Empirical likelihood confidence intervals for the mean of a long-range dependent process

Dan Nordman[1]

Philipp Sibbertsen[2]

Soumendra N. Lahiri[3]

Abstract: This paper considers blockwise empirical likelihood for real-valued linear time processes which may exhibit either short- or long-range dependence. Empirical likelihood approaches intended for weakly dependent time series can fail in the presence of strong dependence. However, a modified blockwise method is proposed for confidence interval estimation of the process mean, which is valid for various dependence structures including long-range dependence. The finite-sample performance of the method is evaluated through a simulation study and compared to other confidence interval procedures involving subsampling or normal approximations.

[1]Dept. of Statistics, Iowa State University, Ames, IA 50011, USA

[2]Fakultät Wirtschaftswissenschaften, Abteilung Statistik, Königsworther Platz 1, 30167 Hannover, Germany

[3]Dept. of Statistics, Iowa State University, Ames, IA 50011, USA

1 Introduction

Empirical likelihood (EL), proposed by Owen (1988, 1990), makes possible likelihood-based statistical inference without a specified distribution for the data. For independent data, the method generates a nonparametric likelihood that shares many qualities associated with parametric likelihood, such as limiting chi-square distributions for log-likelihood ratios [cf. Hall and La Scala (1990), Qin and Lawless (1994)]. Owen (2001) describes further properties and applications of EL, where much research has focused on EL with independent observations. The aim of this paper is to extend EL inference to dependent data, which could exhibit either weak or potentially strong forms of dependence. We show that EL inference on the process mean is possible for a large class of time series, which we next detail.

Let $\mathbb{Z}$ denote the set of integers. We assume that the data $(Y_1, \ldots, Y_n)$ represent a realization from a strictly stationary process $\{Y_t\}$, $t \in \mathbb{Z}$, with mean μ and an integrable spectral density function $f(\lambda), |\lambda| \leq \pi$, that satisfies

$$f(\lambda) \sim \mathcal{C}_d |\lambda|^{-2d}, \quad \text{as } \lambda \to 0, \tag{1}$$

for some $d \in (-1/2, 1/2)$ and positive constant $\mathcal{C}_d > 0$ (where $\sim$ denotes that the ratio of two terms equals one in the limit). There then exists a process $\{\varepsilon_j\}$ of uncorrelated random variables with mean $\mathrm{E}(\varepsilon_j) = 0$ and variance $0 < \mathrm{E}(\varepsilon_j^2) < \infty$ such that $\{Y_t\}$ has a moving average representation of the form [Chapter 16.7, Ibragimov and Linnik (1971)]:

$$Y_t = \mu + \sum_{j \in \mathbb{Z}} b_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{2}$$

where the constants $\{b_j\}$ fulfill $0 < \sum_{j \in \mathbb{Z}} b_j^2 < \infty$. When $d = 0$, we refer to the process $\{Y_t\}$ as weakly or short-range dependent (SRD). The process $\{Y_t\}$ will be called strongly or long-range dependent (LRD) if $d > 0$. For $d < 0$, the process shall be termed "anti-persistent." This classification of $\{Y_t\}$ as SRD or LRD depending on the value of d is common, in which long-range dependence (LRD) entails a pole of f at the origin [Hosking (1981), Beran (1994)]. Formulations of weak dependence based on mixing assumptions often imply the $d = 0$ case of short-range dependence (SRD).

Kitamura (1997) has proposed a blockwise EL method for weakly dependent time series satisfying a mixing condition to ensure SRD. This method applies the Owen (1990) formulation of EL to observational blocks rather than individual observations. As described in Kitamura (1997), blockwise EL involves data blocking techniques used successfully by other

nonparametric likelihood methods for handling weak dependence, such as the block bootstrap and subsampling procedures [Carlstein (1986), Künsch (1989), Politis and Romano (1993)].

However, blocking methods developed for SRD can often fail under LRD and require careful modification. For example, the moving block bootstrap of Künsch (1989) and Liu and Singh (1992) is known to break down under forms of LRD [Lahiri (1993)]. The block-based subsampling methods of Politis and Romano (1993) and Hall and Jing (1996), also developed for weak dependence, must be significantly reformulated for variance estimation with LRD [Hall, Jing and Lahiri (1998), Nordman and Lahiri (2005)]. Since the blockwise EL of Kitamura (1997) also involves data blocking intended for SRD, one might expect strong dependence to create complications with this version of EL as well. Indeed, our results indicate that this blockwise EL formulation can break down under either LRD or anti-persistence, partially because the usual standardization $\sqrt{n}(\bar{Y}_n - \mu)$ of a size n sample mean $\bar{Y}_n$ might fail to produce a non-degenerate limit.

This paper finds that an adapted version of Kitamura's (1997) EL is applicable for confidence intervals with time processes exhibiting very different kinds of dependence structures. We give a modified blockwise EL method for interval estimation of the process mean that is valid for SRD or LRD linear time series satisfying (1). This EL approach also has an advantage over confidence intervals based directly on the sample mean $\bar{Y}_n$ in that estimation of $\text{Var}(\bar{Y}_n)$ is *not* required. In the weak dependence case (e.g., $d = 0$), we establish results on blockwise EL for the mean in a manner alternative to Kitamura (1997), which does not involve mixing conditions to quantify the decay rate of the process $\{Y_t\}$ covariances.

The rest of the paper is organized as follows. Section 2 details some background necessary for the main results, including the blockwise EL construction and the assumptions. The limiting distribution of a blockwise EL ratio under LRD is given in Section 3 and a method of setting EL confidence intervals for the process mean is proposed. Section 4 provides simulation studies of these EL confidence intervals and numerical comparisons to other confidence interval procedures. Section 5 contains proofs of the main results.

2 Preliminaries

To describe the main results, we first review the blockwise EL of Kitamura (1997) in Section 2.1 and outline our assumptions in Section 2.2. In Section 2.3, we also relate some distribution

properties of the sample mean $\bar{Y}_n$ to the behavior of blockwise EL ratios under LRD.

2.1   Construction of blockwise empirical likelihood for the mean

We give an EL function for the process mean $E(Y_t) = \mu$ following the blockwise formulation given in Kitamura (1997) with a maximally overlapping version of blocking. Let $1 \leq \ell \leq n$ be the block length and define blocks $\mathcal{B}_i^\ell = (Y_i, \ldots, Y_{i+\ell-1})$, $1 \leq i \leq N \equiv n - \ell + 1$. Let $M_{\ell i} = \sum_{j=i}^{i+\ell-1} Y_j / \ell$ denote the sample mean of the ℓ variables in $\mathcal{B}_i^\ell$, $1 \leq i \leq N$.

To assess the plausibility of a value for the mean parameter, we assign probabilities $\{p_i\}_{i=1}^N$ to each block sample mean $\{M_{\ell i}\}_{i=1}^N$ to develop a multinomial likelihood function, $\prod_{i=1}^N p_i$. The profile blockwise EL function for μ is given by

$$L_n(\mu) = \sup \left\{ \prod_{i=1}^N p_i : p_i > 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i M_{\ell i} = \mu \right\} \tag{3}$$

where we maximize the product in (3) over distributions or probabilities $\{p_i\}_{i=1}^N$ on $\{M_{\ell i}\}_{i=1}^N$ with expectation μ. The function $L_n(\mu)$ is positive for $\mu \in (\min_{1 \leq i \leq N} M_{\ell,i}, \max_{1 \leq i \leq N} M_{\ell,i})$ and outside of this interval we may define $L_n(\mu) = -\infty$. When positive, $L_n(\mu)$ represents a maximum realized at unique weights $p_i = N^{-1}\{1 + \lambda_\mu(M_{\ell i} - \mu)\}^{-1}$ where $\lambda_\mu \in \mathbb{R}$ is determined by

$$0 = \sum_{i=1}^N \frac{M_{\ell i} - \mu}{1 + \lambda_\mu(M_{\ell i} - \mu)} \equiv g_\mu(\lambda_\mu). \tag{4}$$

Owen (1988, 1990) discusses these and further computational aspects of EL.

Without the "mean μ" linear constraint in (3), the product $\prod_{i=1}^N p_i$ is maximized when each $p_i = N^{-1}$, corresponding to the empirical distribution of $\{M_{\ell i}\}_{i=1}^N$. The profile empirical likelihood ratio for the mean μ is then given by

$$R_n(\mu) = \frac{L_n(\mu)}{N^N} = \prod_{i=1}^N \{1 + \lambda_\mu(M_{\ell i} - \mu)\}^{-1} \tag{5}$$

and a confidence interval for the mean is determined by those values of μ with relatively high EL, namely, intervals of the form

$$\{\mu : R_n(\mu) \geq A\}, \tag{6}$$

where $A > 0$ is chosen to obtain a desired confidence level.

A key feature of EL with independent data is that it allows a nonparametric casting of Wilks's theorem for constructing confidence regions [Wilks (1938)], meaning that the logarithm of the EL ratio has an asymptotic chi-square distribution when evaluated at the true parameter

value [cf. Owen (1988, 1990)]. Kitamura (1997) first showed a similar result applies to the blockwise EL with weakly dependent mixing processes. In Section 3.1, we give the asymptotic distribution of the EL ratio in (5) for SRD and LRD linear time processes fulfilling (1), as required for calibrating the confidence intervals in (6).

2.2 Time process assumptions

We require the following assumptions for proving the our EL results. For two sequences $\{s_n\}$, $\{t_n\}$ of nonzero numbers, we write $s_n \sim t_n$ to denote $\lim_{n\to\infty} s_n/t_n = 1$.

Assumptions:

(A.1). The strictly stationary process $\{Y_t\}$, $t \in \mathbb{Z}$, has a spectral density as in (1). For $0 < |d| < 1/2$, the process autocovariances $r(k) = \mathrm{Cov}(Y_t, Y_{t+k})$, $k \in \mathbb{Z}$ satisfy

$$r(k) \sim \mathcal{C}_d \mathcal{R}_d \cdot k^{-(1-2d)}, \quad \text{as } k \to \infty, \tag{7}$$

with the constant $\mathcal{C}_d$ from (1) and some real $\mathcal{R}_d \neq 0$; if $d = 0$, then f is bounded on every compact subinterval of $(0, \pi]$.

