# Filtered Log-periodogram Regression of long memory processes

Yuanhua Feng[1] and Jan Beran[2]

[1]Faculty of Business Administration and Economics
University of Paderborn
D-33098 Paderborn, Germany
Email:yuanhua.feng@wiwi.upb.de

[2]Department of Mathematics and Statistics
University of Konstanz
D-78457 Konstanz, Germany
Email: J.Beran@uni-konstanz.de

November 2008

**Abstract**

Filtered log-periodogram regression estimation of the fractional differencing parameter $d$ is considered. Asymptotic properties are derived and the effect of filtering on $\hat{d}$ is investigated. It is shown that the estimator by Geweke and Porter-Hudak (1983) can be improved significantly using a simple family of filters. The essential improvement is based on a binary decision that is asymptotically correct with probability one. The idea is closely related to the well known technique of pre-whitening.

*AMS Subject Classification:* 62M10, 62M15

*Keywords.* Filtering, log-periodogram regression, local pre-whitening, fractional differencing parameter, long memory.

# 1 Introduction

Consider estimation of the fractional differencing parameter $d$ for a Gaussian process $X_t$ with spectral density

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_u(\lambda), \tag{1.1}$$

where $d \in [-0.5, 0.5)$, and $f_u(\cdot)$ is a symmetric, positive, continuous function on $[-\pi, \pi]$. Note that $f_u(\cdot)$ is the spectral density of the short-memory process $u_t = (1 - B)^d X_t$. Maximum likelihood estimation of $d$ (Fox and Taqqu 1986, Yajima 1985, Giraitis and Surgailis 1990, Dahlhaus 1989, Beran 1994, 1995) requires knowledge of the model or, if combined with model choice, of the model class (see e.g. Beran et al. 1999). In contrast, semiparametric procedures do not require full specification of $f_u$. A well known semiparametric method is, for instance, the so-called GPH estimator, originally proposed by Geweke and Porter-Hudak (1993). It essentially consists of least squares regression of the log-periodogram versus $\log \lambda$, using small frequencies. It is well known that the GPH estimator does not have ideal finite sample properties due to the dependence of the periodogram ordinates at low frequencies (Künsch, 1986, Hurvich and Beltrao 1993, Robinson, 1995) and a large bias caused by the *relative curvature* of $f_u$ at the origin,

$$\beta_u =: f_u''(0)/f_u(0)$$

(Hurvich et al. 1998). Numerous improvements of the GPH estimator are proposed in the literature. Robinson (1994) proposes a semiparametric average periodogram estimator and Robinson (1995) suggests to exclude a few periodogram ordinates at low frequencies. Giraitis et al. (1997) discuss how to improve the rate of convergence of the GPH estimator. Other versions of this estimator may be found e.g. in Hurvich and Beltrao (1994) and Velasco (1999, 2000). Recently, Shimotsu and Phillips (2002) proposed a pooled GPH estimator to allow the use of a larger number of periodogram ordinates, thus reducing the variance without significant changes of the bias.

In this paper, we propose to improve the performance of the GPH estimator from a new point of view. It is shown that the performance of the GPH estimator can be improved significantly, using a simple family of filters. The essential improvement is based on a binary decision that is asymptotically correct with probability one. More specifically, note that $\beta_u$ does not depend on $d$ and measures, for the short-memory process $u_t$, the *strength of dependence at low frequencies*. If $u_t$ is white noise, then $\beta_u = 0$. On the other hand, $\beta_u = 0$ implies $f_u(0) \neq 0$ and $f_u'(0) = f_u''(0) = 0$. In view of Condition 2 given later, this means that $f_u(\lambda)$ is very flat at

$\lambda = 0$ and the behavior of $u_t$ at low frequencies is quite similar to white noise. The asymptotic performance of the GPH estimator is mainly determined by $\beta_u$ (Hurvich et al. 1998). When $|\beta_u|$ is large, only a very small number of periodogram ordinates can be used to achieve an optimal trade-off between bias and variance, and the resulting mean squared error is quite large. The intuitive reason is that, if $\beta_u$ assumes a large positive value, then $u_t$ ressembles a long-memory process, whereas for extreme negative values of $\beta_u$, $u_t$ is similar to an antipersistent process. The question thus arises whether it is possible to reduce $|\beta_u|$ before estimating $d$, while avoiding full estimation of $f_u$ or its derivatives. The key idea is to achieve a considerable reduction of $|\beta_u|$, without changing $d$, by applying a simple time invariant linear filter. This approach is closely related to the well known pre-whitening technique introduced by Blackman and Tukey (1959) (see also Priestley, 1981, p.556). Here only local pre-whitening is required, in order to whiten the short-memory part at low frequencies. It is therefore not difficult to choose a simple class of filters that improves the performance of the original GPH estimator. It is also worth mentioning that nonparametric kernel and local polynomial regression estimation corresponds to applying a special filter. Our results thus provide a tool for comparing the performance of GPH estimators obtained from a stationary process with estimates based on residuals from nonparametric regression. This was indeed the original motivation of this work.

The paper is organized as follows. In Section 2, the background is explained and the filtered GPH estimator is defined. Asymptotic properties are derived in Section 3. Conditions under which the new estimator performs better than GPH are given. It is shown that not only the finite sample properties but also the rate of convergence can be improved. A one-parameter class of filters is introduced and discussed in Section 4. Final remarks in Section 5 conclude the paper. Proofs are given in the appendix.

## 2 The filtered log-periodogram estimator

Let $Y_t$ be a filtered process obtained from $X_t$ by

$$Y_t = \sum_{j=0}^{\infty} w_j X_{t-j} = W(B) X_t, \qquad (2.2)$$

where $B$ is the backshift operator and $W(B) = \sum_{j=0}^{\infty} w_j B^j$ is a time-invariant linear filter with $w_0 = 1$. Given observations $y_1, ..., y_n$, the filtered GPH estimator $\hat{d}$ is defined as follows. For

Fourier frequencies $\lambda_j = 2\pi j/n$ $(j = 1, ..., k)$, $k = [n/2]$, let

$$I_j = \frac{1}{2n\pi} \left| \sum_{t=1}^{n} y_t e^{it\lambda_j} \right|^2, \ j = 1, ..., k, \tag{2.3}$$

be the periodogram and $Z_j = \log|1 - e^{-i\lambda_j}| = \log|2\sin(\lambda_j/2)|$. Then

$$\log I_j = (\log f_v(0) - C) - 2d \cdot Z_j + \log(f_v(\lambda_j)/f_v(0)) + \epsilon_j, \tag{2.4}$$

where $\epsilon_j = \log(I_j/f_v(\lambda_j))$, $C = 0.577216...$ is Euler's constant, $f_v(\lambda) = f_u(\lambda) f_W(\lambda)$ and

$$f_W(\lambda) = \left| \sum_{j=0}^{\infty} w_j e^{-ij\lambda} \right|^2. \tag{2.5}$$

is the transfer function of $W$ (Priestley, 1981). Equation (2.4) is an approximate linear regression relation between $\log I_j$ and $Z_j$. For a given integer $m \leq k$, we define

$$\hat{d} = -\frac{1}{2} \frac{\sum_{j=1}^{m} (Z_j - \bar{Z}) \log I_j}{\sum_{j=1}^{m} (Z_j - \bar{Z})^2}, \tag{2.6}$$

where

$$\bar{Z} = \frac{1}{m} \sum_{j=1}^{m} Z_j. \tag{2.7}$$

Denote **I** the identity filter with $w_o = 1$ and $w_j = 0$ $(j > 0)$. Then we obtain the original GPH estimator $\hat{d}_{GPH}$, if $W = \mathbf{I}$ is used.

