Leibniz Universität Hannover

Wirtschaftswissenschaftliche Fakultät

Institut für Statistik

Persistente Zeitreihen

Thema:

Non-stationary semiparametric long memory estimation

Prüfer: Prof. Dr. P. Sibbertsen

vorgelegt von:

Name:	Sahra Ghalebikesabi
Studiengang:	Wirtschaftswissenschaft
Fachsemester:	7
Matrikelnummer:	3149150

Contents

List of Tables

List of Figures

1 Introduction

Long Memory processes are found in many major economic applications as discussed in Baillie (1996). According to Beran et al. (2016) a time series $\{X_t\}_{t=1}^{n}$ exhibits long range dependence or long memory, if its spectral density $f_X(\lambda)$ can be written as $f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d}$ (1) with $d \in (0, \frac{1}{2})$ and L_f defining an at zero slowly varying function. In this case, the spectral density is unbounded at $\lambda = 0$ and hence the autocorrelation function is not summable corresponding to Geweke and Porter-Hudak (1983).

In empirical applications, the spectral density is replaced by the periodogram, the non-parametric estimate of $f_X(\lambda)$. For $w(\lambda_j)$ being the discrete Fourier transform at the frequency $\lambda_j = 2\pi j/n$ the j^{th} Fourier periodogram is than defined for all $j \in \{1, ..., [(n-1)/2]\}$ as

$$I_X(\lambda_j) = |w(\lambda_j)|^2 = \frac{1}{2\pi n}\sum_{t=1}^{n} \hat{\gamma}(t)exp(i\lambda_j t) \quad \text{with}$$

$$w(\lambda_j) = \frac{1}{(2\pi n)^{1/2}}\sum_{t=1}^{n} X_t exp(-i\lambda_j t) \quad \text{and} \quad \hat{\gamma}(\tau) = \frac{1}{n}\sum_{t=1}^{n-|\tau|}(X_t - \bar{X})(X_{t+|\tau|} - \bar{X})$$

Further, Hurvich and Ray (1995) define the normalised periodogram as $I_X(\lambda_j)/f_X(\lambda_j)$.

Equation (1) shows that d significantly determines the shape of the spectral density at the origin conferring to Boutahar and Khalfaoui (2011), which raises the question about its estimation. For that purpose, Robinson (1995a) defines an approximation of the spectral density for $\lambda \to 0+$ by $f_X(\lambda) \sim L\lambda^{-2d}$ (2) with L being a positive constant and $\sim$ meaning that the ratio of the left and the right side of the equation tends to one. As the spectral density is only defined in a small region around the origin, this representation is semiparametric. Building upon this model, two estimators were designed for $d \in (-\frac{1}{2}, \frac{1}{2})$, which will be presented in the next section.

For $d \geq \frac{1}{2}$ Robinson (1995a) notes, that the integrability of (2) fails, exposing $\{X_t\}_{t=1}^{n}$ as a non-stationary time series, while $f_X(\lambda)$ is not invertible for $d \leq -\frac{1}{2}$. In section 3 and 4 modified estimators dealing with non-stationary time series will be introduced and discussed, while section 5 is devoted to their comparison in a simulation study.

In the following, we only regard general non-stationary long memory time series $\{X_t\}_{t=1}^{n}$ which get stationary after a finite number of differentiations. The assumptions introduced by Velasco (1999b) are imposed to infer properties of the spectral density in the non-stationary case similar to those of the spectral density of stationary time-series under the assumptions of Robinson (1995a,b).

2 Original Long Memory Estimators

Subsequently, two known estimators of the memory parameter will be presented and their asymptotic properties will be discussed.

2.1 Log Periodogram Estimator

As Velasco (1999b) states, taking logs of (2), adding $\log I(\lambda_j)$ and rearranging delivers the linear expression $\log I(\lambda_j) \approx \log G - 2d \log \lambda_j + \log \frac{I(\lambda_j)}{f_X(\lambda_j)}$ (3) with a slope of $-2d$. Building upon this idea, Geweke and Porter-Hudak (1983) developed the Log periodogram estimator (LPE) $\hat{d}_{GPH}$ also called the GPH-estimator through a simple regression such that with m being the bandwidth number, indicating the greatest Fourier frequency employed, it follows

$$\hat{d}_{GPH} = (\sum_{k=1}^{m} \Lambda_k^2)^{-1} (\sum_{k=1}^{m} \Lambda_k \log I_X(\lambda_k)) \quad \text{with} \quad \Lambda_k = -2\log \lambda_k - \{1/(m-1)\} \sum_{k=1}^{m} -2\log \lambda_k$$

According to Hurvich and Ray (1995), for $-\frac{1}{2} < d < \frac{1}{2}$ it holds true that $m^{1/2}(\hat{d}_{GPH} - d) \overrightarrow{d} N(0, \frac{\pi^2}{24})$ for $m/n \to 0$ and $(m \log m)/n \to 0$ as $n \to \infty$. Kim and Phillips (2006) showed that the LPE is consistent for $d \leq 1$, while it converges in probability to unity for $d > 1$. They did not impose a gaussian times series, but strong restrictions on the number of ordinates included in the regression. According to Kim and Phillips (2006), who sum up the work of Cheung and Lai (1993), the deciding difference lies in the fact that the constituent innovations get persistent for $d > 1$, such that the time series is not mean-reverting anymore.