(A.2). The random variables $\{\varepsilon_j\}$ in (2) are independent and identically distributed (iid).

(A.3). As $n \to \infty$, the block length conditions hold: $\ell^{-1} + \ell^2/n = o(1)$ and if $\mathrm{E}(Y_t) = \mu_0$ denotes the true process mean, then the standardized block means satisfy

$$\left(\frac{\ell}{n}\right)^{1/2-d} \max_{1 \leq i \leq N} \frac{|M_{\ell i} - \mu_0|}{\sqrt{\mathrm{Var} M_{\ell 1}}} \xrightarrow{p} 0, \tag{8}$$

where $\xrightarrow{p}$ denotes convergence in probability.

The process assumptions in (A.1)-(A.2) encompass many types of time series used for describing both SRD and LRD, including two important models for LRD: the fractional Gaussian processes of Mandelbrot and Van Ness (1968) and the fractional autoregressive integrated moving average (FARIMA) models of Adenstedt (1974), Granger and Joyeux (1980) and Hosking (1981).

When $d > 0$, we assume the covariances of the LRD process decay very slowly so that the sum of autocovariances $\sum_{k=1}^{\infty} r(k)$ diverges, whereas the process covariances sum to zero $\sum_{k \in \mathbb{Z}} r(k) = 0$ when $d < 0$. Under Assumption (A.1), the behavior of f in (1) and the autocovariance decay in (7) are closely related when $d \neq 0$. If the covariances $r(k)$ are positive

4

and strictly decreasing [Bingham *et al.* (1987), p. 240] or quasimonotonically convergent to zero [Robinson (1995a), p. 1634], then (1) and (7) are equivalent for prescribing LRD ($d > 0$). In the context of log-periodogram regression for $d \neq 0$ for example, it is not uncommon to assume (1) and (7) both hold with $\mathcal{R}_d = 2\Gamma(1 - 2d)\sin(d\pi)$, where $\Gamma(\cdot)$ denotes the gamma function [cf. Robinson (1995b), Theorem 1]. If $\mathcal{R}_d$ has this form, the autocovariance representation (7) implies the limit of f in (1) for the case $d < 0$ [Robinson (1995a), p. 1634]; when $d > 0$, if f is of bounded variation on compact subintervals of $(0, \pi]$, then the pole of f in (1) implies that (7) holds with $\mathcal{R}_d$ as above [(2.24.V) of Zygmund (1968)].

By our condition for SRD under (A.1) (i.e., $d = 0$), we assume that a finite, positive limit $\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n}(n - k)r(k)$ exists but we do not explicitly require the absolute summability of the covariances $\sum_{k=1}^{\infty} |r(k)| < \infty$, which is a slightly stronger assumption that is often used to categorize SRD.

Assumption (A.3) is a general and mild formulation of block length conditions, which are often stated in terms of the process moments. It holds, for example, if $\mathrm{E}|Y_t|^q < \infty$ for some $q > 2/(1 - 2d)$ with a corresponding block length $\ell = O(n^{1/2 - 1/q(1-2d)})$; when $d = 0$, this formulation matches a blocking assumption of Kitamura (1997) for blockwise EL under SRD. For *any* block ℓ with $\ell^{-1} + \ell^2/n = o(1)$, condition (8) also follows if $\mathrm{E}|Y_t|^q < \infty$ for an even integer $q \geq 4/(1 - 2d)$ (see Section 5, proof of Theorem 2).

2.3   Asymptotic distribution of the sample mean and related issues

To help motivate a theory for EL under LRD, we review some differences in the behavior of the sample mean $\bar{Y}_n = \sum_{i=1}^{n} Y_i/n$ of SRD and LRD data $(Y_1, \ldots, Y_n)$. Theorem 1 describes how the correct scaling a_n for the standardized sample mean $(\bar{Y}_n - \mu)/a_n$ to have a limit distribution depends crucially on the dependence parameter d in (1). In the following, denote convergence in distribution by $\xrightarrow{d}$.

Theorem 1 *Let $\{Y_t\}, t \in \mathbb{Z}$ be a process which satisfies Assumption (A.1), $d \in (-1/2, 1/2)$.*
(a) As $n \to \infty$, $n^{1-2d}\mathrm{Var}(\bar{Y}_n) \to \mathcal{V}_d > 0$, where $\mathcal{V}_d = \mathcal{C}_d\mathcal{R}_d/\{d(1 + 2d)\}$ if $d \neq 0$ and $\mathcal{V}_0 = 2\pi\mathcal{C}_0$ for $d = 0$.
(b) Let $a_n^2 = n^{-1+2d}\mathcal{V}_d$. If Assumption (A.2) holds in addition, then $(\bar{Y}_n - \mu)/a_n \xrightarrow{d} Z$ as $n \to \infty$, where Z denotes a standard normal random variable.

Remark: For $d > 0$, the result in Theorem 1(a) is well-known and follows from the covariance decay in (7) [Taqqu (1977), Lemma A2]; Theorem 1(b) follows from part (a) due to Davydov (1970) and Theorem 18.6.5 of Ibragimov and Linnik (1971).

Under SRD, the variance $\text{Var}(\bar{Y}_n)$ of the sample mean decays at the usual $O(1/n)$ rate associated with iid data. However, under LRD or anti-persistence, the decay rate of $\text{Var}(\bar{Y}_n)$ may be slower or faster than $O(1/n)$, which implies that the usual $\sqrt{n}$ scaling of the sample mean $\sqrt{n}(\bar{Y}_n - \mu)$ may fail to produce a limit distribution. This aspect of LRD complicates inference on the process mean $\text{E}(Y_t) = \mu$ based on either the sample mean $\bar{Y}_n$ or data blocking techniques developed for SRD, such as the block bootstrap, subsampling and even the blockwise EL in Kitamura (1997) [cf. Lahiri (1993), Hall, Jing and Lahiri (1998)].

To recognize potential problems with blockwise EL under LRD, it is helpful to detail some mechanics behind the method with SRD. For inference on the process mean $\text{E}(Y_t) = \mu$, Kitamura (1997) explains that the non-blocking version of EL fails under SRD by neglecting the data dependence structure (i.e., using individual observations) which causes the variance $\text{Var}(\bar{Y}_n)$ of a sample mean to be improperly estimated. To be more precise, under suitable conditions [Owen (1989) with iid data; Kitamura (1997) under mixing; Theorem 2 in Section 3.1], the logarithm of the blockwise EL ratio in (5) may be expanded as

$$-2 \log R_n(\mu) = \frac{(\bar{Y}_n - \mu)^2}{\text{Var}(\bar{Y}_\ell)/N} \left(1 + o_p(1)\right). \tag{9}$$

The $\text{Var}(\bar{Y}_\ell)$ term appears because, in the inner mechanics of the blockwise EL, the collection of smaller length $\ell < n$ data blocks is used to estimate the variance $\text{Var}(\bar{Y}_\ell)$ of a size ℓ sample mean. While a block length $\ell = 1$ is appropriate for (9) to have a chi-square limit with an iid process Y_t (i.e., $N = n$, $\text{Var}(\bar{Y}_\ell) = \text{Var}(Y_1)$), this is not generally true under data dependence. However, the blockwise EL of Kitamura (1997) permits correct variance estimation *under weak dependence* $(d = 0)$ by exploiting the following property of SRD processes:

$$\frac{\text{Var}(\bar{Y}_n)}{\text{Var}(\bar{Y}_\ell)} \sim \frac{\ell}{n} \tag{10}$$

for time series lengths $\ell, n \to \infty$, $\ell/n \to 0$. Hence, under SRD the denominator of (9) can be re-scaled for inference on the full size n sample mean by (10): $\text{Var}(\bar{Y}_n) \sim (\ell/n)\text{Var}(\bar{Y}_\ell)$. As a consequence, blockwise EL statistics in Kitamura (1997) require adjustments involving n/ℓ. Note that, under SRD, subsampling techniques provide similar block-based variance estimates

of $\mathrm{Var}(\bar{Y}_n)$ using (10) [Politis and Romano (1993)]. However, we see from Theorem 1 that (10) may fail outside of the SRD case which implies the blockwise EL of Kitamura (1997) must be modified for LRD.

3 Blockwise empirical likelihood under long-range dependence

3.1 Limit distribution of empirical likelihood ratio

We now give the asymptotic distribution of the blockwise EL ratio in (5) for a general class of SRD or LRD time processes fulfilling (1). As in the EL framework of Kitamura (1997), we require a block correction factor to ensure our EL ratio has a non-degenerate limit. Write $B_n = N^{-1}(n/\ell)^{1-2d}$ for the block adjustment term.

Theorem 2 *Suppose* $\{Y_t\}, t \in \mathbb{Z}$ *satisfies Assumptions* (A.1)-(A.3). *If* $\mathrm{E}(Y_t) = \mu_0$ *denotes the true process mean, then as* $n \to \infty$,

$$-2B_n \log R_n(\mu_0) \overset{d}{\longrightarrow} \chi_1^2 \tag{11}$$

where χ_1^2 *denotes a chi-square random variable with 1 degree of freedom.*

In the case of SRD with $d = 0$ in (1), the block adjustment factor B_n reduces to the one $(n/\ell)N^{-1}$ used by Kitamura (1997) [p. 2089] for weakly dependent processes. However, we find that a blocking adjustment B_n that is appropriate for SRD *or* LRD must generally incorporate the dependence exponent d from (1).

The EL ratio in Theorem 2 involves maximally overlapping blocks, but the result also holds with another common blocking scheme involving non-overlapping blocks given by $\mathcal{B}^\ell_{i,\mathrm{NOL}} = \mathcal{B}^\ell_{1+\ell(i-1)}$ for $1 \le i \le N_{\mathrm{NOL}} \equiv \lfloor n/\ell \rfloor$, where $\lfloor n/\ell \rfloor$ represents the smallest integer not exceeding n/ℓ. Theorem 2 then follows upon substituting N_{NOL} for N and using block sample means from $\mathcal{B}^\ell_{i,\mathrm{NOL}}$, $1 \le i \le N_{\mathrm{NOL}}$, in (5).