# 3 Properties of the filtered GPH estimator

The main objective is to study the effect of the filter $W$ on $\hat{d}$, and to see how to choose $W$ in order to achieve a considerable improvement compared to $\hat{d}_{GPH}$. The asymptotic properties of $\hat{d}$ are obtained by extending Theorem 1 of Hurvich et al. (1998). The following assumptions will be needed.

**Condition 1** $m \to \infty$ and $(m \log m)/n \to 0$ as $n, m \to \infty$.

**Condition 2** $0 < C_o \leq f_u(\lambda) \leq C_1 < \infty$, $f'_u(0) = 0$, $|f''_u(\lambda)| < C_2 < \infty$ and $|f'''_u(\lambda)| < C_3 < \infty$ for all $\lambda$ in some neighborhood of zero, and suitable finite constants $C_j$ $(j = 0, 1, 2, 3)$.

**Condition 3** $W$ is a causal, time-invariant filter with $w_0 = 1$, $\sum\limits_{j=0}^{\infty} w_j e^{-ij\lambda} \neq 0$ for all $\lambda \in [-\pi, \pi]$ and $\sum\limits_{j=1}^{\infty} j^3 |w_j| < \infty$.

**Remark 1** *The asssumption* $f_u'(0) = 0$ *is natural since it follows from symmetry of* $f_u(\lambda)$ *and the existence of* $f_u''(\lambda)$ *in a neighborhood of* $\lambda = 0$.

**Remark 2** *The relative curvature of the transfer function* $f_W$ *(or the filter* $W$*) is*

$$\beta_W = f_W''(0)/f_W(0). \tag{3.1}$$

*Under Condition 3, Theorem 1 shows that Condition 2 carries over to* $f_W$.

**Remark 3** *The relative curvature is invariant under linear transformation in the sense that* $\beta_{au+b} = \beta_u$ *for any* $a \neq 0, b \in \mathbb{R}$.

**Remark 4** *The fact that* $d$ *can be estimated from* $Y_t$ *instead of* $X_t$ *is based on the well known result (see e.g. Theorem 4.3.1 in Fuller, 1996) that under Condition 3, the spectral density of* $Y_t$ *exists and is given by*

$$\begin{aligned} f_Y(\lambda) &= f(\lambda) f_W(\lambda) \\ &= |1 - e^{-i\lambda}|^{-2d} f_u(\lambda) f_W(\lambda) \\ &=: |1 - e^{-i\lambda}|^{-2d} f_v(\lambda), \end{aligned} \tag{3.2}$$

*where*

$$f_v(\lambda) = f_u(\lambda) f_W(\lambda) \tag{3.3}$$

*is the spectral density of the (short-memory) process* $v_t = (1 - B)^d Y_t$.

Let $\beta_v$ denote the relative curvature of $f_v$. The performance of the original GPH estimator will be improved, if $W$ is such that $|\beta_v| < |\beta_u|$. Therefore, in a first step, a formula for $\beta_v$ is needed. The value of $\beta_v$ depends in turn on the relative curvature $\beta_W$. The following theorem provides a connection between simple sufficient conditions on the weights $w_j$ and regularity conditions for $f_W$, as well as a formula for the relative curvature of a convolution of filters.

**Theorem 1** *i) Under Condition 3,* $f_W(\lambda)$ *satisfies Condition 2.*
*ii) Let* $W$ *be the convolution of* $k$ *filters,* $W_1$, ..., $W_k$, *all of which satisfy Condition 3. Denote by* $\beta_W$, $\beta_{W_1}, .., \beta_{W_k}$ *the relative curvatures of* $W$, $W_1$, ..., $W_k$. *Then*

$$\beta_W = \sum_{j=1}^{k} \beta_{W_j}. \tag{3.4}$$

An equally simple formula can be obtained for the inverse filter. Under Condition 3, the inverse filter $W^I$, defined by $W^I \otimes W = W \otimes W^I = \mathbf{I}$, exists. Here, $\otimes$ denotes the convolution and $\mathbf{I}$ is the identity filter defined before. The relative curvature of $W^I$ is given by

**Lemma 1** *Let $W$ be a filter with relative curvature $\beta_W$. Then, under Condition 3, the transfer function of its inverse filter $W^I$ satisfies Condition 2 and has relative curvature $\beta_W^I = -\beta_W$.*

Sufficient assumptions that imply Condition 2 for $f_u$ are given by

**Corollary 1** *Let $u_t$ be a stationary process with Wold representation*

$$u_t = \Psi(B)\epsilon_t = \sum_{j=1}^{\infty} \psi_j \epsilon_{t-j} + \epsilon_t \tag{3.5}$$

*where $\sum_{j=1}^{\infty} j^3 |\psi_j| < \infty$ and $\epsilon_t$ are identically distributed uncorrelated zero mean random variables with finite variance. Then Condition 2 holds for the spectral density $f_u(\lambda)$ of $u_t$.*

The relative curvature of $f_v(\lambda)$ now follows from Theorem 1:

**Corollary 2** *Let $f_v(\lambda) = f_u(\lambda) f_W(\lambda)$ be as defined in (3.3). Assume that $f_u(\lambda)$ satisfies Condition 2 and the filter $W$ satisfies Conditions 3. Then $f_v(\lambda)$ satisfies Condition 2 and its relative curvature*

$$\beta_v = \frac{f_v''(0)}{f_v(0)}$$

*is given by*

$$\beta_v = \beta_u + \beta_W, \tag{3.6}$$

*where $\beta_u$, $\beta_W$ are the relative curvatures of $f_u(\lambda)$ and $f_W(\lambda)$, respectively.*

The proof of Corollary 2 is straightforward and is omitted. Note that $v_t$ can be obtained by applying the convoluted filter $W \otimes \Psi$ to $\epsilon_t$. Therefore (3.6) follows immediately from (3.4).