2.2 Local Whittle Estimator

An asymptotically better behaving estimator is obtained if the log likelihood is maximised or $-\frac{1}{n}$ times the log likelihood is minimised. The latter is represented by the Whittle approximation of the Gaussian likelihood by Whittle (1953). This minimisation equals the optimisation of $Q(G,d) = \frac{1}{m} \sum_{j=1}^{m} \{logG\lambda_j^{-2d} + \frac{I_X(\lambda_j)}{G\lambda_j^{-2d}}\}$ for $d \in \Theta = [\nabla_1, \nabla_2]$ and $0 < G < \infty$. Building upon this, rearranging delivers the local whittle estimator or also called the Gaussian semiparametric estimator (GSE) which is given by

$$\hat{d}_{LW} = arg \min_{d \in \Theta} R_m(d) \quad \text{with} \quad R_m(d) \quad = \log \hat{G}_m(d) - d(\frac{2}{m} \sum_{j=1}^{m} \log \lambda_j) \text{ and } \hat{G}_m(d) = \frac{1}{2} \sum_{j=1}^{m} \frac{I_n(\lambda_j)}{\lambda_j^{-2d}}$$

While Robinson (1995b) proved its asymptotic properties for $d \in (-\frac{1}{2}, \frac{1}{2})$, Velasco (1999a) deduced the same under similar assumptions for a wider range of d. As stated in Dahlhaus

(1988) the minimisation of the Whittle approximation can be interpreted as the minimisation of the information divergence between the spectral density and the periodogram. Thus, as $I_X(\lambda_j)$'s bias increases with d, the GSE could only be shown to be consistent for $d \in (-\frac{1}{2}, 1)$ if $\frac{1}{m} + \frac{m}{n} \to 0$ as $n \to \infty$, whereas its asymptotical normality with

$$m^{1/2}(\hat{d}_{GSE} - d) \overrightarrow{d} N(0, \frac{1}{4}) \quad \text{with} \quad \frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \quad \text{as} \quad n \to \infty$$

could only be inferred for $d \in (-\frac{1}{2}, \frac{3}{4})$.

Consequently, the GSE seems to be more efficient than the LPE, although the GPH-estimator as the result of a regression is easier to calculate than the LW-estimator which builds upon numerically time consuming optimisation. However, the asymptotic properties of both estimators can be further improved by pooling and trimming. These methods and particularly their application on the LPE will be introduced in the next subsection.

2.3 Trimming and Pooling

Robinson (1995a) introduced a modified GPH estimator $\hat{d}^t_{GPH}$ with improved asymptotic properties gained through pooling (using the sum of adjacent periodogram ordinates) and trimming (dropping the first Fourier frequencies). While the periodogram's variance is reduced by pooling according to Faÿ et al. (2009), corresponding to Velasco (1999a,b) trimming reduces the autocorrelation between w_j and w_k and thus the estimator's bias. The intuition behind trimming is that the normalised periodogram $I_X(\lambda)/f_X(\lambda)$ and thus the error term in the log-periodogram regression as seen in (3) is neither asymptotically independent nor identically distributed, when j is fixed for increasing n as stated in Robinson (1995a). With $l < m$ being the trimming number, indicating the smallest frequency employed, and J the pooling number, indicating the number of periodogram ordinates that are summed up, Robinson (1995a) derives for $(m-l)/J$ being an integer the expression

$$Y_k^{(J)} = \log(\sum_{j=1}^{J} I_X(\lambda_{k+j-J})), \quad k = l+J, l+2J, \ldots, m \quad \text{leading to}$$

$$\hat{d}^t_{GPH} = (\sum_k \Lambda_k^2)^{-1}(\sum_k \Lambda_k Y_k^{(J)}) \quad \text{with} \quad \Lambda_k = -2\log\lambda_k - \{J/(m-l)\}\sum_k -2\log\lambda_k.$$

While Robinson (1995a) proved the unbiasedness and consistency of this estimator for $d \in (-\frac{1}{2}, \frac{1}{2})$, Velasco (1999b) shows that the asymptotic properties of the DFTs in the non-stationary case resemble exactly those in the stationary case under certain assumptions, inter alia gaussian innovations. As Velasco (1999b) states, for $d < 1$ the bias of the periodogram decreases with j, such that consistency only holds, if the number of used Fourier frequencies increases with n. Accordingly, he showed that $\hat{d}^t_{GPH}$ is consistent for $\frac{\log m}{l^{2(1-d)}} + \frac{1}{m-l} + \frac{(\log n)^2}{m} + \frac{m}{n} \to 0$ as $n \to \infty$. Further under even stronger assumptions Velasco (1999b) deduces for $d \in [\frac{1}{2}, \frac{3}{4})$ with $\psi'(x) = (d/dx)\log\Gamma(x)$

$$m^{1/2}(\hat{d}-d)\overrightarrow{d}N(0, \frac{J}{4}\psi'(J)) \quad \text{if} \quad \frac{m^{1/2}\log m}{l^{2(1-d)}} + \frac{l(\log n)^2}{m} + \frac{m^{1+1/2*\alpha}}{n} \to 0 \quad \text{as} \quad n \to \infty.$$

3 Tapered Estimators

Outside of the ranges analysed in the section before, the estimates of d converge more slowly as the expectation of the periodogram converges more slowly as well. This problem is solved by Velasco (1999b,a) with tapering.