3.2 Empirical likelihood confidence intervals for the mean

When the dependence structure is unknown, we require a consistent estimator of d in the block adjustment B_n without rigid model assumptions and several such estimators $\hat{d}_n$ are available based on the periodogram (see Section 3.3). Provided with an estimator $\hat{d}_n$ of d, we use (11)

to set an EL confidence interval for the process mean μ. Namely, an approximate two-sided $100(1-\alpha)\%$ confidence interval for μ, of the form (6), is given by

$$\mathcal{I}_n(1-\alpha) = \left\{ \mu \in \mathbb{R} : -2\widehat{B}_n \log R_n(\mu) \leq \chi^2_{1;1-\alpha} \right\}, \qquad \widehat{B}_n = N^{-1}(n/\ell)^{1-2\hat{d}_n}, \qquad (12)$$

where $\chi^2_{1;1-\alpha}$ represents the $1-\alpha$ percentile of a χ^2_1 random variable, $0 < \alpha < 1$. One-sided approximate $100(1-\alpha)\%$ lower/upper confidence bounds for μ correspond to the lower/upper endpoints of the interval $\mathcal{I}_n(1-2^{-1}\alpha)$. Under mild consistency conditions on $\hat{d}_n$, the above EL interval has asymptotically correct coverage for any dependence type d under (1), as detailed in Theorem 3.

Theorem 3 *Suppose* $\{Y_t\}, t \in \mathbb{Z}$ *satisfies Assumption* (A.1)-(A.3) *and the estimator* $\hat{d}_n$ *of* $d \in (-1/2, 1/2)$ *fulfills* $|\hat{d}_n - d| \log n \xrightarrow{p} 0$. *If* $\mathrm{E}(Y_t) = \mu_0$ *denotes the true process mean, then as* $n \to \infty$,

$$P\Big(\mu_0 \in \mathcal{I}_n(1-\alpha)\Big) \longrightarrow 1 - \alpha.$$

EL confidence intervals for the mean μ have an advantage over intervals set immediately with $\bar{Y}_n$ under Theorem 1 because direct variance estimation of $\mathrm{Var}(\bar{Y}_n)$ is not needed. That is, an approximate $100(1-\alpha)\%$ confidence interval based on a normal approximation

$$\bar{Y}_n \pm \frac{z_{1-\alpha/2}}{n^{1/2-d}} \sqrt{\frac{\mathcal{C}_d \mathcal{R}_d}{d(1+2d)}}, \qquad \mathcal{R}_d = \begin{cases} 2\Gamma(1-2d)\sin(d\pi) & d \neq 0 \\ 2\pi & d = 0, \end{cases} \qquad (13)$$

would require at least estimation of the constant $\mathcal{C}_d$ in (1) in addition to an estimate of the exponent d. (Above $z_{1-\alpha/2}$ denotes $(1-\alpha/2)$ percentile of a standard normal and we assume a form for the covariance decay in (7).) However, semi-parametric estimates for $\mathcal{C}_d$, which do not require strong assumptions on f, are known to be less accurate (i.e., slower convergence rates) than corresponding estimators for d, as in the case of log-periodogram regression [Robinson (1995b)].

The "sampling window" (SW) method of Hall, Jing and Lahiri (1998) is another nonparametric method for confidence interval estimation that uses a data-blocks [see Nordman and Lahiri (2005) for the SW applied to linear processes]. Both the EL and SW methods require a block length choice ℓ as well as a second tuning parameter related to variance estimation. Specifically, the EL approach requires an estimate of d (avoiding direct estimation of $\mathrm{Var}(\bar{Y}_n)$) while the SW procedure requires a parameter $[\theta \in (0,1)$ in Hall, Jing and Lahiri (1998)] for a device to directly estimate $\mathrm{Var}(\bar{Y}_n)$ based on data-blocks. A potential advantage to estimating

d in the EL approach, rather than $\text{Var}(\bar{Y}_n)$ directly, is greater precision in the resulting confidence intervals for linear time processes satisfying (1); numerical studies in Section 4 support that EL intervals are usually much *shorter* than SW intervals for such processes. However, the EL method is not immediately applicable whenever the sample mean of the underlying time process has a non-normal limit (as the EL method involves chi-square calibrations), while the SW may be valid. Hall, Jing and Lahiri (1998) consider the SW for a class of transformed processes producing non-normal limits for $\bar{Y}_n$.

3.3 Estimation of d

For the EL interval in (12), we describe two possible frequency domain estimators $\hat{d}_n$ for the dependence parameter d. Define the periodogram $I_n(\lambda) = (2\pi n)^{-1}|\sum_{t=1}^n Y_t e^{\lambda t \sqrt{-1}}|^2$, $\lambda \in [0, \pi]$ and let m be a bandwidth parameter such that $m \to \infty, m/n \to 0$ as $n \to \infty$.

The popular Geweke-Porter-Hudak (GPH) estimator of d is found by a regression of the log-periodogram against the first m Fourier frequencies:

$$\hat{d}_{m,n}^{GPH} = \frac{\sum_{j=1}^m (g_j - \bar{g}) \log(I_n(\lambda_j))}{\sum_{j=1}^m (g_j - \bar{g})^2},$$

where $\bar{g} = \sum_{j=1}^m g_j/m$ with $g_j = -2 \log|1 - e^{\lambda_j \sqrt{-1}}|$, $\lambda_j = 2\pi j/n$, $j = 1, \ldots, m$ [Geweke and Porter-Hudak (1983)]. Consistency of the estimator, with a rate of convergence $(\hat{d}_{m,n}^{GPH} - d) = O_p(m^{-1/2})$, can be obtained without knowledge of the distribution of the data generating process [Robinson (1995b), Hurvich *et al.* (1998)]. However, because the representation (1) of the spectral density holds only near the origin, a trade-off between the bias and variance of $\hat{d}_{m,n}^{GPH}$ must be made in choosing the optimal bandwidth m. Whereas Geweke and Porter-Hudak(1983) proposed $m = n^{1/2}$, which is still used in many applications, Hurvich *et al.* (1998) showed that a bandwidth $m = O(n^{4/5})$ is mean squared error (MSE)-optimal.

A second possibility for $\hat{d}$, proposed by Künsch (1987), is the so-called Gaussian semiparametric estimator (GSE) or local Whittle estimator

$$\hat{d}_{m,n}^{GSE} = \underset{d \in [-1/2 + \epsilon, 1/2 - \epsilon]}{\arg\min} \log\left(\frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_n(\lambda_j)\right) - \frac{2d}{m} \sum_{j=1}^m \log(\lambda_j),$$

involving a small fixed $\epsilon > 0$. Robinsion (1995a) established particularly mild conditions for the consistency of $\hat{d}_{m,n}^{GSE}$ at a rate $m^{-1/2}$ in probability. Compared to the GPH estimator that only requires a linear regression, the GSE is more complicated to compute but, under certain

conditions, its asymptotic variance $\mathrm{Var}(\hat{d}_{m,n}^{GSE}) \sim 1/(4m)$ is smaller relative to $\mathrm{Var}(\hat{d}_{m,n}^{GPH}) \sim \pi^2/(24m)$, as $n \to \infty$. The MSE-optimal bandwidth for $\hat{d}_{m,n}^{GSE}$ is also known to be $m = O(n^{4/5})$ [Andrews and Sun (2004)].

See Moulines and Soulier (2003) for further theoretical properties of $\hat{d}_{m,n}^{GPH}, \hat{d}_{m,n}^{GSE}$ and for other possible estimators of d.

4 Numerical studies

4.1 Coverage accuracy of one-sided EL bounds

We first conducted a simulation study of the finite-sample coverage accuracy of one-sided blockwise EL confidence bounds. The design and results of the study are next described.

4.1.1 Simulation parameters

Let $\{\tilde{Y}_t\}$, $t \in \mathbb{Z}$, represent a FARIMA$(0, d, 0)$ series

$$\tilde{Y}_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \varepsilon_{t-j}, \tag{14}$$

involving iid (mean zero) innovations $\{\varepsilon_t\}$, $t \in \mathbb{Z}$ and the gamma function $\Gamma(\cdot)$. To examine the performance of the EL method for several time series fulfilling (1), we considered FARIMA processes $Y_t = \varphi Y_{t-1} + \tilde{Y}_t + \vartheta \tilde{Y}_{t-1}$, $t \in \mathbb{Z}$, formed with one of the following ARMA filters, innovation distributions and d values:

- $\varphi = \vartheta = 0$ (Filter 1); $\varphi = -0.3$, $\vartheta = 0.1$ (Filter 2); $\varphi = 0.7$, $\vartheta = -0.3$ (Filter 3);

- $\{\varepsilon_t\}$ in (14) has a standard normal or $t-$distribution with 3 degrees of freedom;

- $d = -0.1, 0, 0.1, 0.25, 0.4$.

The processes $\{Y_t\}$ above fulfill (1) with various decay rates d based on innovations that may be Gaussian or non-Gaussian with heavy tails. Each process also has mean zero $\mathrm{E}(Y_t) = 0$.

We obtained size n time stretches $(Y_1, \ldots, Y_n)$ from each FARIMA series above by using a sample $\tilde{\mathcal{Y}}_n = (\tilde{Y}_1, \ldots, \tilde{Y}_n)$ from the process (14) as the innovations in the appropriate ARMA model. Gaussian samples $\tilde{\mathcal{Y}}_n$ were simulated by the circulant embedding method of Wood and Chan (1994) with FARIMA$(0, d, 0)$ covariances [see Beran (1994)], whereas non-Gaussian $\tilde{\mathcal{Y}}_n$ were generated by truncating the moving average expression in (14) after the first $M = 1000$

terms and then using $n + M$ independent $t-$variables to build an approximate truncated series [see Bardet *et al* (2003), p. 590 for details]. The sample sizes considered were $n = 100, 500, 1000$.