The implication of (3.6) is that the filter $W$ leads to an improved estimator of $d$, if

$$|\beta_u + \beta_W| < |\beta_u|.$$

This is quite a weak condition, particularly when $|\beta_u|$ is large, where, i.e. where the GPH estimator performs poorly. The ideal filter would be, of course, the unknown inverse filter

$\Psi^I$. However, full pre-whitening for all frequencies is not necessary. It is sufficient to achieve $f_W(\lambda) \approx const \cdot f_u^{-1}(\lambda)$ in a neighborhood of the origin so that $f_v''(0) \approx 0$ and $\beta_v \approx 0$. We will see below that this can be achieved for a wide range of filters, and detailed knowledge of $f_u$ is not required. A related idea is the use of a taper as proposed e.g. by Velasco (2000). However, the filtered GPH estimator is different, because a taper is not a time-invariant filter.

Using Corollary 2 of Theorem 1 in Hurvich et al. (1998) may easily be extended to the filtered estimate $\hat{d}$:

**Theorem 2** *Under Conditions 1 to 3 we have*

$$E(\hat{d} - d) = \frac{-2\pi^2 \beta_v}{9} \frac{m^2}{n^2} [1 + o(1)] + O\left(\frac{m^3 \log m}{n^3}\right) + O\left(\frac{\log^3 m}{m}\right), \tag{3.7}$$

$$\text{var}(\hat{d}) = \frac{\pi^2}{24m} [1 + o(1)], \tag{3.8}$$

$$\begin{aligned} MSE(\hat{d}) &= \frac{4\pi^4 \beta_v^2}{81} \frac{m^4}{n^4} [1 + o(1)] + \frac{\pi^2}{24m} [1 + o(1)] \\ &+ O\left(\frac{m^6 (\log m)^2}{n^6}\right). \end{aligned} \tag{3.9}$$

**Remark 5** *Both, the bias and the MSE of $\hat{d}$, depend on $\beta_v$. For a given filter, the formula for the MSE of $\hat{d}$ is analgous to the one for $\hat{d}_{GPH}$, with $\beta_u$ in the dominating term of the bias being replaced by $\beta_v$.*

**Remark 6** *If $\beta_v \neq 0$, then the bias of $\hat{d}$ is dominated by the term of order $O((m/n)^2)$. On the other hand, if $W$ is such that $\beta_W = -\beta_u$, then $\beta_v = 0$ and the dominating term is of the smaller order $O\left(\log m \cdot (m/n)^3\right)$.*

The following proposition shows when $\hat{d}$ performs better than $\hat{d}_{GPH}$. Note that, the case $\beta_u = 0$ is excluded, since there the original GPH estimator performs well and filtering does not lead to any improvements.

**Corollary 3** *Let $\beta_u \neq 0$, and $W$ such that*

$$-2 < \frac{\beta_W}{\beta_u} < 0 \tag{3.10}$$

*or equivalently*

$$-1 < \frac{\beta_v}{\beta_u} < 1. \tag{3.11}$$

*Then the dominating term in the bias of $\hat{d}$ is smaller than the corresponding term for $\hat{d}_{GPH}$.*

The proof of Corollary 3 is straightforward and is therefore omitted. The result leads to the following rule for choosing a filter:

- If $\beta_u > 0$, then $\hat{d}_{GPH}$ can be improved by applying a filter with $-2\beta_u < \beta_W < 0$.
- If $\beta_u < 0$, then $W$ should be such that $0 < \beta_W < -2\beta_u$.

**Remark 7** *The effect of filtering can be explained in detail as follows. All filters satisfying Condition 3 may be grouped according to the value of $\beta_W$. Consider the case $\beta_u < 0$, i.e. $f_u''(0) < 0$ and $f_u$ is convex at zero. Then the dominating term in the bias of $\hat{d}_{GPH}$ is positive. Applying a filter with $\beta_W < 0$ increases the positive bias term so that $\hat{d}_{GPH}$ is corrected in the wrong direction. On the other hand, if $W$ is such that $0 < \beta_W < -\beta_u$, then the dominating bias term of $\hat{d}$ is still positive but smaller than for $\hat{d}_{GPH}$. In the optimal case with $\beta_W = -\beta_u$, the first term in (3.7) vanishes and the bias of $\hat{d}$ is even of a smaller order of magnitude. If $-\beta_u < \beta_W < -2\beta_u$, then the dominating bias term of $\hat{d}$ is negative but with a smaller absolute value. If $\beta_W > -2\beta_u$, then the dominating bias term of $\hat{d}$ is negative and its absolute value is larger. Thus, $\hat{d}_{GPH}$ is over-corrected.*

Theorem 2 implies an optimal choice of the number $m$ of Fourier frequencies used in the regression estimate. Assume that $f_v''(0) \neq 0$. Then, in analogy to Hurvich et al. (1998), the optimal $m$ that minimizes the dominating part of MSE($\hat{d}$) is given by

$$m^{opt} = \left( \frac{27\beta_v^{-2}}{128\pi^2} \right)^{1/5} n^{4/5}. \tag{3.12}$$

The value of $m^{opt}$ strongly depends on $\beta_v$ and hence on $\beta_W$. If $\beta_W$ satisfies (3.10) resp. (3.11), then $\beta_v^2 < \beta_u^2$, $m^{opt} > m_{GPH}^{opt}$ and $\text{MSE}_{opt}(\hat{d}) < \text{MSE}_{opt}(\hat{d}_{GPH})$ asymptotically. The latter follows since both, the squared bias and the variance, are reduced by filtering. If a filter with $\beta_W = -\beta_u$ could be chosen, then $m^{opt}$ would even be of a larger order than $O(n^{4/5})$ so that the rate of convergence of $\hat{d}$ would be faster than the one of $\hat{d}_{GPH}$.

# 4 A simple class of filters

To make the results in the previous section applicable, we consider a specific class of filters. Let

$$F_{(\alpha,k)}(B) = 1 - \alpha \sum_{j=1}^{\infty} \frac{j^{-k}}{\zeta(k)} B^j = \sum_{j=0}^{\infty} \psi_j B^j, \tag{4.1}$$

where $-1 < \alpha < 1$, $k > 4$, $\psi_0 = 1$,

$$\psi_j = -\frac{\alpha j^{-k}}{\zeta(k)}, j = 1, ... \tag{4.2}$$

and

$$\zeta(k) = \sum_{j=1}^{\infty} j^{-k} \tag{4.3}$$

is Riemann's $\zeta$-function. Note that $\zeta$ decreases monotonically in $k$ and $\lim_{k\to\infty} \zeta(k) = 1$. The factor $\zeta^{-1}(k)$ is introduced in oder that $\sum_{j=1}^{\infty} \psi_j = \alpha$.