3.1 Introduction to Tapering

Corresponding to Velasco (1999b), a taper is a smoothing function assigning each observation X_t a weight h_t, leading to a DFT as

$$\begin{aligned} w^T(\lambda_j) &= \frac{1}{\sqrt{2\pi \sum_{t=1}^n h_t^2}} \sum_{t=1}^n h_t X_t exp(i\lambda_j t) \\ &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \frac{h_t\sqrt{n}}{\sqrt{\sum_{t=1}^n h_t^2}} X_t exp(i\lambda_j t) \end{aligned}$$

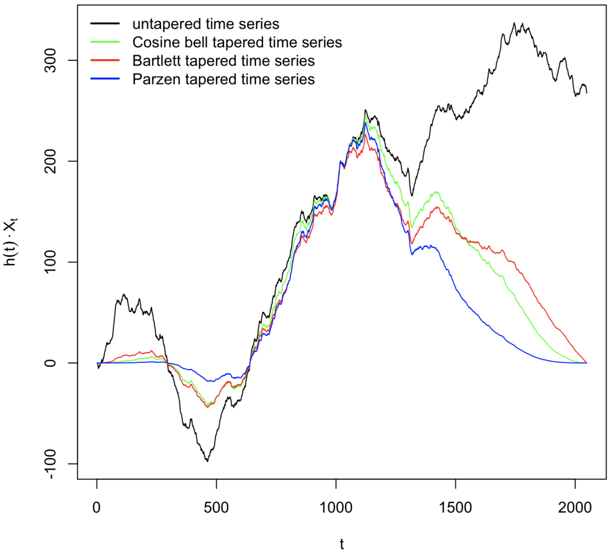

Figure 1: Tapered $ARFIMA(0, 1.85, 0)$ for $n = 2048$

Hence, tapering a time series equals multiplying each observation X_t with $\frac{h_t\sqrt{n}}{\sqrt{\sum_{t=1}^n h_t^2}}$. For the original time series, it holds $h_t \equiv 1$ for all t. The taper usually takes values around 1 for the central part of the data, and decays smoothly to 0 at both the beginning and the end of the sample, weighting down the influence of the low frequencies and thus of non-stationarities as explained in Sibbertsen (2004). As seen in Figure 1, this leads to a smoothing of the time series, dampening peaks and troughs and changing the course of the time series especially for boundary values such that the raw and tapered data are hardly comparable.

Faÿ et al. (2009) state that tapering was primarily used in nonparametric spectral analysis of

short memory time series for a reduced bias due to frequency domain leakage, where part of the spectrum leaks into adjacent frequencies (for further reference see Tukey (1967) or Dahlhaus (1983)). For non-stationary time series, Velasco (2000) proposed to taper the data to reduce the leakage in the periodogram ordinates from the zero frequency and to pool the periodogram to obtain better behaved regressors. In Dahlhaus (1988), this improvement is shown to be due to the fact, that the tapered periodogram's bias converges faster to 0 with j increasing, whence Dahlhaus et al. (1997) follow that tapering reduces the leakage and bias of the periodogram introduced by the non-stationary behaviour of the boundaries.
For an exemplary taper, the effects on the LPE will be analysed next.

3.2 Cosine Bell Taper

According to Velasco (1999b), the asymmetric version of the cosine bell's taper sequence is defined by $h_t^{CB} = \frac{1}{2}(1 - cos[2\pi t/n])$, which is graphically represented in Figure 2.
Velasco (1999b) refers to Bloomfield (2004) and Percival and Walden (1993), when inferring, that the tapered Fourier transform can be written as $w^T(\lambda_j) = \frac{1}{\sqrt{6}}[-w(\lambda_{j-1}) + 2w(\lambda_j) - w(\lambda_{j+1})]$, a linear combination of untapered Fourier transforms for $2 \leq j \leq n-2$. Consequently, as Velasco (2000) states, tapering destroys the orthogonality relations between DFTs, as the adjacent tapered Fourier transforms have two common components. To avoid correlation to be able to use Robinson (1995a)'s results, only every third Fourier frequency can be included in the calculation of $\hat{d}_{cb}^T$ leading to its variance increasing by a factor of 3 as presented in Velasco (1999b). The GPH is than given as before, but $Y_k^{(J)}$ is replaced by $Y_{k,g}^{(T,J)} = log(\sum_{j=1}^{J} I^T(\lambda_{k+g(j-J)}))$ for all $k = l + gJ, l + 2gJ, ..., m$, $I^T(\lambda_j) = |w^T(\lambda_j)|$ and $g = 3$.