For each process and sample size, we calculated lower and upper approximate 95% one-sided confidence bounds with the blockwise EL method from (12) using overlapping blocks of length $\ell = C_1 n^{0.4}$, $C_1 \in \{1/2, 1, 2\}$ as well as the two estimators $\hat{d}_{m,n}^{GPH}$ and $\hat{d}_{m,n}^{GSE}$ of d from Section 3.3 with bandwidths $m = C_2 n^{0.8}$, $C_2 \in \{1/8, 1/4, 1/2, 1\}$. (All applications of the GSE were automated with a minimization over $|d| < 0.499$.) The order of the block length was selected as a compromise between size $O(n^\tau)$ blocks, $\tau \leq 1/3$, (optimal for subsampling and block bootstrap methods under SRD [Hall and Jing (1996)]) and size $O(n^{1/2})$ blocks chosen for subsampling methods under LRD [Hall, Jing and Lahiri (1998)]. (Repeating the simulation study with $\ell = C_1 n^{0.5}$ did not change the results significantly.) Coverage probabilities for the EL confidence bounds for the process mean $\mu = 0$ were approximated by an average over 1000 simulation runs for each process and sample size n.

4.1.2   Summary of results: one-sided EL bounds

It suffices to summarize numerical findings only for the upper confidence bounds, as the lower bounds closely matched the upper bounds in performance at each level combination in the study. Figures 1-3 display the coverage accuracy of upper 95% EL confidence bounds for the mean of processes based on t and Gaussian innovations; the graphics are differentiated by the estimator of d (GPH/GSE) used with the same two bandwidths $m = n^{0.8}/4$ or $n^{0.8}/2$.

Based on the results, we make some observations on the effects of the following factors:

- *d estimator:* There are generally few differences in the EL coverage accuracies across GSE and GPH estimators.

- *innovation type:* The EL intervals perform similarly for both t-distributed and Gaussian innovations, except for the case $d = 0.4$ where Gaussian innovations appear to have lower coverage probabilities. (See also Table 2 for this behavior.)

- *bandwidth m:* Performance differences in bandwidths $m = n^{0.8}/4$ or $n^{0.8}/2$ appear only with Filter 3, which introduces a high positive autoregressive component. For this filter, the EL confidence bounds become conservative when $m = n^{0.8}/2$ and perform better with a shorter bandwidth $m = n^{0.8}/4$. Because a process with a high positive autoregressive

coefficient can mimic long memory, a shorter bandwidth may help to refine estimation of the memory exponent d in this situation as the influence of the short memory component (most apparent for large frequencies) is down weighted by this strategy. For a discussion of the bias problem of the log-periodogram estimator in presence of additional short memory, see Davidson and Sibbertsen (2005). The EL coverage accuracies with bandwidths $m = C_2 n^{0.8}$, $C_2 \in \{1/2, 1\}$, were comparable throughout the study, as were $m = C_2 n^{0.8}$, $C_2 \in \{1/8, 1/4\}$.

- *block length ℓ:* The coverage probabilities for the blockwise EL bounds tend to decrease slightly as the block sizes increase. This behavior is most apparent with the smallest sample size $n = 100$ and also under the strongest forms of LRD $d = 0.4$. The smaller block sizes $\ell = C_1 n^{0.4}$, $C_1 \in \{1/2, 1\}$, seem to yield the most accurate EL intervals.

- *sample size n:* With the smallest sample size $n = 100$, the coverage accuracy of the EL bounds quickly deteriorates as the memory parameter d increases. This type of behavior is well-known in location estimation with strong LRD processes due to the persistence of the series [cf. Beran (1994)], so that a size $n = 100$ sample may be too small for inference in these cases. However, the coverage probabilities of EL intervals became closer to the nominal level as the sample sizes increase and the accuracy is often quite good even for a sample size of $n = 500$.

The simulation evidence here indicates that a block size $\ell = C_1 n^{0.4}$, $C_1 \in \{1/2, 1\}$, along with GPH estimator of d with bandwidth $m = n^{0.8}/4$, may be recommended for the coverage accuracy of the blockwise EL confidence bounds for the process mean. The performance of the EL method generally depends on the strength of the underlying dependence structure, but improves with increasing sample sizes.

It is presently unclear if an optimal block size ℓ exists for the blockwise EL intervals (12). Under weak dependence, optimal block lengths are available for implementing some block-based methods, such as the block bootstrap [cf. Lahiri (2003), Chapter 7]. However, more research on block lengths is required both with blockwise EL and with strong dependence.

4.2   Comparisons with other procedures: two-sided confidence intervals

We next compared the performance of the EL method to two other confidence interval procedures for the process mean μ: the normal approximation interval from (13) and the sampling

window (SW) method of Hall, Jing and Lahiri (1998).

For the EL method, we chose the GPH estimator of d with a bandwidth of $m = n^{0.8}/4$ and a block length $\ell = n^{0.4}/2$ or $n^{0.4}$. For the SW approach, we considered block lengths $\ell = n^{0.5}/2$ or $n^{0.5}$ with a smoothing parameter $\theta = 0.8$, as suggested by numerical studies in Hall, Jing and Lahiri (1998) and Nordman and Lahiri (2005). For the normal approximation intervals in (13), we used the GSE method for d and $\mathcal{C}_d$ which guarantees estimates in the parameter space (-1/2,1/2) of the long-memory parameter; the GPH estimator performed poorly in comparison with the normal approximation intervals, especially for processes involving $d = 0.4$ in our simulations. The bandwidth for the GSE was set to $m = n^{0.8}/4$, which seemed to produce the best overall intervals with the normal approximation.

Using iid standard normal and centered chi-square $\chi_1^2 - 1$ (with one degree of freedom) innovations $\{\varepsilon_t\}$ in (14), we simulated the FARIMA processes described in Section 4.1.1. For each of the three interval procedures above, we computed coverage probabilities for two-sided 90% confidence intervals as well as the average length of the intervals, based on 1000 simulation runs for each process and sample size $n = 500, 1000$. In addition, where Gaussian innovations were used, we considered two non-linear transformations:

$$\text{Filter 4:} \quad Y_{t,1}^3 \qquad\qquad \text{Filter 5:} \quad \sin(Y_{t,3})$$

using the Gaussian innovation-based FARIMA processes generated under filter 1 (denoted $Y_{t,1}$) and filter 3 (denoted $Y_{t,3}$). The sample mean $\bar{Y}_n$ also has a limiting normal distribution for these transformed Gaussian processes which have a Hermite rank of 1 [Taqqu (1975)]. The results with Gaussian innovations are presented in Tables 1-2 and chi-square innovations in Table 3.

To summarize our findings on the performance differences of the three methods:

- No one single confidence interval procedure emerges as clearly superior. The performances of the methods often depend on the process filter used as well as strength d of the process dependence.

- Generally, EL intervals tend to error on the side of anti-conservatism while SW intervals tend to be conservative. In many instances in Tables 1 and 3, the coverage inaccuracies of both EL and SW methods are of a similar degree, with the coverage probabilities for EL intervals falling below the nominal 90% level by an amount similar to which coverage probabilities of SW intervals fall above 90%. In the extreme cases in Table 1, the coverage

13

accuracy of EL intervals drop far below the nominal level for Processes 1 or 4 exhibiting the strongest forms of LRD ($d = 0.25, 0.4$), while the SW method performs much better; at the same time, the SW intervals are extremely conservative intervals for Processes 3 and 5, while the EL intervals perform better.

- Related to the point above, the expected length of two-sided EL intervals is usually much shorter than SW intervals (often half as long on average). This supports the differing behaviors in the coverage probabilities of both methods. However, in instances where both SW and EL methods exhibited similar coverage (see Table 3 for example), the EL intervals are still much shorter.

- From Tables 2 and 3, the normal approximation intervals generally tend to be much more anti-conservative than the EL method. In addition, as the long memory component d increases, the expected lengths of the normal approximation intervals explode, perhaps due to instabilities in directly substituting estimates of d into (13).

The EL method generally offers large improvements upon confidence intervals set by the normal approximation. The EL approach also appears to be competitive with the SW method for the linear processes considered here. A benefit of the EL approach, compared to the SW, seems to be much shorter confidence intervals. The trade-off is that EL intervals generally exhibit coverage inaccuracies in terms of undercoverage, as opposed to the more conservative SW method.

5 Proofs

For the proofs, we denote the positive integers as $\mathbb{N}$ and use C to denote a generic positive constant that does not depend on n.

5.1 Proof of Theorem 1(a)

When $d > 0$, the result is standard and follows from the covariance decay in (7) [Lemma A2 of Taqqu (1977)]. When $d = 0$, we use the nonnegative Fejer kernel $K_n(\lambda) = (2\pi n)^{-1} \sum_{j=-n}^{n}(n - |j|)e^{j\lambda\sqrt{-1}}$, $|\lambda| \leq \pi$, to write

$$n\mathrm{Var}(\bar{Y}_n) - 2\pi\mathcal{C}_0 = 2\pi \int_{-\pi}^{\pi} K_n(\lambda)\big[f(\lambda) - \mathcal{C}_0\big]d\lambda,$$

Table 1: Coverage probabilities of approximate two-sided 90% confidence intervals for $\mu = 0$ based on EL and SW methods for various block lengths ℓ; expected length of intervals denoted beneath. EL method uses GPH estimator with $m = n^{0.8}/4$ and $\ell = C_1 n^{0.4}$; SW method uses $\theta = 0.8$ and $\ell = C_1 n^{0.5}$. Filtered processes based on standard normal innovations, with filters 4, 5 based on transformations.