**Lemma 2** *The filter $F_{(\alpha,k)}$ defined above satisfies Condition 3.*

Denote by $f_\alpha(\lambda)$ the transfer function of $F_{(\alpha,k)}$ and by $\beta(\alpha, k)$ the relative curvature of $f_\alpha(\lambda)$. We have

**Theorem 3** *i) The relative curvature of $f_\alpha(\lambda)$ is equal to*

$$\beta(\alpha, k) = \frac{2\alpha}{(1-\alpha)^2}\frac{\zeta(k-2)}{\zeta(k)} = \frac{2\alpha}{(1-\alpha)^2}q(k) \tag{4.4}$$

*and increases monotonically in $\alpha$, where $q(k) = \zeta(k-2)/\zeta(k)$. Moreover, $\beta(0, k) = 0$ for any $k > 4$.*

*ii) For any $\beta_u < 0$ there exit two unique $0 < \alpha_0 < \alpha_1$ such that $\beta(\alpha_0, k) = -\beta_u$ and $\beta(\alpha_1, k) = -2\beta_u$.*

Theorem 3 shows that, for $\beta_u < 0$, the range of $\alpha$ where filtering improves the GPH-estimate of $d$ is $0 < \alpha < \alpha_1$. The larger $|\beta_u|$ the larger is this range, due to the monotonicity of $\beta(\alpha, k)$. A similar result can be obtained for positive values of $\beta_u$, however only if $\beta_u$ is not too large. If $\beta_u >> 0.5$, then an $\alpha_0$ such that $\beta(\alpha_0, k) = -\beta_u$ does not exit. The reason is that $\lim_{\alpha\to-1} \beta(\alpha, k) \geq -0.5 \inf_k q(k) > -\infty$ so that $\beta(\alpha, k)$ cannot reach $-\beta_u$ for any $\alpha \in (-1, 1)$. As a result, the correction by $F_{(\alpha,k)}$ will always be suboptimal, independently of the choice of $\alpha$. For the case $\beta_u > 0$, we therefore propose to use instead the inverse filter of $F_{(\alpha,k)}$, denoted by

$$F^I_{(\alpha,k)}(B) = \sum_{j=0}^{\infty} \varphi_j B^j. \tag{4.5}$$

Lemma 2 implies the existence of $F^I_{(\alpha,k)}$. Exact and approximate formulas for the coefficients $\varphi_j$ are given by

**Theorem 4** *i) The coefficients of $F^I_{(\alpha,k)}(B)$ are given by $\varphi_0 = 1$, $\varphi_1 = -\psi_1$ and, for $j \geq 1$,*

$$\varphi_j = -\sum_{l=1}^{j} \psi_l \varphi_{j-l}. \tag{4.6}$$

*ii) For any $j = 0, 1, ...$*

$$\varphi_j = \frac{\alpha^j}{\zeta^j(k)} + O(2^{-k}). \tag{4.7}$$

Denote the relative curvature of $F^I_{(\alpha,k)}$ by $\beta^I(\alpha, k)$. Lemma 1 and Theorem 3 imply

**Corollary 4** *i) The relative curvature of $F^I_{(\alpha,k)}$ is given by*

$$\beta^I(\alpha, k) = -\frac{2\alpha}{(1-\alpha)^2} q(k) \tag{4.8}$$

*and decreases monotonically in $\alpha$, where $q(k)$ is as defined in Theorem 3. Moreover, $\beta^I(0, k) = 0$ for any $k > 4$.*

*ii) For any $\beta_u > 0$ there exit two unique $0 < \alpha_0 < \alpha_1$ such that $\beta^I(\alpha_0, k) = -\beta_u$ and $\beta^I(\alpha_1, k) = -2\beta_u$.*

These results lead to the following rule:

- If $\beta_u > 0$, then apply $F^I_{(\alpha,k)}$.
- If $\beta_u < 0$, then apply $F_{(\alpha,k)}$.

Two tuning parameters are not specified in this rule, namely $k$ and $\alpha$. The first parameter $k$ is not critical, in the sense that for any given $k > 4$, there always exist two unique $|\alpha_0| < |\alpha_1|$ such that the results *ii*) in Theorem 3 and Corollary 4 hold. We may therefore prefer a simplified filter where $k$ does not occur. The simplest solution is to consider the limits of $F_{(\alpha,k)}$ and $F^I_{(\alpha,k)}$ as $k$ tends to infinity. We first define the limit of a sequence of filters.

**Definition 1** *Let $W_k$, $k = 1, 2, ...$, with weights $w_{jk}$ ($j = 0, 1, 2, ...$), $w_{0k} \equiv 1$, denote a sequence of filters. Then a given filter $W$ with weights $w_j$, $j = 0, 1, ...$ is called the limit of $W_k$, as $k \to \infty$, if $\lim_{k\to\infty} w_{jk}$ exist for all $1 \leq j < \infty$ and*

$$\sup_j |\lim_{k\to\infty} w_{jk} - w_j| = 0. \tag{4.9}$$

Following Definition 1 it is clear that if the limit of a sequence of filters exists, then it is unique. Furthermore, it can also be shown that, if the relative curvatures of $W_k$ are $\beta_{W_k}$ $(k = 1, 2, ...)$, and that of $W$ is $\beta_W$, then

$$\lim_{k\to\infty} \beta_{W_k} = \beta_W.$$

The following lemma shows that the limits of $F_{(\alpha,k)}$ and $F^I_{(\alpha,k)}$ exist and are two well known classes of filters.

**Lemma 3** *For any $-1 < \alpha < 1$ and the sequences of filters $F_{(\alpha,k+4)}$ and $F^I_{(\alpha,k+4)}$, $k = 1, 2, ...,$ we have*

*i)* $\lim_{k\to\infty} F_{(\alpha,k)}(B) = (1 - \alpha B) =: F_{(\alpha,MA)}(B),$

*ii)* $\lim_{k\to\infty} F^I_{(\alpha,k)}(B) = (1 - \alpha B)^{-1} =: F_{(\alpha,AR)}(B),$

*iii)* $F_{AR}(\alpha) = F^I_{MA}(\alpha).$

Thus, $F_{(\alpha,MA)}(B)$ and $F_{(\alpha,AR)}(B)$ are the first order moving average (MA) and the first order autoregressive (AR) filters. By taking limits, we obtain

**Lemma 4** *Denote the relative curvature of $F_{(\alpha,MA)}$ and $F_{(\alpha,AR)}$ by $\beta_{MA}(\alpha)$ and $\beta_{AR}(\alpha)$ respectively. Then*

$$\beta_{MA}(\alpha) = \frac{2\alpha}{(1-\alpha)^2} \tag{4.10}$$

*and*

$$\beta_{AR}(\alpha) = \frac{-2\alpha}{(1-\alpha)^2}. \tag{4.11}$$

The proof of this lemma is omitted. Note that $\beta_{MA}(\alpha)$ increases monotonically in $\alpha$, with $\lim_{\alpha\to 1} \beta_{MA}(\alpha) = \infty$ and $\lim_{\alpha\to -1} \beta_{MA}(\alpha) = -0.5$. Thus, for any $\beta_u < 0$ there exists a unique $\alpha_0$ such that $\beta_{MA}(\alpha) = -\beta_u$, whereas this is not the case if $\beta_u > 0.5$. On the other hand, $\beta_{AR}(\alpha)$ decreases monotonically in $\alpha$ with $\lim_{\alpha\to 1} \beta_{AR}(\alpha) = -\infty$ and $\lim_{\alpha\to -1} \beta_{AR}(\alpha) = 0.5$. Therefore, for any $\beta_u > 0$ there exists a unique $\alpha_0$ such that $\beta_{AR}(\alpha) = -\beta_u$, whereas no such $\alpha$ exists if $\beta_u < -0.5$. This motivates the following definitions

