GPH estimator was defined using $\log|1-\exp(-i\lambda)|$ as the regressor for d. For the asymptotic statements given in our theorems, this difference is of no importance.

[2]Note that we do not include Assumption 3(b) in AS into our AS3. It is used in Robinson (1995a) to obtain a law of large number for ε_t^2 without fourth moment assumptions. But AS3(a) yields this law of large number, since it implies that ε_t^2-1 is a bounded variance martingale difference sequence.

In AS4 note the following. If $s > 2+2r$ is assumed as in AG2 above, Assumption 4 in AS boils down to $m^{0.5+2r}/n^{2r} \to \infty$ and $m^{2.5+2r}/n^{2+2r} = O(1)$ as $n\to\infty$. If we want to be certain that in (2.10) $\tau_r b_{2+2r} m^{2+2r}/n^{2+2r}$ is the dominant higher order bias expression, we have to sharpen assumption $m^{2.5+2r}/n^{2+2r} = O(1)$ to $m^{2.5+2r}/n^{2+2r} \to \chi \in (0,\infty)$ in AS4. The reason is that $m^{2.5+2r}/n^{2+2r} = O(1)$ is compatible with $m^{2.5+2r}/n^{2+2r} = o(1)$ in which case it is not clear that $\tau_r b_{2+2r} m^{2+2r}/n^{2+2r}$ is the dominant higher order bias expression: there could be other $o_p(1)$ terms that converge to zero slower than $\tau_r b_{2+2r} m^{2+2r}/n^{2+2r}$.

[3]In the discussion of the local Whittle estimator, we mean by "asymptotic bias" the difference of the mean of the asymptotic distribution of the estimator and the true value of d. In contrast, for the log-periodogram estimator, "asymptotic bias" refers to the asymptotic behavior of $E\widehat{d}_{LP} - d$. For technical reasons, we can not use this definition for the Whittle case. The same comment applies to "asymptotic variance". For further discussion of the difference of these two concepts we refer to Davidson and MacKinnon (1993, p.124).

[4]If $m^{2.5+2r}/n^{2+2r} = o(1)$ then $\sqrt{m}(\widehat{d}_{LP}(r,m)-d) \to_d N(0,\frac{\pi^2}{24}c_r)$ and similar modifications hold for the other asymptotic normality results below. We focus on the case $m^{2.5+2r}/n^{2+2r} \to \chi \in (0,\infty)$ in our discussion here because this is the MSE-optimal growth rate for m.

[5]W.l.o.g. we impose $l_j > l_{j-1}$ and for convenience, we always pick l such that the increments $l_{j+1}-l_j$ are equal, for $j=1,...,K-1$. Furthermore, m and l_K are picked such that $[l_K m] \le n/2$. Assuming that l is a vector with rational components $l_i = p_i/q_i$ (for relatively prime integers p_i and q_i and $i=1,...,K$) and that m contains all the q_i as factors, it follows that $m_i = l_i m$. To simplify the exposition, we make this assumption from now on. Of course, all the results below go through without this assumption. For example, take the bias formula in Theorem 3.2: because $m_i^{2k}/m^{2k} = l_i^{2k} + O(1/m)$, condition (3.3) implies that for $k=1,...,r$ we have $\sum_{i=1}^K w_i m_i^{2k}/n^{2k} = O(m^{2k-1}/n^{2k})$ which is dominated by $O((\ln^3 m)/m)$.

[6]To simplify notation, we do not explicitly list K and w among the parameters that the weighted estimator depends on. The dependence on K is implicit

through l and below we describe a procedure of how to pick w given r, m, l, and δ.

[7] k-step procedures based on Newton-Raphson-type iterations are probably more common. We could also define a k-step (Newton-Raphson-type) local Whittle estimator by the recursion $\widehat{d}_{LW,j} := \widehat{d}_{LW,j-1} - H_m(\widehat{d}_{LW,j-1})^{-1} S_m(\widehat{d}_{LW,j-1})$, for $j = 1, 2, ..., k$, where we used the same notation as in (3.12). The performance of the Gauss-Newton- and Newton-Raphson-type estimators are very similar. In fact, Lemma 2(b) and (d) of AS show that for $\eta > 0$ we have $\sup_{|\overline{d}-d| \leq \eta \ln^{-5} m} m^{-1} \left| H_m(\overline{d}) - J_m \right| = o_p(1)$ which together with (A.6) and (A.7) can be used to establish the equivalence of the two procedures. We use J_m and not $H_m(\widehat{d})$ because it is technically more convenient and computationally simpler because J_m does not depend on any estimated parameter.

[8] Note that there are functions $h \in C^1(U)$ for which Definition 4.1 does not apply. Assume h is such that its T is finite with $R_T(\lambda) = |\lambda|^{2T} / \log |1/\lambda|$ for $\lambda \neq 0$ and $R_T(0) = 0$. Then, there is no $\tau > 2T$ and positive $C_{\min}$ such that (4.1) holds.

[9] Assumption AS4 implies that $O((\ln^2 m/m)(m^{2.5}/n^2)) = o(1)$ and thus the statement follows by noting that $m^{2.5}/n^2$ is the highest order term of T_4 that hits $O(\ln^2 m/m)$. For similar reasons we can ignore the $O(\ln m/m)$ terms in T_4 in equation (A.20).