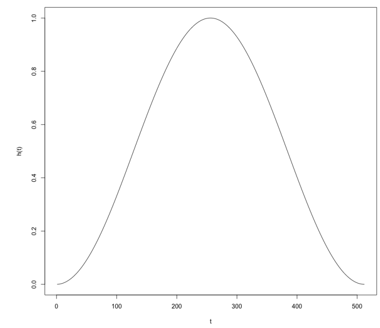

Figure 2: Course of the cosine bell taper

As a linear combination of shift-invariant functions, w^T is again shift-invariant. Thus the tapered periodogram is shift invariant. Otherwise, replacing the mean by its estimate would destroy the estimators' performance according to Beran et al. (2016). Building on this representation, Velasco (1999b) infers, that the tapered periodogram has improved asymptotic properties compared to the usual periodogram as the bias of the periodogram on the tails is reduced. Velasco (1999b) proves that the expectation of the normalised periodogram diverges for $d < \frac{3}{2}$ at any

λ_j, such that

$$m^{1/2}(\hat{d}^T_{GPH} - d)\vec{d}N(0, \frac{3J}{4}\psi'(J)) \quad \text{only holds for} \quad \frac{m^{1/2}}{l} + \frac{l(\log n)^2}{m} + \frac{m^{1+1/2\alpha}}{n} \quad \text{as} \quad n \to \infty$$

and $d \in [\frac{1}{2}, \frac{3}{2})$, $\mu = 0$, where μ is the expectation of the time series and α being such, that the spectral density of the innovations u can be written as $f_u(\lambda) = G\lambda - 2(d-1) + O(\lambda^{-2(d-1)+\alpha})$ as $\lambda \to 0+$. Thanks to tapering, these asymptotic properties for non-stationary time series hold under milder conditions than imposed for $\hat{d}^t_{GPH}$. What is more, Velasco (1999b) found a general class of tapers with even improved asymptotic properties which shall be presented in the following.

3.3 Taper of Order p by Velasco (1999a,b)

Velasco (1999b) calls $\{h^{Vel}_t\}^n_{t=1}$ a sequence of data tapers of order p, if

- $\sum^n_{t=1} h^{Vel^2}_t = bn$ holds for a function $b(n) = b$ for any positive n with $0 < b < \infty$
- For $\frac{n}{p} = N$, the Dirichlet kernel D^T_p can be written as

  $$D^T_p(\lambda) \equiv \sum_{t=1}^{n} h^{Vel}_t exp(i\lambda t) = \frac{\alpha(\lambda)}{n^{p-1}}(\frac{sin[n\lambda/2p]}{sin[\lambda/2]})^p,$$

  with $\alpha(\lambda)$ being a positive complex function with a bounded modulus and $p-1$ in modulus bounded derivatives as n increases for $\lambda \in [-\pi, \pi]$.

Taper of this definition have the property $\sum^n_{t=1} h^{Vel}_t(1 + t + ... + t^s)exp(i\lambda_j t) = 0$ for $s \leq p-1$, which allows to estimate the memory parameter of a time series invariantly to a polynomial deterministic trend of order s, as it would be removed in the calculation of $w^T(\lambda_j)$. Although this equation also holds true for the cosine bell taper for $s = 0$, it does not belong to this class of taper sequence.

An example for $n = 4N$ with N being an integer is the Parzen window

$$h^{Parzen}_t = \begin{cases} 2(1 - |[2t-n]/n|)^3, & 1 \leq t \leq N \quad or \quad 3N \leq t \leq 4N, \\ 1 - 6(([2t-n]/n)^2 - |[2t-n]/n|^3), & N < t < 3N \end{cases}$$

for $j = 4, 8, ..., n-4, s = 3$ and $p = 4$ as presented by Velasco (1999b). Zhurbenko and Trush (1979) present a general class of Kolmogorov data tapers of order p for any $p = 1, 2, ...$ For

$p = 4$ the taper resembles the Parzen window and for $p = 2$ the Bartlett's or also triangular window with $h_t^{Bartlett} = 1 - |\frac{t - \frac{n-1}{2}}{\frac{n-1}{2}}|$. According to Velasco (1999b), the LPE is adjusted as follows:

$$\hat{d^T_{GPH,p}} = (\sum_k \Lambda_{kp}^2)^{-1}(\sum_k \Lambda_{kp} Y_{kp}^{(T,1)}) \quad \text{and} \quad Y_{kp}^{(t,1)} = log I_p^T(\lambda_{kp}), k = l, l+\eta, l+2\eta, ..., m\eta,$$

such that still m observations (as many observations as in the untapered case) are included for fixed l and m. Thus, for $p \geq \lfloor d + \frac{1}{2} \rfloor + 1$, Gaussian innovations and further assumptions as in Velasco (1999b), for $n \to \infty$ it follows

$$m^{1/2}(\hat{d^T_{p,GPH}} - d)\vec{d}N(0, \pi^2/24) \quad \text{if} \quad \frac{(m\eta)^{1/2}}{l^{max\{1,\alpha\}}} + \frac{m^{1/(2p-1)}}{\eta} + \frac{(m\eta)^{1+1/2\alpha}}{n} + \frac{l(\log n)^2}{m\eta} \to 0$$

What is more, Nouira et al. (2009) showed in a Monte Carlo study, that the optimal standard deviation is obtained for $p = [d + \frac{1}{2}] + 1$. For $\mu = 0$, the above distribution holds already for $p > d + \frac{1}{4}$ with $(m\eta)^{1/2} l^{2(d-p)} \log m \to 0$ as $n \to \infty$. Velasco (1999b) infers that $\hat{d^T_{GPH,p}}$ is consistent and asymptotically normal even for $d \leq -\frac{1}{2}$ for p big enough.