**Definition 2** *For $\alpha \in (0, 1)$, $\tau \in \{-1, 1\}$, define*

$$F_{\alpha,\tau}(B) = \begin{cases} F_{(\alpha,MA)}(B) & , \textit{ for } \tau = 1, \\ F_{(\alpha,AR)}(B) & , \textit{ for } \tau = -1. \end{cases} \tag{4.12}$$

**Definition 3** *Let* $\alpha \in (0,1)$*,* $\tau$ *=-sign*$\beta_u$*, and* $Y_t = F_{\alpha,\tau}(B)X_t$*. Then* $\hat{d} = \hat{d}(\alpha)$ *based on* $y_1, ..., y_n$ *is called the* $\alpha-$*filtered GPH estimator.*

**Remark 8** *Note that* $\hat{d}(0) = \hat{d}_{GPH}$*.*

For fixed $\alpha$, the calculation of $\hat{d}(\alpha)$ requires the knowledge or estimation of the sign of $\beta_u$. Estimating a sign corresponds to a 0-1-decision, and is therefore easier than estimating $\beta_u$ itself. Asymptotically, sign$\beta_u$ can be estimated correctly with probability one, and the asymptotic mean squared error of $\hat{d}$ is not affected. The only remaining problem is therefore the choice of $\alpha$. Theoretically, the range of $\alpha-$values that improve estimation of $d$ follows from the results above:

**Theorem 5** *i) The relative curvature of the filter* $F_{\alpha,\tau}(B)$ *is given by*

$$\beta_F(\alpha, \tau) = \beta_F(\alpha \cdot \tau) = \frac{2\alpha\tau}{(1-\alpha)^2} \tag{4.13}$$

*As a function of* $\alpha \cdot \tau$*,* $\beta_F(\alpha \cdot \tau)$ *is antisymmetric and monotonically increasing, with* $\beta_F(0) = 0$*,* $\lim_{\alpha\tau \to -1} \beta_F(\alpha\tau) = -\infty$ *and* $\lim_{\alpha\tau \to 1} \beta_F(\alpha\tau) = \infty$*.*

*ii) For any* $-\infty < \beta_u < \infty$ *there exist two unique* $0 \leq \alpha_0 \leq \alpha_1$ *such that* $\beta_v(\alpha_0\tau) = 0$ *and* $\beta_v(\alpha_1\tau) = -\beta_u$ *where* $\tau$ *=-sign*$\beta_u$*. For any* $\alpha \leq \alpha_1$ *we have*

$$\lim_{n\to\infty} \frac{MSE_{opt}(\hat{d}(\alpha))}{MSE_{opt}(\hat{d}_{GPH})} \leq 1.$$

*More explicitely,* $\alpha_0 = \alpha_1 = 0$ *for* $\beta_u = 0$ *and, for* $\beta_u \neq 0$*,*

$$\alpha_0 = \frac{1 + |\beta_u| - \sqrt{1 + 2|\beta_u|}}{|\beta_u|}, \tag{4.14}$$

$$\alpha_1 = \frac{1 + 2|\beta_u| - \sqrt{1 + 4|\beta_u|}}{2|\beta_u|}. \tag{4.15}$$

**Remark 9** *The solutions* $\alpha_0$ *and* $\alpha_1$ *do not depend on* $n$ *and* $m$*. In the special case where* $u_t$ *itself is an AR(1) process with coefficient* $\varphi > 0$*, we have* $\alpha_0 = \varphi$*. Similarily, if* $u_t$ *is an MA(1) process with coefficient* $\psi < 0$*, then* $\alpha_0 = -\psi$*. In these two cases, the short-memory process* $u_t$ *is pre-whitened completely and* $\hat{d}(\alpha)$ *is essentially a parametric estimator.*

A small simulation study confirms the theoretical results and the use of the filter $F_{\alpha,\tau}(B)$ defined in (4.12). This will not be reported here to save space. However, how do we chose $\alpha$

practically, i.e. if $\beta_u$ is unknown? First of all note that $\beta_v^2(\alpha)$ is concave with minimum zero at $\alpha_0$. Also, the optimal value of $\alpha$ does not depend on $m$ or $n$. In contrast, the optimal number of frequencies, $m^{opt}$, defined in (3.12) strongly depends on $\beta_v$ and hence on $\alpha$. Instead of trying to find $m^{opt}$ directly, it is therefore much easier to first estimate $\alpha_o$ by minimizing an estimate of the dominating term of the squared bias. The latter can be done using any reasonable $m$. Once $\alpha$ is selected, $m$ can be chosen by applying, for instance, the algorithm in Hurvich and Beltrao (1994) to the filtered data. An explicit development of a data-driven algorithm and its practical implementation is beyond the aim of the current paper and will be discussed elsewhere.

# 5 Final remarks

In this paper, we discussed the effect of filtering on the GPH-estimator of the fractional parameter $d$. A modification was proposed based on a class of filters characterized by a one-dimensional parameter $\alpha$ and the sign of $\beta_u$. The method focusses on eliminating the main source of bias, and provides the basis for a simple but effective data-driven algorithm. Since nonparametric regression estimators correspond to special filters, the results also provide a tool for comparing the performance of estimators obtained from a stationary process with those based on residuals from nonparametric regression.

## **Appendix:** Proofs of results

**Proof of Theorem 1.** *i*) Note that

$$\begin{aligned} f_W(\lambda) &= \left| \sum_{j=0}^{\infty} w_j e^{-ij\lambda} \right|^2 \\ &= \left( \sum_{j=0}^{\infty} w_j \cos(j\lambda) \right)^2 + \left( \sum_{j=1}^{\infty} w_j \sin(j\lambda) \right)^2 . \end{aligned} \tag{A.1}$$

$f_W(\lambda)$ is bounded from zero, since $\sum_{j=0}^{\infty} w_j e^{-ij\lambda} \neq 0$ for $\lambda \in [-\pi, \pi]$. And $f_W(\lambda)$ is bounded from above, because $\sum_{j=1}^{\infty} |w_j| < \infty$ which is satisfied, because $\sum_{j=1}^{\infty} |w_j| \leq \sum_{j=1}^{\infty} |j|^3 |w_j| < \infty$. Under this condition $f'_W(\lambda)$ exists and is given by

$$\begin{aligned} f'_W(\lambda) &= -2 \left( \sum_{j=0}^{\infty} w_j \cos(j\lambda) \right) \left( \sum_{j=0}^{\infty} j w_j \sin(j\lambda) \right) \\ &\quad +2 \left( \sum_{j=1}^{\infty} w_j \sin(j\lambda) \right) \left( \sum_{j=1}^{\infty} j w_j \cos(j\lambda) \right) \\ &= -2 w_0 \left( \sum_{j=1}^{\infty} j w_j \sin(j\lambda) \right) , \end{aligned} \tag{A.2}$$

which is antisymmetric and bounded on $[-\pi, \pi]$ with $f'_W(0) = 0$. Straightforward calculation shows that $f_W^{(k)}(\lambda)$ exists and is bounded, if $\sum_{j=1}^{\infty} |j|^k |w_j| < \infty$. Hence, under the condition $\sum_{j=1}^{\infty} |j|^3 |w_j| < \infty$, we have both $|f''_W(\lambda)| < \infty$ and $|f'''_W(\lambda)| < \infty$ for all $\lambda \in [-\pi, \pi]$. Thus Condition 2 is satisfied.