References

Agiakloglou, C., P. Newbold & M. Wohar (1993). Bias in an Estimator of the Fractional Difference Parameter. *Journal of Time Series Analysis* 14, 235–246.

Andrews, D.W.K. & P. Guggenberger (2003). A Bias-Reduced Log-periodogram Regression Estimator for the Long-Memory Parameter. *Econometrica* 71(2), 675-712.

Andrews, D.W.K. & Y. Sun (2004). Adaptive Local Polynomial Whittle Estimation of Long-range Dependence. *Econometrica* 72(2), 569-614.

Brillinger, D.R. (1975). *Time Series, Data Analysis and Theory.* San Francisco: Holden-Day.

Davidson, R. & J.G. MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford University Press.

Geweke, J. & S. Porter-Hudak (1983). The Estimation and Application of Long-memory Time Series Models. *Journal of Time Series Analysis* 4, 221–237.

Giraitis, L., P.M. Robinson & A. Samarov (2000). Adaptive Semiparametric Estimation of the Memory Parameter. *Journal of Multivariate Analysis* 72, 183–207.

Gray, H.L. & W.R. Schucany (1972). *The Generalized Jackknife Statistic.* Dekker, New York.

Guggenberger, P. & Y. Sun (2003). Bias-Reduced Log-Periodogram and Whittle Estimation of the Long-Memory Parameter Without Variance Inflation. Previous version of this paper, available at http://www.econ.ucla.edu/people/faculty/Guggenberger.html.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge, UK.

Hidalgo, J. (2005). Semiparametric estimation for stationary processes whose spectra have an unknown pole. Forthcoming in *Annals of Statistics.*

Hurvich, C.M. (2001). Model Selection for Broadband Semiparametric Estimation of Long Memory in Time Series. *Journal of Time Series Analysis* 22, 679–709.

Hurvich, C.M. & J. Brodsky (2001). Broadband Semiparametric Estimation of the Memory Parameter of a Long-Memory Time Series Using Fractional Exponential Models. *Journal of Time Series Analysis* 22, 221-249.

Hurvich, C.M. & R.S. Deo (1999). Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-memory Time Series. *Journal of Time Series Analysis* 20, 331–341.

Hurvich, C.M., R.S. Deo & J. Brodsky (1998). The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-memory Time Series. *Journal of Time Series Analysis* 19, 19–46.

Hurvich, C.M, E. Moulines & P. Soulier (2002). The FEXP estimator for potentially non-stationary linear time series. *Stochastic Processes and Their Applications* 97, 307-340.

Iouditsky, A., E. Moulines & P. Soulier (2001). Adaptive Estimation of the Fractional Differencing Coefficient. *Bernoulli* 7, 699-731.

Kim, C.S. & P.C.B. Phillips (1999a). Log Periodogram Regression: The Nonstationary Case. Working paper, Cowles Foundation, Yale University.

———(1999b). Modified Log Periodogram Regression. Working paper, Cowles Foundation, Yale University.

Künsch, H.R. (1987). Statistical Aspects of Self-similar Processes. *Proceedings of the First World Congress of the Bernoulli Society* 1, 67-74, ed. by Yu. Prohorov and V.V. Sazanov. Utrecht: VNU Science Press.

Lepski, O.V. (1990). On a Problem of Adaptive Estimation in Gaussian White Noise. *Theory of Probability and Its Applications* 35, 454-466.

Lobato, I.N. (1999). A semiparametric two step estimator in a multivariate long memory model. *Journal of Econometrics* 90, 129-153.

Moulines, E. & P. Soulier (1999). Broadband Log-Periodogram Regression of Time Series With Long-Range Dependence. *Annals of Statistics* 27(4), 1415-1439.

———(2000). Data Driven Order Selection for Projection Estimator of the Spectral Density of Time Series With Long Range Dependence. *Journal of Time Series Analysis* 21(2), 193-218.

Phillips, P.C.B. and K. Shimotsu (2004). Local Whittle Estimation in Nonstationary and Unit Root Cases. *Annals of Statistics* 32(2), 656-692.

Robinson, P.M. (1995a). Gaussian Semiparametric Estimation of Long Range Dependence. *Annals of Statistics* 23, 1630–1661.

———(1995b). Log-periodogram Regression of Time Series with Long Range Dependence. *Annals of Statistics* 23, 1048–1072.

Robinson, P.M. & M. Henry (2003). Higher-order Kernel Semiparametric M-estimation of Long Memory. *Journal of Econometrics* 114(1), 1-27.

Shimotsu, K. & P.C.B. Phillips (2002). Exact Local Whittle Estimation of Fractional Integration. Forthcoming in *Annals of Statistics,* available at http://cowles.econ.yale.edu.

Velasco, C. (1999). Non-stationary Log-periodogram Regression. *Journal of Econometrics* 91, 325–371.

———(2000). Non-Gaussian Log-Periodogram Regression. *Econometric Theory* 16, 44-79.

Velasco, C. & P.M. Robinson (2000). Whittle Pseudo-maximum Likelihood Estimation for Nonstationary Time Series. *Journal of the American Statistical Association* 95, 1229–1243.

Walter, W. (1992). *Analysis I.* Springer, Berlin.