C_1	n	filter	$d = -0.1$ EL	$d = -0.1$ SW	$d = 0$ EL	$d = 0$ SW	$d = 0.1$ EL	$d = 0.1$ SW	$d = 0.25$ EL	$d = 0.25$ SW	$d = 0.4$ EL	$d = 0.4$ SW
1/2	500	1	83.3	96.3	84.8	96.5	83.9	95.0	83.5	92.9	79.4	89.5
			0.1	0.2	0.18	0.35	0.32	0.63	0.84	1.51	1.78	3.22
		3	93.7	99.5	93.2	98.5	93.3	99.1	94.6	97.8	92.1	94.3
			0.62	1.7	1.19	2.84	2.47	4.87	6.02	11.26	9.99	21.87
		4	82.1	93.6	84.0	95.6	83.3	92.5	77.6	86.9	67.3	78.3
			0.49	1.18	0.7	1.59	1.12	2.21	2.71	4.99	5.79	14.58
		5	88.4	99.8	90.5	99.2	86.6	99.3	85.7	99.1	90.9	100.0
			0.22	0.75	0.28	0.94	0.28	0.99	0.27	0.97	0.35	1.5
1/2	1000	1	84.3	95.9	85.9	97.0	86.4	96.4	83.8	95.5	83.3	90.2
			0.06	0.13	0.11	0.26	0.23	0.52	0.67	1.53	1.71	3.7
		3	92.2	99.6	93.8	99.2	94.5	98.7	95.9	97.9	95.6	95.8
			0.36	1.17	0.76	2.19	1.58	4.05	4.94	10.41	10.4	23.94
		4	84.9	97.2	84.4	96.6	85.3	96.0	81.4	89.2	68.1	84.1
			0.33	0.94	0.45	1.27	0.75	1.97	2.13	4.56	5.75	16.32
		5	90.4	99.7	92.6	99.1	89.4	99.3	85.0	98.7	91.3	99.6
			0.14	0.46	0.18	0.59	0.2	0.69	0.19	0.66	0.24	0.96
1	500	1	83.5	96.1	85.2	95.4	85.1	93.8	83.7	94.8	77.9	88.9
			0.09	0.19	0.17	0.35	0.31	0.64	0.78	1.63	1.54	3.43
		3	92.5	99.5	94.2	99.0	93.7	97.1	93.1	96.8	88.7	92.7
			0.63	1.48	1.17	2.6	2.23	4.52	5.24	10.56	8.43	19.61
		4	83.2	96.4	82.5	94.4	80.0	93.2	79.1	87.6	63.3	82.5
			0.49	1.2	0.65	1.55	1.01	2.26	2.33	4.83	4.66	14.21
		5	91.6	98.1	90.4	98.9	87.7	98.0	88.2	97.5	91.6	99.4
			0.23	0.57	0.28	0.73	0.28	0.77	0.27	0.76	0.35	1.1
1	1000	1	87.0	94.8	87.4	93.2	84.8	92.1	85.1	91.1	82.3	85.6
			0.06	0.11	0.11	0.21	0.22	0.43	0.65	1.23	1.57	2.91
		3	93.4	97.8	95.4	96.9	95.8	97.3	96.0	95.5	91.6	90.5
			0.37	0.78	0.74	1.43	1.56	2.73	4.51	6.95	9	15.84
		4	85.9	93.1	86.7	93.3	83.9	93.2	80.9	85.8	69.6	77.9
			0.32	0.62	0.45	0.86	0.71	1.48	1.92	3.23	4.89	10.07
		5	92.0	96.9	92.7	98.0	88.1	97.5	87.5	96.6	91.9	97.1
			0.14	0.32	0.18	0.4	0.2	0.44	0.19	0.43	0.24	0.57

Table 2: Coverage probabilities of approximate two-sided 90% confidence intervals for $\mu = 0$ based on normal approximation method (using GSE with $m = n^{0.8}/4$); expected length of intervals denoted beneath. Filtered processes based on standard normal innovations.

		$n = 500$					$n = 1000$				
	d	-0.1	0	0.1	0.25	0.4	-0.1	0	0.1	0.25	0.4
	1	76.7	75.9	77.8	77.1	74.3	77.9	77.4	74.9	76.9	79.2
		0.07	0.12	0.23	1.14	13.32	0.04	0.08	0.16	0.47	8.03
	3	87.6	92.8	91.3	93.3	95.6	89.1	90.7	90.9	94.0	96.6
filter		0.47	0.93	3.29	52.19	191.88	0.29	0.56	1.21	13.99	181.87
	4	76	75.3	71.6	70.0	69.6	76.4	75.9	76.5	71.8	67.9
		0.35	0.47	0.68	2.76	20.68	0.23	0.32	0.51	1.21	8.33
	5	83.9	84.1	81.1	78.5	85.1	87.2	85.5	82.9	79.7	86.5
		0.17	0.21	0.22	0.20	0.28	0.11	0.14	0.15	0.14	0.18

Table 3: Coverage probabilities of approximate two-sided 90% confidence intervals for $\mu = 0$ based on EL, SW and normal approximation methods; expected length of intervals denoted beneath. EL method uses GPH estimator with $m = n^{0.8}/4$ and $\ell = n^{0.4}$; SW method uses $\theta = 0.8$ and $\ell = n^{0.5}$; normal approximation intervals use the GSE with $m = n^{0.8}/4$. Filtered processes based on centered chi-square $\chi_1^2 - 1$ innovations.

n	filter	$d = 0$			$d = 0.25$			$d = 0.4$		
		NORM	EL	SUB	NORM	EL	SUB	NORM	EL	SUB
	1	77.0	83.6	95.0	77.2	84.8	95.4	82.7	85.6	94.8
		0.17	0.23	0.59	1.2	1.11	2.64	19.41	2.24	5.43
500	2	75.8	84.9	94.9	75.2	81.8	94.5	84.2	85.9	92.6
		0.14	0.2	0.47	1.26	0.92	2.21	15.83	1.88	4.55
	3	91.1	94.1	97.9	93.1	95.0	97.6	98.1	94.1	95.4
		1.63	1.7	4.17	73.37	7.66	16.21	281.17	12.27	30.69
	1	77.4	87.3	94.9	81.1	86.7	94.9	88.1	91.6	94.5
		0.12	0.16	0.35	0.68	0.94	2.02	12.47	2.27	4.72
1000	2	78.7	86.4	93.8	79.7	86.9	94.5	88.0	92.1	93.6
		0.1	0.14	0.27	0.56	0.78	1.61	10.25	1.92	3.9
	3	90.7	95.5	97.6	95.3	96.3	96.9	98.7	98.1	95.2
		0.83	1.07	2.29	18.72	6.18	11.11	254.14	13.1	24.65

since $\int_{-\pi}^{\pi} K_n(\lambda)d\lambda = 1$. For a fixed $0 < \delta \leq \pi$, it holds that $|f(\lambda) - \mathcal{C}_0|$ is bounded on $\delta \leq |\lambda| \leq \pi$ and $\int_{\delta \leq |\lambda| \leq \pi} K_n(\lambda)d\lambda \to 0$ as $n \to \infty$ [Brockwell and Davis (1987), Section 2.11] so that

$$\limsup_{n \to \infty} |n\text{Var}(\bar{Y}_n) - 2\pi\mathcal{C}_0| \leq 2\pi \sup_{0 < |\lambda| \leq \delta} |f(\lambda) - \mathcal{C}_0|. \tag{15}$$

Letting $\delta \to 0$, the righthand side above converges to zero by (1), establishing the case $d = 0$. When $d < 0$, we use (7) and $\sum_{k \in \mathbb{Z}} r(k) = 0$ to deduce that

$$A_{1n} \equiv \sum_{k=-n}^{n} r(k) = -2 \sum_{k=n+1}^{\infty} r(k) \sim \frac{\mathcal{C}_d \mathcal{R}_d}{d} n^{2d}, \quad A_{2n} \equiv 2n^{-1} \sum_{k=1}^{n} kr(k) \sim \frac{2\mathcal{C}_d \mathcal{R}_d}{1 + 2d} n^{2d},$$

as $n \to \infty$ by Lemma A2 of Taqqu (1977). Since $n\text{Var}(\bar{Y}_n) = A_{1n} - A_{2n}$, Theorem 1(a) now follows for $d < 0$. $\square$

5.2 Proof of Theorems 2 and 3

To prove Theorem 2, we require some preliminary results. For LRD/SRD linear processes with *bounded* innovations, Proposition 1 implies that block means at distant lags are asymptotically uncorrelated, based on a condition suggested by Hall, Jing and Lahiri (1998), Theorem 2.4.

Proposition 1 *Suppose $\ell^{-1} + \ell/n = o(1)$ and $\{Y_t\}, t \in \mathbb{Z}$, satisfies the conditions of Theorem 1(b) with bounded innovations, i.e., $P(|\epsilon_t| \leq C) = 1$ for some $C > 0$. Then, for any integers $a, b \in \mathbb{N} \cup \{0\}$ and $0 < \epsilon < 1$,*

$$\max_{n\epsilon \leq i \leq n} \left| \text{E}\left[a_\ell^{-a}(M_{\ell 1} - \mu)^a \cdot a_\ell^{-b}(M_{\ell i} - \mu)^b\right] - \text{E}(Z^a) \cdot \text{E}(Z^b) \right| = o(1), \quad \text{as } n \to \infty,$$

where $M_{\ell i} = \sum_{j=i}^{i+\ell-1} Y_j/\ell$, $i \geq 1$, $\text{E}(Y_t) = \mu$, and Z is a standard normal variable.