*ii*) Let $W$ be the convolution of $W_1$, ..., $W_k$, $k \geq 2$. Let $f_W(\lambda)$, $f_{W_1}(\lambda)$, ..., $f_{W_k}(\lambda)$ denote their transfer functions, all of them have zero derivative at $\lambda = 0$ and bounded second derivatives. Following the idea of Remark 4 (cf. the proof of Theorem 4.3.1 in Fuller, 1996) it is easy to show that, if $k = 2$, $f_W(\lambda) = f_{W_1}(\lambda) f_{W_2}(\lambda)$. This can then be generalized to

$$f_W(\lambda) = \prod_{j=1}^{k} f_{W_j}(\lambda). \tag{A.3}$$

Under the assumptions of Theorem 1 we have

$$f''_W(\lambda) = \sum_{j=1}^{k} f''_{W_j}(\lambda) \prod_{l \neq j} f_{W_1}(\lambda) + \sum_{j=1}^{k} \sum_{l=1, l \neq j}^{k} f'_{W_j}(\lambda) f'_{W_1}(\lambda) \prod_{m \neq j, m \neq l} f_{W_m}(\lambda).$$

Note that $f'_{W_j}(0) = 0$ for all $j$. We have

$$\beta_W = \frac{f''_W(0)}{f_W(0)}$$

$$
\begin{aligned}
&= \frac{\sum_{j=1}^{k} f''_{W_j}(0) \prod_{l \neq j} f_{W_1}(0)}{\prod_{j=1}^{k} f_W(0)} \\
&= \sum_{j=1}^{k} \frac{f''_{W_j}(0)}{f_{W_j}(0)} \\
&= \sum_{j=1}^{k} \beta_{W_j}. \qquad \text{(A.4)}
\end{aligned}
$$

Theorem 1 is proved. ◇

**Proof of Lemma 1.** Let $f_I(\lambda)$ be the transfer function of $\mathbf{I}$. We have $f_I(\lambda) \equiv 1$ and

$$f_{W^I}(\lambda) = \frac{1}{f_W(\lambda)}, \qquad \text{(A.5)}$$

which is bounded from zero and bounded above, if $f_W(\lambda)$ is. Furthermore,

$$f'_{\mathrm{W^I}}(\lambda) = \frac{-f'_{\mathrm{W}}(\lambda)}{f^2_{\mathrm{W}}(\lambda)}. \qquad \text{(A.6)}$$

Hence $f'_{\mathrm{W^I}}(0) = 0$ provided $f'_{\mathrm{W}}(0) = 0$. Similarly, $f''_{\mathrm{W^I}}(\lambda)$ and $f'''_{\mathrm{W^I}}(\lambda)$ are bounded in a neighbourhood of $\lambda = 0$, if $f_{\mathrm{W}}(\lambda)$ satisfies Condition 2.

Following (A.5), it can be further shown that

$$f''_{\mathrm{W^I}}(0) = \frac{-f''_{\mathrm{W}}(0)}{f^2_{\mathrm{W}}(0)}. \qquad \text{(A.7)}$$

Also,

$$
\begin{aligned}
\beta_{\mathrm{W^I}} &= \frac{f''_{\mathrm{W^I}}(0)}{f_{\mathrm{W^I}}(0)} \\
&= \frac{-f''_{\mathrm{W}}(0)}{f^2_{\mathrm{W}}(0)} f_{\mathrm{W}}(0) \\
&= -\frac{f''_{\mathrm{W}}(0)}{f_{\mathrm{W}}(0)} \\
&= -\beta_{\mathrm{W}}. \qquad \text{(A.8)}
\end{aligned}
$$

Lemma 1 is proved. ◇

**Proof of Corollary 1.** Note that $u_t$ is a MA($\infty$) process with iid innovations $\epsilon_t$ and absolutely summable coefficients. It is well known that (see e.g. Corollary 4.3.1.1 in Fuller, 1996) the spectral density of such a process is given by

$$f_{\mathrm{u}}(\lambda) = \frac{\sigma^2_\epsilon}{2\pi} \left( \sum_{j=0}^{\infty} \psi_j e^{-ij\lambda} \right) \left( \sum_{j=0}^{\infty} \psi_j e^{ij\lambda} \right)$$

$$= \frac{\sigma_\epsilon^2}{2\pi}\left(\sum_{j=0}^{\infty}\psi_j\left[\cos(j\lambda) - i\sin(j\lambda)\right]\right)\left(\sum_{j=0}^{\infty}\psi_j\left[\cos(j\lambda) + i\sin(j\lambda)\right]\right)$$

$$= \frac{\sigma_\epsilon^2}{2\pi}\left\{\left(\sum_{j=0}^{\infty}\psi_j\cos(j\lambda)\right)^2 + \left(\sum_{j=1}^{\infty}\psi_j\sin(j\lambda)\right)^2\right\}, \tag{A.9}$$

which is up to a positive constant factor the same as the transfer function of the linear filter with $\psi_j$ as coefficients. Hence results in Theorem 1 $i$) hold for $f_u$. ◇

**Proof of Theorem 2.** Here only a very brief description as a connection between the proofs of Theorem 2 here and Theorem 1 in Hurvich et al. (1998) will be given. For more details we refer the reader to original proof.

Corollary 2 ensures that under Conditions 1 to 3 the conditions of Theorem 1 in Hurvich et al. (1998) are all fulfilled. Let $I_j^*$ denote the periodogram ordinates at $\lambda_j$ of the original process $X_t$ and define $\epsilon_j^* = \log(I_j^*/f(\lambda_j))$. Following the idea of Remark 4 and (A.3) it can be shown that

$$I_j = I_j^* f_{\mathrm{W}}(\lambda_j)[1 + O_p(n^{-1})] \tag{A.10}$$

and

$$\epsilon_j = \epsilon_j^*[1 + O_p(n^{-1})], \tag{A.11}$$

where the $O_p(n^{-1})$ term is due to the finite sample. This shows that the asymptotic variances of $\hat{d}$ and $\hat{d}_{\mathrm{GPH}}$ are the same up to a $O(n^{-1})$ term, which is given in (3.8). See also (5) in Hurvich et al. (1998).