Under slightly stronger assumptions Velasco (1999a) defines the modified GSE with a taper of order p by

$$(\hat{G}_p, \hat{d^T_{LW,p}}) = arg \min_{0<G<\infty, d\in\Theta} Q_p(G,d) \text{ for } Q_p(G,d) = \frac{p}{m}\sum_{j=1}^{m/p}\{log(G\lambda_{jp}^{-2d}) + \frac{I_p^T(\lambda_{jp})}{G\lambda_{jp}^{-2d}}\} \text{ and}$$

$$\hat{d^T_{LW,p}} = arg \min_{d\in\Theta} R_p(d) \text{ where } R_p(d) = log\hat{G}_p(d) - 2d\frac{p}{m}\sum_{j=1}^{m/p} log(\lambda_{pj}), \hat{G}_p(d) = \frac{p}{m}\sum_{j=1}^{m/p}\lambda_{jp}^{2d} I_p^T(\lambda_{jp})$$

for $\Theta = [\nabla_1, \nabla_2]$ with $\frac{1}{2} < \nabla_1 < \nabla_2 < d^*$, $d^* < p + \frac{1}{2}$ and m/p being an integer. The properties for the periodogram in the non-stationary case are only equivalent to the stationary case, if no more than the frequencies whose index are a multiple of p are included which explains the sums in the formulas as discussed in Velasco (1999a). As before the periodogram is an unbiased estimate of the spectral density if j is growing slowly with n for a certain p corresponding to Velasco (1999a,b). Under these assumptions $\hat{d^T_{LW,p}}$ converges in probability against d for $d \in [\nabla_1, \nabla_2], \nabla_1 > \frac{1}{2}$ and $p \geq \lfloor \nabla_2 + \frac{1}{2} \rfloor + 1$ if $\frac{1}{m} + \frac{m}{n} \to 0$ as $n \to \infty$.

Additionally, Velasco (1999a) shows under further assumptions that for

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \quad \text{as} \quad n \to \infty \quad \text{it holds that} \quad m^{1/2}(\hat{d^T_{p,LW}} - d)\vec{d}N(0, \frac{1}{4}p\Phi)$$

$$\text{with} \quad \Phi = \lim_{n\to\infty}(\sum_{t=1}^{n}(h_t^{Vel})^2)^{-2}\sum_{k=0}^{(n-p)/p}\{\sum_{t=1}^{n}(h_t^{Vel})^2 cos(t\lambda_{kp})\}^2 \text{ almost always greater than 1.}$$

The variance of this estimators exceeds that of their untapered counterpart because of the correlation of the tapered periodogram ordinates which had not been orthogonalised as Velasco (1999a) states.

3.4 Differentiation

According to Robinson (1995a), the memory parameter d^d of a differenced data set ΔX_t with $\Delta X_t = X_t - X_{t-1} \forall t \in \{1,...,T\}$ is one unit smaller than that of the original data. Considering this, Robinson (1995a) proposed to differentiate an observed non-stationary time series z times until the memory parameter $d^{d(z)}$ of the z^{th} differenced time series lies in $(-\frac{1}{2},\frac{1}{2})$ such that it can be estimated consistently with the GPH estimator. The memory parameter d is then given by $\hat{d} = \hat{d}^{d(z)} + z$. But as Hurvich and Ray (1995) showed the GPH estimator is in fact not invariant to first differences in contrast to the tapered GPH-estimator.
Further, differencing the data z times assures that the memory estimate is invariant with respect to time trends of order z corresponding to Hurvich and Chen (2000).
However, the possibility of overdifferencing, i.e. obtaining $d^{d(z)} < -\frac{1}{2}$, endangers the asymptotic properties of the estimator as non-invertible processes could lead to detrimental leakage and thus introduce estimation bias to cite Hurvich and Ray (1995). Hence, in order to choose an appropriate z additional information on the size of d would be needed. Still, there exist tapered estimators which rely on differentiation in order to enable more efficient estimation of the memory parameter for non-stationary time series, one of which shall be presented next.

3.5 Taper of Order p by Hurvich and Chen (2000)

In Hurvich and Chen (2000) a new class of tapers for the LW estimator with preferable asymptotic properties is presented. The new class of tapers is defined by

$$h_t^{HUC} = 0.5[1 - exp\frac{i2\pi(t-1/2)}{n}], \quad \text{which delivers} \quad w_j^T = \sqrt{2}\{0.5w_j - 0.5w_{j+1}exp(-i\pi/n)\}.$$

The data is first differenced, such that the new memory parameter is $d^d = d - 1$. The GSE is then applied on the tapered differenced data. Although there is a loss of efficiency through differencing, the estimator is valid in the noninvertible spectrum. Under similar assumptions as Velasco (1999a), Hurvich and Chen (2000) follow the consistency of the estimator for $\frac{1}{m} + \frac{m}{n} \to 0$ and even its asymptotic normality

$$m^{1/2}(\hat{d}_{GSE}^{HUC} - d)\overrightarrow{d}N(0, 15/40) \quad \text{for} \quad \frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0 \quad \text{as} \quad n \to \infty$$

for $d \in (-\frac{1}{2}, \frac{3}{2})$. According to Faÿ et al. (2009), this results from the fact that Hurvich and Chen (2000)'s taper mixes together adjacent periodogram ordinates, such that Fourier frequencies are mixed in each tapered periodogram ordinate.