Proof of Proposition 1. The case $d > 0$ is treated in Lemma 2 of Nordman and Lahiri (2005) which requires $\text{Var}(\bar{Y}_n)/\text{Var}(\bar{Y}_\ell) = (\ell/n)^{1-2d} = o(1)$ and $n^2\text{Var}(\bar{Y}_n) \to \infty$. Using the same arguments, the cases $d = 0$ and $d < 0$ follow with a modification to show (for the $a = b = 1$ situation) that $\Delta_{n,\epsilon} \equiv \max_{n\epsilon \leq i \leq n} |\text{Cov}(M_{\ell 1}, M_{\ell i})|/a_\ell^2 = o(1)$ holds for a fixed $0 < \epsilon < 1$. When $|d| > 0$, we may use (7) and Theorem 1(a) to show

$$\Delta_{n,\epsilon} \leq C\ell^{1-2d} \max_{n\epsilon \leq i \leq n} \max_{|j| \leq \ell} |r(i+j)| = O\left((\ell/n)^{1-2d}\right) = o(1).$$

17

When $d = 0$ and $a_\ell^2 = 2\pi\mathcal{C}_0\ell^{-1}$ by Theorem 1(a), we have with the Fejer kernel K_ℓ that

$$\frac{\text{Cov}(M_{\ell 1}, M_{\ell i})}{a_\ell^2} = \frac{1}{\mathcal{C}_0} \int_{-\pi}^{\pi} K_\ell(\lambda) \cdot e^{(i-1)\lambda\sqrt{-1}} [f(\lambda) - \mathcal{C}_0] d\lambda, \qquad i > \ell,$$

holds since $\int_{-\pi}^{\pi} K_\ell(\lambda) \cdot e^{(i-1)\lambda\sqrt{-1}} d\lambda = 0$ for $i > \ell$; then, $\Delta_{n,\epsilon} = o(1)$ follows as $\ell \to \infty$ from the same argument as in (15) using $|e^{(i-1)\lambda\sqrt{-1}}| = 1$, $|\lambda| \leq \pi$, $i \in \mathbb{N}$. $\square$

Proposition 2 *Suppose $\ell^{-1} + \ell/n = o(1)$ and $\{Y_t\}, t \in \mathbb{Z}$, satisfies Theorem 1(b) conditions with $E(Y_t) = \mu$. Let $\bar{M}_{n\mu} = N^{-1}\sum_{i=1}^{N}(M_{\ell i} - \mu)$ and $\hat{a}_{\ell\mu}^2 = N^{-1}\sum_{i=1}^{N}(M_{\ell i} - \mu)^2$. As $n \to \infty$,*
(a) $\bar{M}_{n\mu}/a_n \xrightarrow{d} Z$, a standard normal variable;
(b) $\hat{a}_{\ell\mu}^2/a_\ell^2 \xrightarrow{p} 1$;
(c) $P\left(\min_{1\leq i \leq N} M_{\ell i} < \mu < \max_{1\leq i \leq N} M_{\ell i}\right) \to 1$.

Proof of Proposition 2. Noting that the first and last ℓ observations in $(Y_1, \ldots, Y_n)$ contribute less to $\bar{M}_{n\mu}$, we may write $N\bar{M}_{n\mu} = n(\bar{Y}_n - \mu) - H_n$ where $\ell H_n = \sum_{j=1}^{\ell}(\ell - j)(Y_j + Y_{n-j+1} - 2\mu) = \sum_{i=1}^{\ell-1}\sum_{j=1}^{i}(Y_j + Y_{n-j+1} - 2\mu)$. Using Holder's inequality and $n^2a_n^2 = n^{1+2d}\mathcal{V}_d$, $n \in \mathbb{N}$, we find

$$\text{E}(H_n^2) \leq \frac{4}{\ell}\sum_{i=1}^{\ell-1}\text{Var}\left(\sum_{j=1}^{i}Y_j\right) = O\left(\frac{1}{\ell}\sum_{i=1}^{\ell-1}i^2a_i^2\right) = O\left(\ell^{1+2d}\right) = o(n^2a_n^2)$$

so that $(na_n)^{-1}H_n \xrightarrow{p} 0$ follows. Then applying Slutsky's theorem with Theorem 1(b) and $n/N \to 1$, we have the distributional result for $\bar{M}_{n\mu}/a_n$ in Proposition 2(a).

To prove Proposition 2(b), let $\text{E}(\varepsilon_t^2) = \sigma^2$ and denote the indicator function as $I\{\cdot\}$. For each $c \in \mathbb{N}$ and $t \in \mathbb{Z}$, define variables $\varepsilon_{t,c} = \varepsilon_t I\{|\varepsilon_t| \leq c\} - \text{E}(\varepsilon_t I\{|\varepsilon_t| \leq c\})$ and $Y_{t,c} = \mu + \sum_{j\in\mathbb{Z}} b_j(\sigma\varepsilon_{t-j,c}/\sigma_c)$, where w.l.o.g. $\sigma_c^2 = \text{E}(\varepsilon_{t,c}^2) > 0$. For each $n, c, i \in \mathbb{N}$, write analogs $M_{\ell i,c} = \sum_{j=i}^{i+\ell-1}Y_{j,c}/\ell$ and $\hat{a}_{\ell\mu,c}^2 = N^{-1}\sum_{i=1}^{N}(M_{\ell i,c} - \mu)^2$ with respect to $Y_{1,c}, \ldots, Y_{n,c}$. The processes $\{Y_t\}$ and $\{Y_{t,c}\}$ have identical means μ and covariances because each series involves the same linear filter with iid innovations of mean 0 and variance σ^2. Since $\text{E}|\hat{a}_{\ell\mu}^2 - \hat{a}_{\ell\mu,c}^2| \leq \text{E}|\{(M_{\ell 1} - \mu) + (M_{\ell 1,c} - \mu)\}(M_{\ell 1} - M_{\ell 1,c})|$ holds, we may apply Holder's inequality to show that for each $c \in \mathbb{N}$,

$$\text{E}|\hat{a}_{\ell\mu}^2 - \hat{a}_{\ell\mu,c}^2| \leq 4\text{Var}(\bar{Y}_\ell)\{1 - \sigma_c^{-1}\sigma^{-1}\text{E}(\varepsilon_t\varepsilon_{t,c})\}^{1/2}, \tag{16}$$

using $\text{Var}(\bar{Y}_\ell) = \text{Var}(M_{\ell 1}) = \text{Var}(M_{\ell 1,c})$ and $\text{Var}(M_{\ell 1} - M_{\ell 1,c}) = \text{Var}(\bar{Y}_\ell) \cdot \text{Var}(\sigma^{-1}\varepsilon_t - \sigma_c^{-1}\varepsilon_{t,c})$ by the iid property of the innovations. For each $c \in \mathbb{N}$, we find that $\text{E}(\hat{a}_{\ell\mu,c}^2) = \text{Var}(\bar{Y}_\ell) \sim a_\ell^2$ and,

through applying Proposition 1, that $\text{Var}(\hat{a}_{\ell\mu,c}^2) = o(a_\ell^4)$ follows as in the proof of Theorem 2.5 of Hall, Jing and Lahiri (1998). From this and (16), we deduce that

$$\limsup_{n\to\infty} \text{E}\left|\frac{\hat{a}_{\ell\mu}^2}{a_\ell^2} - 1\right| \leq \limsup_{n\to\infty} \frac{1}{a_\ell^2}\left\{\text{E}|\hat{a}_{\ell\mu}^2 - \hat{a}_{\ell\mu,c}^2| + \left[\text{Var}(\hat{a}_{\ell\mu,c}^2)\right]^{1/2} + |\text{E}(\hat{a}_{\ell\mu,c}^2) - a_\ell^2|\right\}$$

$$\leq 4\{1 - \sigma_c^{-1}\sigma^{-1}\text{E}(\varepsilon_t\varepsilon_{t,c})\}^{1/2}$$

for each $c \in \mathbb{N}$. Because $\lim_{c\to\infty}\sigma_c^{-1}\text{E}(\varepsilon_t\varepsilon_{t,c}) = \sigma$, $\hat{a}_{\ell\mu}^2/a_\ell^2 \xrightarrow{p} 1$ now follows in Proposition 2(b).

For $x \in \mathbb{R}$, let $\Phi(x) = P(Z \leq x)$ denote the distribution function of a standard normal Z and define $\widehat{\Phi}_n(x) = N^{-1}\sum_{i=1}^N\{(M_{\ell i} - \mu)/a_\ell \leq x\}$. Fix $\delta > 0$. To show Proposition 2(c), it suffices to prove that if $\delta' = \delta$ or $-\delta$, then $\widehat{\Phi}_n(\delta') \xrightarrow{p} \Phi(\delta')$ as $n \to \infty$, since $0 < \Phi(\pm\delta) < 1$. Using Proposition 1, it can be shown that $\lim_{n\to\infty}\text{E}|\widehat{\Phi}_n(\delta') - \Phi(\delta')| = 0$, for $\delta' = \pm\delta$, following the proof of Theorem 1 of Nordman and Lahiri (2005). □

Proof of Theorem 2. Applying Lemma 2(c) with $\mu = \mu_0$, we find $P(R_n(\mu_0) > 0) \to 1$ as $n \to \infty$ so that, with probability approaching 1 as $n \to \infty$, $R_n(\mu_0)$ can be written as in (5) with probabilities $p_i = N^{-1}\{1 + \lambda_{\mu_0}(M_{\ell i} - \mu_0)\}^{-1}$, $1 \leq i \leq N$, where $g_{\mu_0}(\lambda_{\mu_0}) = 0$ in (4).

Let $Z_{n\mu_0} = \max_{1 \leq i \leq N}|M_{\ell i} - \mu_0|$. Since $\text{Var}(M_{\ell 1}) \sim a_\ell^2 = \ell^{-1+2d}\mathcal{V}_d$, we find that

$$a_\ell^{-2}a_n \cdot Z_{n\mu_0} = o_p(1), \tag{17}$$

using (8) by assumption. We remark that, when $\ell^2/n = o(1)$ and $\text{E}|Y_t|^q < \infty$ for an even integer $q \geq 4/(1-2d)$, the random variable in (8) is bounded by

$$\left(\frac{\ell}{n}\right)^{1/2-d} n^{1/q}\left(\frac{1}{n}\sum_{i=1}^N \frac{(M_{\ell 1} - \mu_0)^q}{\{\text{Var}(M_{\ell 1})\}^{q/2}}\right)^{1/q} = o(n^{-1/4+d/2+1/p})O_p(1) = o_p(1),$$

where $O_p(1)$ follows for the sum by using $\text{E}(M_{\ell 1} - \mu_0)^q \leq C\{\text{Var}(M_{\ell 1})\}^{q/2}$ (by Assumption (A.2) and lemma 4, Davydov (1970)) with Markov's inequality.

Let $\bar{M}_{n\mu_0} = N^{-1}\sum_{i=1}^N(M_{\ell i} - \mu_0)$ and $\hat{a}_{\ell\mu_0}^2 = N^{-1}\sum_{i=1}^N(M_{\ell i} - \mu_0)^2$. Similar to step (2.12) of Owen (1990), it follows from $g_{\mu_0}(\lambda_{\mu_0}) = 0$ in (4) that

$$|\lambda_{\mu_0}| = O_p(a_n/a_\ell^2), \tag{18}$$

using (17) with $|\bar{M}_{n\mu_0}|/\hat{a}_{\ell\mu_0}^2 = O_p(a_n/a_\ell^2)$ from Proposition 2.