The bias of the GPH estimator is quantified by (4) in Hurvich et al. (1998). All of the details of the proof of Theorem 1 in Hurvich et al. (1998) related to the bias part stay asymptotically unchanged by replacing $f_u(\lambda)$ with $f_v(\lambda)$, because $f_v(\lambda)$ also satisfies the corresponding assumptions. The dominating part of the bias is given in Lemma 1 there. Note in particular that the $o\left(\frac{m^2}{n^2}\right)$ term there is indeed of the order $O\left(\frac{m^3\log m}{n^3}\right)$ (see the proof of Lemma 1 in Hurvich et al., 1998). This concludes the proof of Theorem 2. ◇

**Proof of Lemma 2.** For $F_{(\alpha,k)}$ we have $\sum_{j=0}^{\infty}\psi_j e^{-ij\lambda} > 0$ for all $\lambda \in [-\pi, \pi]$, because $\phi_j$ are of one sign for $j > 0$ and $\sum_{j=1}^{\infty}|\psi_j| = |\alpha| < 1$. Furthermore, the condition $k > 4$ ensures that

$$\sum_{j=1}^{\infty} j^3|\psi_j| \leq \frac{\alpha}{\zeta(k)}\sum_{j=1}^{\infty} j^{-2} < \infty.$$

◇

**Proof of Theorem 3.** *i*) Following (A.1) we have

$$f_\alpha(0) = \left(\sum_{j=0}^{\infty} \psi_j\right)^2 = (1-\alpha)^2. \tag{A.12}$$

Following (A.2) we can obtain

$$f''_{\mathrm{W}}(\lambda) = -2\left(\sum_{j=1}^{\infty} j^2 \psi_j \cos(j\lambda)\right). \tag{A.13}$$

Hence,

$$\begin{aligned}\beta(\alpha,k) &= \frac{f''_\alpha(0)}{f_\alpha(0)} \\ &= \frac{2\alpha}{(1-\alpha)^2}\frac{\sum_{j=1}^{\infty} j^{k-2}}{\zeta(k)} \\ &= \frac{2\alpha}{(1-\alpha)^2}\frac{\zeta(k-2)}{\zeta(k)}.\end{aligned} \tag{A.14}$$

*ii*) Straightforward calculation shows that

$$\beta'(\alpha,k) = \frac{2}{(1-\alpha)^3}\frac{\zeta(k-2)}{\zeta(k)} > 0 \text{ for all } \alpha \in (-1,1).$$

Thus, $\beta(\alpha,k)$ is continuous, increases monotonically in $\alpha$ with $\beta(0,k) = 0$ and $\beta(\alpha,k) \to \infty$ as $\alpha \to 1$. Hence, for any $\beta_u < 0$, unique solutions $\alpha_1$ and $\alpha_2$ exist such that $\beta(\alpha_1,k) = -\beta_u$ and $\beta(\alpha_2,k) = -2\beta_u$ which satisfy $0 < \alpha_1 < \alpha_2 < 1$, because $0 < -\beta_u < -2\beta_u < \infty$. $\diamond$

**Proof of Theorem 4.** *i*) By definition we have

$$F_{(\alpha,k)}(B) \otimes F^{\mathrm{I}}_{(\alpha,k)}(B) = \left(\sum_{i=0}^{\infty} \psi_i B^i\right)\left(\sum_{j=0}^{\infty} \varphi_j B^j\right) \equiv 1. \tag{A.15}$$

The first part can be proved by matching the orders on the left- and right-hand sides of (A.15). Thus, $\psi_0\varphi_0 = 1$, implying $\varphi_0 = 1$, and for any $j > 0$, (4.6) follows from

$$\sum_{l=0}^{j} \psi_l \varphi_{j-l} = 0.$$

*ii*) Now we will give more detailed calculations to show that simple closed form formulae for $\varphi_j$ might not exists. For $j = 1$ we have

$$\varphi_1 = -\psi_1 = \frac{\alpha}{\zeta(k)},$$

for $j=2$

$$\varphi_2 = -\psi_1\varphi_1 - \psi_2 = \frac{\alpha^2}{\zeta^2(k)} + \frac{\alpha 2^{-k}}{\zeta(k)},$$

for $j=3$

$$\begin{aligned}\varphi_3 &= -\psi_1\varphi_2 - \psi_2\varphi_1 - \psi_3 \\ &= \frac{\alpha}{\zeta(k)}\left(\frac{\alpha^2}{\zeta^2(k)} + \frac{\alpha 2^{-k}}{\zeta(k)}\right) - \frac{\alpha^2 2^{-k}}{\zeta^2(k)} + \frac{\alpha 3^{-k}}{\zeta(k)} \\ &= \frac{\alpha^3}{\zeta^3(k)} + \frac{\alpha 3^{-k}}{\zeta(k)}\end{aligned}$$

and for $j=4$

$$\begin{aligned}\varphi_4 &= -\psi_1\varphi_3 - \psi_2\varphi_2 - \psi_3\varphi_1 - \psi_4 \\ &= \frac{\alpha}{\zeta(k)}\left(\frac{\alpha^3}{\zeta^3(k)} + \frac{\alpha 3^{-k}}{\zeta(k)}\right) + \frac{\alpha 2^{-k}}{\zeta(k)}\left(\frac{\alpha^2}{\zeta^2(k)} + \frac{\alpha 2^{-k}}{\zeta(k)}\right) - \frac{\alpha 3^{-k}}{\zeta(k)}\frac{\alpha}{\zeta(k)} + \frac{\alpha 4^{-k}}{\zeta(k)} \\ &= \frac{\alpha^4}{\zeta^4(k)} + \frac{\alpha^3 2^{-k}}{\zeta^3(k)} + \frac{\alpha^2 4^{-k}}{\zeta^2(k)} + \frac{\alpha 4^{-k}}{\zeta(k)}.\end{aligned}$$

For $j > 4$ we can see that the first term of $-\psi_1\varphi_{j-1}$ is always $\frac{\alpha^j}{\zeta^j(k)}$, which is the dominating term of $\varphi_j$. All other terms are at most of the order $O(2^{-k})$. ◇

**Proof of Lemma 3.** *i*) Note that $\psi_1 \to \alpha$ and $\psi_j \to 0$ for $j > 1$, as $k \to \infty$. Thus, $F_{(\alpha,k)}(B) \to (1-\alpha B)$ as $k \to \infty$.

*ii*) Following *ii*) of Theorem 4 we have $\varphi_j \to \alpha^j$ for all $j = 0, 1, ...,$ as $k \to \infty$. These are the coefficients of $(1-\alpha B)^{-1}$. Thus, $F^{\mathrm{I}}_{(\alpha,k)}(B) \to (1-\alpha B)^{-1}$ as $k \to \infty$.

*iii*) This result is obvious. ◇

**Proof of Theorem 5.** *i*) Results in this part can be obtained by simply combining the two formulae given in Lemma 4 and by checking their limits.