Hurvich and Chen (2000) define their taper of order p by $\{(h_t^{HUC})^p\}_t$. The tapers of order p of Hurvich and Chen (2000) and Velasco (1999a) share similar properties as stated in Faÿ et al. (2009). Since Hurvich and Chen (2000)'s taper is complex-valued, it does not fall into the class regarded by Velasco (1999b). Another difference is that tapers of Velasco (1999a) are not exactly orthogonal according to Arteche and Velasco (2005). What is more, the linear representation of the DFT indicate the estimator's shift-invariance in mean without imposing assumptions as strict as in Velasco (1999a) according to Faÿ et al. (2009). Hurvich and Chen (2000) proclaim that the tapered GSE's variance as introduced in Velasco (1999a) cannot undercut $\frac{21}{40m}$ with Kolmogorov tapering while their estimator has only a variance of $\frac{15}{40m}$.

For invariance in respect to polynomial trends of order p, the time series has to be differenced p times and its DFTs to be tapered by$\{(h_t^{HUC})^p\}_t$. This is achieved in Hurvich and Chen (2000) under less strict assumptions than in Velasco (1999b). Further, HUC's taper's asymptotic variance is shown to be approximately $\frac{(p\pi/2)^{1/2}}{4m}$. In comparison, Velasco (1999b) had to use tapers of order $p+1$ to estimate invariantly to polynomial trends of order p and omit all frequencies whose index were not a multiple of $p+1$, which led to an asymptotic variance of at least $\frac{p}{4m}$. Regardless of that, an evident disadvantage is the need of an upper bound of d and the danger of overdifferencing.

4 Exact Local Whittle Estimation

We consider the fractional process X_t generated by the model $(1-L)^d X_t = u_t \mathbf{1}_{t\geq 1}$ with $(d)_k = d \cdot (d+1) \cdot \ldots \cdot (d+k-1)$ leading to $X_t = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} u_{t-k}$. The likelihood of the stationary innovation u_t is then given by Shimotsu (2010a) as $\sum_{j=1}^{m} \log f_u(\lambda_j) + \sum_{j=1}^{m} \frac{I_u(\lambda_j)}{f_u(\lambda_j)}$. If $d \in (-\frac{1}{2}, \frac{1}{2})$, $I_u(\lambda_j)$ can be approximated by $\lambda_j^{2d} I_X(\lambda_j)$. Building upon this, the LW-estimator can be obtained. For $d > \frac{1}{2}$ this approximation does not hold anymore. Thus, Shimotsu et al. (2005) propose to use instead the exact equality $I_u(\lambda_j) = I_{\Delta^d X}(\lambda_j)$, which is why the estimator basing on this idea is called the exact local whittle estimator. Intuitively, the ELW is obtained out of the GSE by replacing the approximation $I_u(\lambda_j) \approx \lambda_j^{2d} I_X(\lambda_j)$ by the exact equation $I_u(\lambda_j) = I_{\Delta^d X}(\lambda_j)$, which results directly from the imposed model. Hence, the ELW also possesses the same asymptotic properties for an even wider range of d.

Using this in the likelihood, adding Jacobian $\sum_{j=1}^{m} \log |D_n(exp(i\lambda_j);d)|^{-2}$ and approximating $f_u(\lambda_j) \sim G$ and $|D_n(exp(i\lambda_j);d)|^2 \sim \lambda_j^{2d}$ delivers

$$Q_m(G,d) = \frac{1}{m}\sum_{j=1}^{m}[\log G\lambda_j^{-2d} + \frac{1}{G}I_{\Delta^d x}(\lambda_j)] \quad \text{and} \quad \Delta^d X_t = (1-L)^d X_t = \sum_{k=0}^{t}\frac{(-d)_k}{k!}X_{t-k}$$

To obtain the estimates of G and d, Shimotsu et al. (2005) minimise $Q_m(G,d)$, such that

$$(\hat{G}_{ELW}, \hat{d}_{ELW}) = arg \min_{G\in(0,\infty), d\in[\Delta_1,\Delta_2]} Q_m(G,d) \quad \text{and} -\infty < \Delta_1 < \Delta_2 < \infty.$$

Resolution according to $\hat{d}_{ELW}$ bears $\hat{d}_{ELW} = arg\min_{d\in\Delta_1,\Delta_2} R(d)$, where

$$R(d) = \log \hat{G}(d) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j \quad \text{and} \quad \hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m} I_{\Delta^d x}(\lambda_j).$$

Under the same assumptions as for the local whittle estimator in Robinson (1995b) and

$$\frac{1}{m} + \frac{m(\log m)^{1/2}}{n} + \frac{\log n}{m^\gamma} \to 0 \quad \text{for any } \gamma > 0 \quad \text{as well as} \quad \Delta_2 - \Delta_1 \leq \frac{9}{2}$$

Shimotsu et al. (2005) conclude the consistency of the ELW-estimator, if the mean is known. The last assumption hereby ensures that $R(d) - R(d_0)$ for d_0 being the true memory parameter converges to a non-random function.