Define $\theta_i = \lambda_{\mu_0}(M_{\ell i} - \mu_0)$ for $1 \leq i \leq N$. The equation $0 = g_{\mu_0}(\lambda_{\mu_0})$ can be solved for $\lambda_{\mu_0} = (\bar{M}_{n\mu_0} + I_{n\mu_0})/\hat{a}_{\ell\mu_0}^2$, where $I_{n\mu_0} = N^{-1}\sum_{i=1}^N \theta_i^2(M_{\ell i} - \mu_0)/(1 + \theta_i)$ satisfies

$$|I_{n\mu_0}| \leq |\lambda_{\mu_0}|^2 Z_{n\mu_0}\hat{a}_{\ell\mu_0}^2 \max_{1 \leq i \leq N}|1 + \theta_i|^{-1} = o_p(a_n),$$

19

by (17)-(18), Proposition 2(b), and $\max_{1 \le i \le N} |\theta_i| \le |\lambda_{\mu_0}| Z_{n\mu_0} = o_p(1)$. When $|\lambda_{\mu_0}| Z_{n\mu_0} < 1$, a Taylor expansion gives $\log(1 + \theta_i) = \theta_i - \theta_i^2/2 + \nu_i$ for each $1 \le i \le N$ where

$$|\nu_i| \;\le\; |\lambda_{\mu_0}|^3 Z_{n\mu_0} (M_{\ell i} - \mu_0)^2 (1 - |\lambda_{\mu_0}| Z_{n\mu_0})^{-3}. \tag{19}$$

Using this expansion of $\log(1 + \theta_i)$, $1 \le i \le N$, in (5) along with the expression for λ_{μ_0} and $B_n = N^{-1} a_n^{-2} a_\ell^2$, we now write

$$-2B_n \log R_n(\mu_0) \;=\; Q_{1n} - Q_{2n} + Q_{3n},$$

where $Q_{1n} = (a_n^{-1} \bar{M}_{n\mu_0})^2 / (a_\ell^{-2} \hat{a}_{\ell\mu_0}^2) \xrightarrow{d} \chi_1^2$ by Proposition 2; $Q_{2n} = (a_n^{-1} I_{n\mu_0})^2 / (a_\ell^{-2} \hat{a}_{\ell\mu_0}^2) = o_p(1)$ applying $|I_{n\mu_0}| = o_p(a_n)$; and for $Q_{3n} = -2B_n \sum_{i=1}^{N} \nu_i$, we may bound

$$|Q_{3n}| \;\le\; B_n \sum_{i=1}^{N} |\nu_i| \;\le\; 2 a_\ell^2 a_n^{-2} |\lambda_{\mu_0}|^3 Z_{n\mu_0} \hat{a}_{\ell\mu_0}^2 (1 - |\lambda_{\mu_0}| Z_{n\mu_0})^{-3} \;=\; o_p(1)$$

by (19). Theorem 2 follows now by applying Slutsky's theorem. □

Proof of Theorem 3. From Theorem 2 and $\ell = O(n)$ by assumption, it suffices to show $\widehat{B}_n/B_n = (n/\ell)^{2(d-\hat{d}_n)} \xrightarrow{p} 1$, which follows from $|d - \hat{d}_n| \log(n) \xrightarrow{p} 0$. □

Acknowledgments

References

ADENSTEDT, R. K. (1974) On large-sample estimation for the mean of a stationary sequence. *Ann. Statist.* 2, 1095-1107.

ANDREWS D. W. K. and SUN, Y. (2004) Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* 72, 569 - 614.

BARDET, J., LANG, G., OPPENHEIM, G., PHILIPPE, S. and TAQQU, M. S. (2003) Generators of long-range dependent processes: a survey. In P. Doukhan, G. Oppenheim and M. S. Taqqu (eds.), *Theory and Applications of Long-range Dependence*, pp. 579-624. Birkhäuser, Boston.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987) *Regular variation.* University Press, Cambridge.

BERAN, J. (1994) *Statistical methods for long memory processes.* Chapman & Hall, London.

BROCKWELL, P. J. and DAVIS, R. A. (1987) *Time series: theory and methods.* Springer-Verlag, New York.

CARLSTEIN, E. (1986) The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.* 14, 1171-1179.

DAVIDSON, J. and SIBBERTSEN, P. (2005) Tests of Bias in Log-Periodogram Regression. *Discussion Paper 317/2005, University of Hannover.*

DAVYDOV, Y. A. (1970) The invariance principle for stationary processes. *Theor. Prob. Appl.* 15, 487-498.

GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long-memory time series models. *J. Time Series Anal.* 4, 221-238.

GRANGER, C. W. J. and JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* 1, 15-29.

HALL, P. and LA SCALA, B. (1990) Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* 58, 109-127.

HALL, P. and JING, B.-Y. (1996) On sample reuse methods for dependent data. *J. R. Stat. Soc., Ser. B* 58, 727-737.

HALL, P., JING, B.-Y. and LAHIRI, S. N. (1998) On the sampling window method for long-range dependent data. *Statist. Sinica* 8, 1189-1204.

HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165-176.

HURVICH, C. M., DEO, R. and BRODSKY, J. (1998) The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *J. Time Series Anal.* 19, 19-46.

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971) *Independent and stationary sequences of random variables.* Wolters-Noordhoff, Groningen.

KITAMURA, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* 25, 2084-2102.

KÜNSCH, H. R. (1987) Statistical aspects of self-similar processes. *Probability Theory and Applications* 1, 67-74.

KÜNSCH, H. R. (1989) The jackknife and bootstrap for general stationary observations. *Ann. Statist.* 17, 1217-1261.

LAHIRI, S. N. (1993) On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* 18, 405-413.

LAHIRI, S. N. (2003) *Resampling methods for dependent data.* Springer, New York.

LIU, R. Y. and SINGH, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap,* 225-248, R. LePage and L. Billard (editors). John Wiley & Sons, New York.

MANDELBROT, B. B. and VAN NESS, J. W. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* 10, 422-437.

MOULINES, E. and SOULIER, P. (2003) Semiparametric spectral estimation for fractional processes. In P. Doukhan, G. Oppenheim and M. S. Taqqu (eds.), *Theory and Applications of Long-range Dependence*, pp. 251-301. Birkhäuser, Boston.

NORDMAN, D. J. and LAHIRI, S. N. (2005) Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory* 21, 1087-1111.

OWEN, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237-249.

OWEN, A. B. (1990) Empirical likelihood confidence regions. *Ann. Statist.* 18, 90-120.

OWEN, A. B. (2001) *Empirical likelihood.* Chapman & Hall, London.

POLITIS, D. N. and ROMANO, J. P. (1993) On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes Appl.* 45, 155-167

ROBINSON, P.M. (1995a) Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* 23, 1630-1661.

ROBINSON, P.M. (1995b) Log-periodogram regression of time series with long range dependence. *Ann. Statist.* 23, 1048-1072.

QIN, J. and LAWLESS, J. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.* 22, 300-325.

TAQQU, M. S. (1975). Weak convergence to fractional brownian motion and to the Rosenblatt process. *Z. Wahr. Verw. Gebiete* 31, 287-302.

TAQQU, M. S. (1977) Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Z. Wahr. Verw. Gebiete* 40, 203-238.

WILKS, S. S. (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* 9, 60-62.

WOOD, A. T. A. and CHAN, G. (1994) Simulation of stationary Gaussian processes in $[0,1]^d$. *J. Comput. Graph. Statist.* 3, 409-432.