*ii*) Note that for $\beta_{\mathrm{u}} = 0$ we have obviously $\alpha_0 = \alpha_1 = 0$. Hence we will only consider the case with $\beta_{\mathrm{u}} \neq 0$. For $\beta_{\mathrm{v}} = 0$ we have $\beta_{\mathrm{F}}(\alpha\tau) = -\beta_{\mathrm{u}}$. $\alpha_0$ can be obtained by solving the equation

$$\frac{2\alpha_0\tau}{(1-\alpha_0)^2} = -\beta_{\mathrm{u}}. \tag{A.16}$$

For $\beta_{\mathrm{u}} > 0$ we have $\tau = -1$ and

$$\frac{2\alpha_0}{(1-\alpha_0)^2} = \beta_{\mathrm{u}}. \tag{A.17}$$

Straightforward calculation leads to the solution

$$\alpha_0 = \frac{1 + \beta_{\mathrm{u}} \pm \sqrt{1 + 2\beta_{\mathrm{u}}}}{\beta_{\mathrm{u}}}. \tag{A.18}$$

Note that $\alpha_0 < 1$. Hence, the unique solution is

$$\alpha_0 = \frac{1 + \beta_\mathrm{u} - \sqrt{1 + 2\beta_\mathrm{u}}}{\beta_\mathrm{u}}. \tag{A.19}$$

For $\beta_\mathrm{u} < 0$ we have $\tau = 1$, leading to the unique solution

$$\alpha_0 = -\frac{1 - \beta_\mathrm{u} - \sqrt{1 - 2\beta_\mathrm{u}}}{\beta_\mathrm{u}}. \tag{A.20}$$

For $\beta_\mathrm{v} = -\beta_\mathrm{u}$ we have $\beta_\mathrm{F}(\alpha\tau) = -2\beta_\mathrm{u}$. Similarly the unique solution of

$$\frac{2\alpha_1 \tau}{(1 - \alpha_1)^2} = -2\beta_\mathrm{u} \tag{A.21}$$

is

$$\alpha_1 = \frac{1 + 2\beta_\mathrm{u} - \sqrt{1 + 4\beta_\mathrm{u}}}{2\beta_\mathrm{u}} \tag{A.22}$$

for $\beta_\mathrm{u} > 0$ and

$$\alpha_1 = -\frac{1 - 2\beta_\mathrm{u} - \sqrt{1 - 4\beta_\mathrm{u}}}{2\beta_\mathrm{u}} \tag{A.23}$$

for $\beta_\mathrm{u} < 0$.

For $\beta_\mathrm{u} > 0$ we have

$$\alpha_1 - \alpha_0 = \frac{2\sqrt{1 + 2\beta_\mathrm{u}} - (1 + \sqrt{1 + 4\beta_\mathrm{u}})}{2\beta_\mathrm{u}}. \tag{A.24}$$

Let $T_1 = 2\sqrt{1 + 2\beta_\mathrm{u}}$ and $T_2 = 1 + \sqrt{1 + 4\beta_\mathrm{u}}$. $T_1$ and $T_2$ are both positive, and

$$\begin{aligned} T_1^2 - T_2^2 &= 2(1 + 2\beta_\mathrm{u} - \sqrt{1 + 4\beta_\mathrm{u}}) \\ &= 2(\sqrt{1 + 4\beta_\mathrm{u} + 4\beta_\mathrm{u}^2} - \sqrt{1 + 4\beta_\mathrm{u}}) > 0. \end{aligned}$$

Thus, $0 < \alpha_0 < \alpha_1 < 1$ for $\beta_\mathrm{u} > 0$. This also follows from the monotonicity of $\beta_\mathrm{F}(\alpha)$. Similarly, we have $0 < \alpha_0 < \alpha_1 < 1$ for $\beta_\mathrm{u} < 0$, too. Furthermore, all of these solutions can be summarized in the form given in (4.14) and (4.15).

For any $\alpha \le \alpha_1$ we have $\beta_v^2 \le \beta_u^2$. Following (3.12) we have $m_{\hat{d}(\alpha)}^{opt} \ge m_{\hat{d}_{GPH}}^{opt}$. This results in that the two dominating terms of $MSE_{opt}(\hat{d}(\alpha))$ as given in (3.9) are both asymptotically smaller than those for $MSE_{opt}(\hat{d}_{GPH})$, respectively. We have therefore

$$\lim_{n \to \infty} \frac{MSE_{opt}(\hat{d}(\alpha))}{MSE_{opt}(\hat{d}_{GPH})} \le 1.$$

◇

**References:**


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

Beran, J., 1995. Maximum likelihood of estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. R. Statist. Soc.* B, 57, 659–672.

Beran, J., Bhansali, R.J. and Ocker, D., 1999. On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. Biometrika, 85, 921–934.

Blackman, R.B. and Tukey, J.W., 1959. *The measurements of power spectra from the point of view of communications engineering.* Dover, New York. (Originally published in *Bell Systems Tech. J., 37, 1958.*)

Dahlhaus, R., 1989. Efficitive parameter estimation for self-similar processes. *Ann. Statist.*, 17, 1749–1766.

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

Geweke, J. and Porter-Hudak, S., 1983. The estimation of long memory time series models. *J. Time Ser. Anal.*, 4, 221–238.

Giraitis, L., Robinson, P.M. and Samarov, A., 1997. Rate optimal semiparametric estimation of the memory parameter of Gaussian time series with long-range dependence. *J. Time Ser. Anal.*, 18, 49–60.

Giraitis, L. and Surgailis, D., 1990. A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittel's estimate. *Probab. Th. Rel. Fields*, 86, 87–104.

Hurvich, C.M. and Beltrao, K.I., 1993. Asymptotics for low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Ser. Anal.*, 14, 455–472.

Hurvich, C.M. and Beltrao, K.I., 1994. Automatic semiparametric estimation of the memory parameter of a long-memory time series. *J. Time Ser. Anal.*, 15, 285–302.

Hurvich, C.M., Deo, R. and Brodsky, J., 1998. The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Ser. Anal.*, 19, 19–46.

Künsch, H., 1986. Discrimination between monontone trends and long-range dependence. J. Appl. Probab., 23, 1025–30.

Priestley, M.B., 1981. *Spectral analysis and time series* (Volumes 1 and 2). Academic Press, London.

Robinson, P.M., 1994. Semiparametric analysis of long-memory time series. *Ann. Statist.*, 22, 515–539.

Robinson, P.M., 1995. Log periodogram regression of time series with long-range dependence. *Ann. Statist.*, 23, 1048–1072.

Shimotsu, K. and Phllips, P.C.B., 2002. Pooled log periodogram regression. *J. Time Ser. Anal.*, 23, 57–93.

Velasco, C., 1999. Non-stationary log-periodogram regression. *J. Econometrics*, 91, 325–371.

Velasco, C., 2000. Non-Gaussian log-periodogram regression. *Econometric Theory*, 16, 44–79.