Under some further assumptions introduced in Shimotsu et al. (2005), which are similar, but slightly stronger than the analogue assumptions for the LW, Shimotsu et al. (2005) deduce

$$m^{1/2}(\hat{d}-d)\overrightarrow{d}N(0,\frac{1}{4}) \quad \text{as} \quad n\to\infty \quad \text{for} \quad \frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} + \frac{\log n}{m^\gamma} \to 0 \quad \text{for any} \quad \gamma > 0.$$

The advantage of the need of neither differencing nor tapering is confronted by the disadvantage that d could only be shown to be a consistent estimator if its admissible values are beforehand restricted to an interval of length $\frac{9}{2}$.

Additionally, according to Faÿ et al. (2009), the ELW estimator is neither time-shift invariant nor invariant upon addition of a constant in the data. Further if the mean is not known and has to be estimated, the asymptotic properties of the ELW estimator deter (for further reference see A.2 in the appendix). However, as $n \to \infty$ there is no need to estimate μ_0 as in Shimotsu (2010b), as the signal on d by X_t^0 dominates that of μ_0 if $X_t = \mu_0 + X_t^0$.

5 Simulation Study

After asymptotical properties of the modified estimators were analysed, a Monte Carlo study was conducted in order to explore their finite sample behaviour.

5.1 Influence of d, m and n

For this reason, for each $d \in \{0.35, 0.55, 0.75, 1.25, 1.75, 2.75, 3.75\}$ 1000 replications of an $ARFIMA(0,d,0)$ were calculated for each sample size $n \in \{512, 1024, 2048\}$ and bandwidth $m \in (n^{0.5}, n^{0.6}, n^{0.7}, n^{0.8}, n^{0.868})$. In the following, the parameter $e = \log_n m$ instead of m is used to refer to the bandwidth parameter. In order to have comparable results, these estimates were selected similarly to those in Boutahar and Khalfaoui (2011). The inclusion of $n^{0.868}$ is based on Nouira et al. (2009), who argued that the optimal bandwidth of memory estimators rises for the non-stationary case. The samples were simulated with `fracdiff.sim` for $d \in (-0.5, 0.5)$ and afterwards integrated accordingly.

Thereupon, the pooled and trimmed GPH of Robinson (1995a) for $l = 3$ and J being the greatest divisor of $m - l$ smaller than $min\{(m-l)/6, 10\}$ was applied on the raw and the cosine bell (cb) tapered data. Further, for $\eta = 1$ and $l = 3$ the bartlett (b) and parzen (p) tapered LPE were calculated as well as the GSE of the untapered, cosine bell, bartlett, parzen and Hurvich and Chen (2000) (huc) tapered data. In order to calculate the latter estimates, the interval to be optimised over was restricted to $[0.25, 3.95]$. The choice of l roots in the remark of Arteche and Velasco (2005), who state that a tapering reduces the bias significantly for $l \geq p - 1$ which is less strict than in the untapered case. Moreover, simulations show that l can be chosen independently of n, as Arteche and Velasco (2005) argue. Last but not least, the ELW was calculated through an optimisation over $[-3.5, 3.5]$ in Matlab using code provided by Shimotsu. An exemplary output can be seen in the Table 1 whereas all results are presented in the appendix (A.3).

Evidently, the LPE obtains a bias of approximately $1 - d$ for $d > 1$ corresponding to the theoretical result of Hurvich and Ray (1995) that it converges to unity for this interval. As Velasco (1999b) showed, the GPH estimator seems consistent for $d < 1$ because of its relatively low bias which is smaller than that of all of its tapered counterparts. For $d > 1$ the bias of the tapered LPEs fall below that of the untapered LPE. Nonetheless, there exist obvious differences in the magnitude of the bias of the tapered estimators. While the bartlett tapered GPH undercuts the bias of the other tapered GPHs for $d < 1$, the cosine bell taper provides the smallest bias for $1 < d < 3$ and the parzen taper for $d = 3.75$ which goes along with Velasco (1999b)'s asymptotic theory for tapers of order p with $p \geq \lfloor d + \frac{1}{2} \rfloor + 1$. What is more, both tapered and untapered LPE estimators tend to have a positive bias for low values of d and a negative bias for greater values of d. As expected, the standard deviation for all values of d increases with