ZYGMUND, A. (1968) *Trigonometric Series* Volumes I, II. Cambridge University Press, Cambridge, U.K.
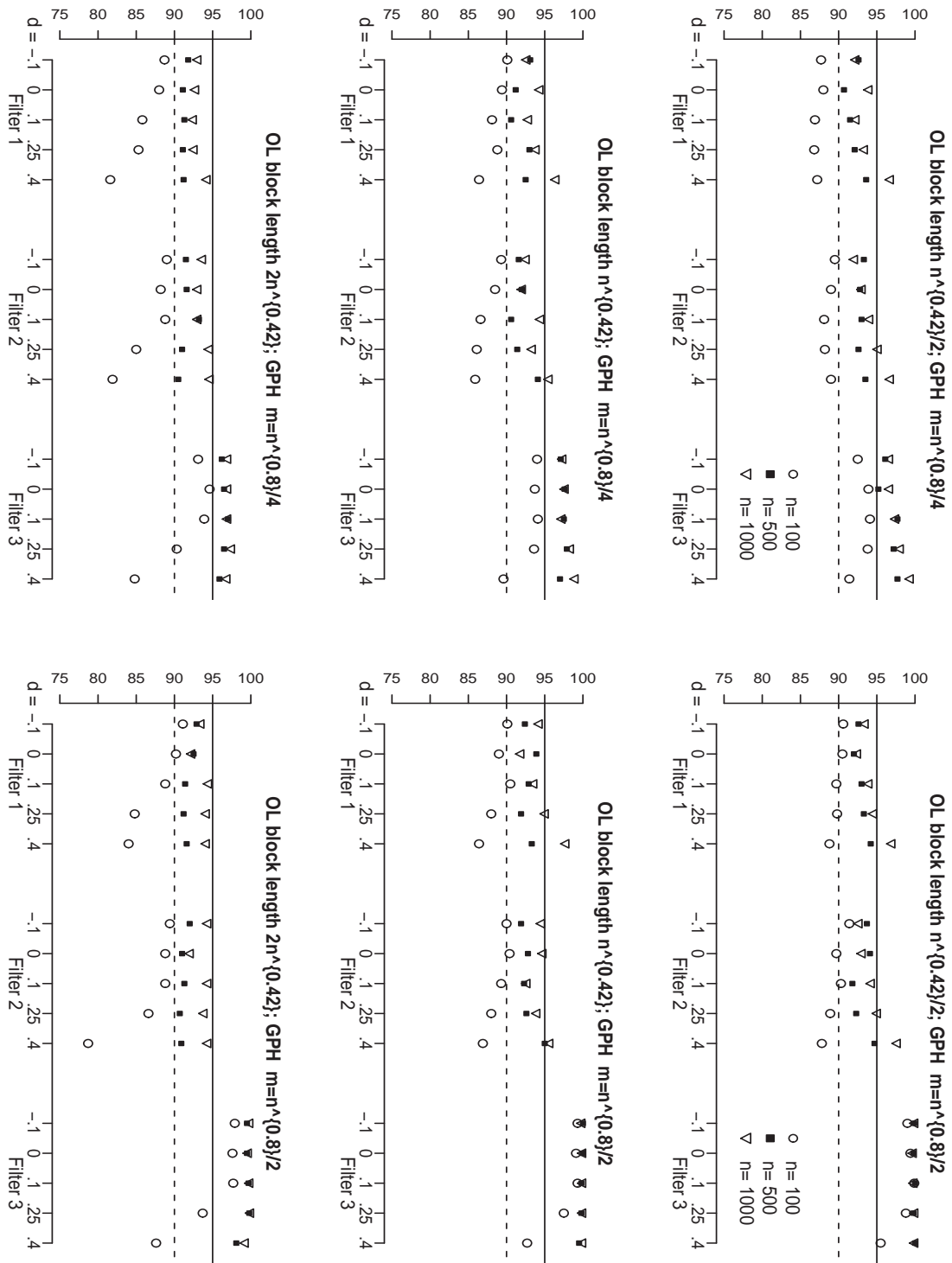
Figure 1: *t-innovations:* Coverage percentages for approximate 95% one-sided upper EL confidence bounds using GPH estimator with bandwidth $m = C_2 n^{0.8}$, $C_2 \in \{1/4, 1/2\}$.
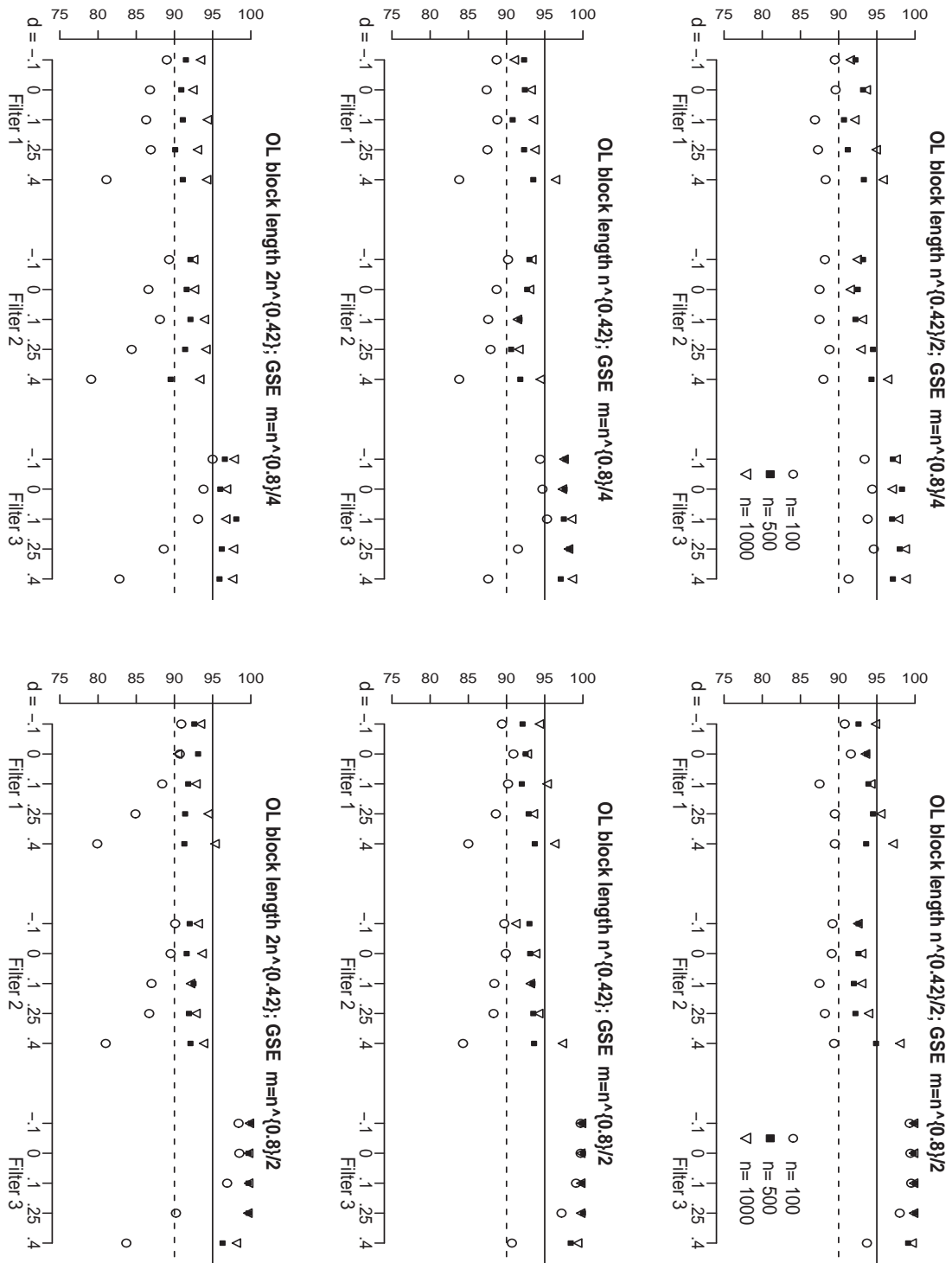
Figure 2: *t-innovations:* Coverage percentages for approximate 95% one-sided upper EL confidence bounds using GSE with bandwidth $m = C_2 n^{0.8}$, $C_2 \in \{1/4, 1/2\}$.
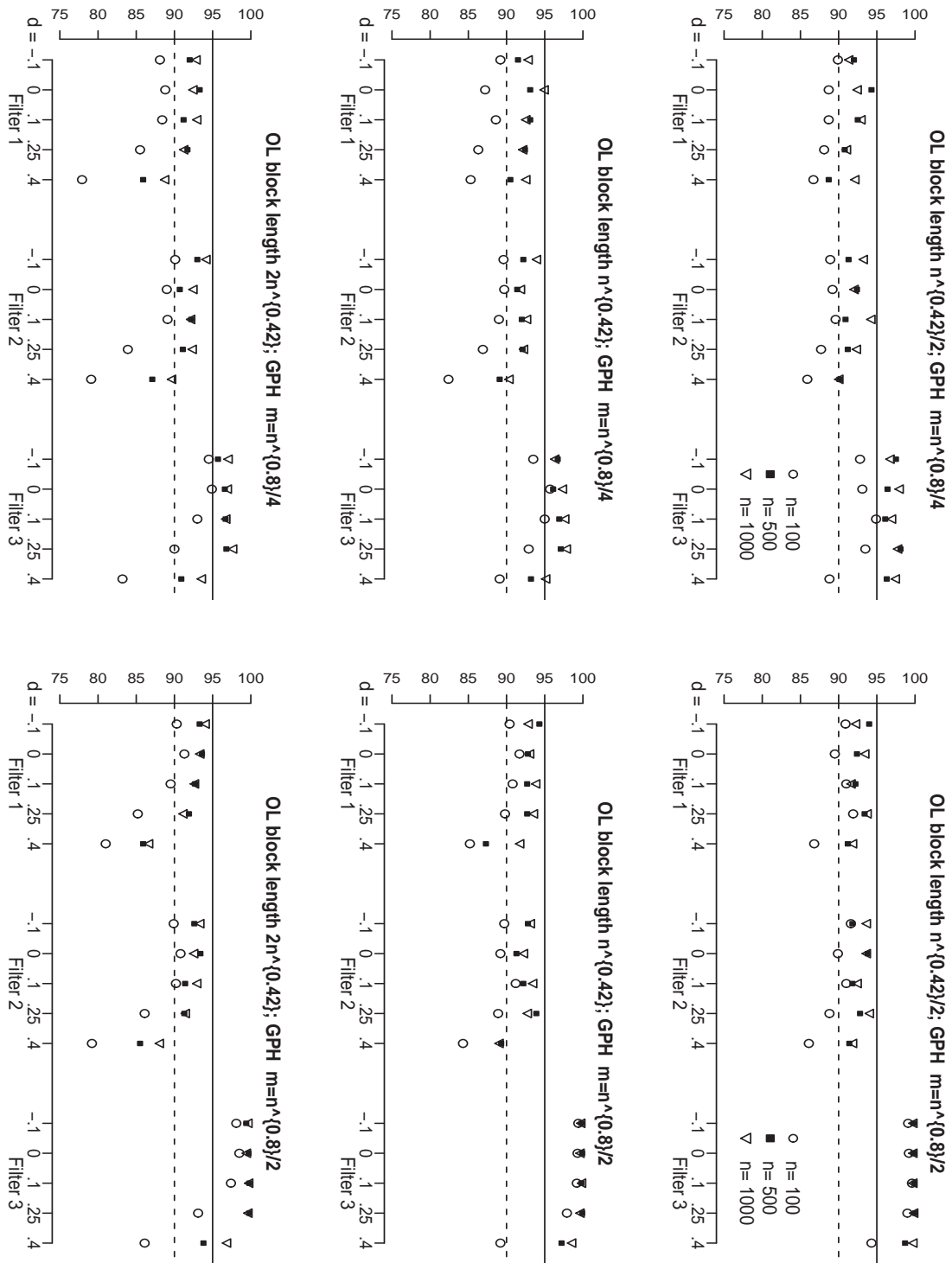
Figure 3: *Gaussian-innovations:* Coverage percentages for approximate 95% one-sided upper EL confidence bounds using GPH estimator with bandwidth $m = C_2 n^{0.8}$, $C_2 \in \{1/4, 1/2\}$.