	GPH	cbGPH	bGPH	pGPH	LW	cbLW	bLW	pLW	hucLW	ELW
Bias for d=0.35	0.0004	0.0074	0.0010	0.0068	-0.0071	-0.0042	-0.0094	-0.0092	-0.0012	-0.0005
Sd for d=0.35	0.0341	0.0617	0.0486	0.0642	0.0252	0.0347	0.0376	0.0518	0.0311	0.0389
Bias for d=0.55	0.0039	0.0058	0.0039	0.0172	-0.0062	0.0012	-0.0120	-0.0120	-0.0022	0.0066
Sd for d=0.55	0.0348	0.0604	0.0486	0.0655	0.0264	0.0367	0.0369	0.0523	0.0303	0.0416
Bias for d=0.75	0.0161	0.0080	0.0071	0.0252	0.0004	-0.0028	-0.0135	-0.0149	-0.0012	-0.0011
Sd for d=0.75	0.0393	0.0597	0.0466	0.0622	0.0308	0.0350	0.0368	0.0526	0.0310	0.0390
Bias for d=1.25	-0.2111	0.0004	0.0160	0.0483	-0.2088	0.0069	-0.0175	-0.0209	0.0023	-0.0010
Sd for d=1.25	0.0737	0.0616	0.0465	0.0653	0.0699	0.0398	0.0361	0.0498	0.0296	0.0391
Bias for d=1.75	-0.7351	-0.0028	0.0377	0.0831	-0.7372	0.0301	-0.0090	-0.0286	0.0185	-0.0010
Sd for d=1.75	0.0762	0.0605	0.0535	0.0683	0.0912	0.0539	0.0431	0.0539	0.0362	0.0391
Bias for d=2.75	-1.7467	0.1978	-0.7140	0.2027	-1.7573	0.1487	-0.7495	-0.0285	0.1063	-0.0011
Sd for d=2.75	0.0349	0.1148	0.0796	0.0985	0.0511	0.0835	0.0956	0.0554	0.0737	0.0391
Bias for d=3.75	-2.7485	-0.7464	-1.7279	0.3643	-2.7656	-0.7289	-1.7740	0.0068	-0.6613	-0.0018
Sd for d=3.75	0.0117	0.0337	0.0338	0.0451	0.0255	0.0514	0.0481	0.0643	0.0837	0.0400

Table 1: Bias and standard deviation of memory estimates of an $ARFIMA(0,d,0)$ with $n = 2048$ and $m = 208 \approx n^{0.7}$

the tapering which results out of the necessary spacing of Fourier frequencies with $g = 3$ for the cosine bell taper, $p = 2$ for the bartlett taper and $p = 4$ for the parzen taper. Subsequently, the bGPH dominates with the smallest standard deviation, followed by the cosine bell tapered LPE. As before, the GSE seems to be unbiased for $d < 1$ with a lower standard deviation than the GPH, but converging to unity for $d > 1$ with a higher standard deviation than the LPE. Again, the higher order parzen taper optimises higher values of d well, whereas the lower order bartlett taper obtains its minimum bias for smaller memory parameters. As there are no frequencies omitted in the calculation of cbLW, but tapering still induces correlation of adjacent DFTs, the standard deviation of the cosine bell tapered GSE tends to exceed that of the untapered one. However, the bartlett and parzen tapered GSEs optimise an with p spaced objective function and thus provide a higher standard deviation with the latter varying more than the former. For $d \in \{0.75, 1.25\}$ the tapered LW according to Hurvich and Chen (2000) performs indeed more efficiently than the other tapered estimators and provides the smallest standard deviation for an even wider range. As stated by Beran et al. (2016), the asymptotic variance of the tapered LW-estimator undercuts that of the tapered GPH-estimator. Nonetheless, the variance increases for both respectively to the case without tapering. The ELW prevails over all other presented estimators with a bias near to zero (independently of the magnitude of d) and a constantly small standard deviation, although the range to be optimised over was wider than $\frac{9}{2}$.

Comparing these results with those in A.3 for other sample sizes and bandwidth parameters, it can be summed up that the bias and standard deviation typically decrease for higher sample sizes. Nonetheless, few expectations remain. For instance for $d \in \{0.35, 0.55, 0.75\}$ and $b = 0.868$ the parzen tapered LPE reaches its minimal bias for $n = 512$, which however does not indicate faster convergence as this gain is confronted by a higher standard deviation.

As expected, the standard deviation decreases for all d, n and estimators as more frequencies are included in the calculation. As m decreases, the short memory effects should vanish and hence the estimates should indicate the true long-memory dependence according to Taqqu et al. (1995). Anyhow, for decreasing m estimates get scattered and unreliable. However, the bandwidth minimising the bias of the estimates depends strongly on the estimators they are based on and is thus not uniform which is confirmed by Velasco (2000). For example, the ELW takes its minimal bias for $e = 0.6$, whereas the local whittle estimator of the untapered and bartlett or parzen tapered observations performs best for $e = 0.7$. The LW tapered with Hurvich and Chen (2000)'s taper or the cosine bell taper can be minimised for $e = 0.8$ just like the bartlett tapered GPH. Nouira et al. (2009)'s simulation result, that $e = 0.868$ delivers the smallest possible bias for the GPH estimator is reinforced in this simulation even for the cosine bell and parzen tapered LPEs, although the GPH attains its optimal bandwidth of $e = 0.8$ for $n = 2048$, which goes along with Hurvich and Ray (1995)'s theoretical results. Although Geweke and Porter-Hudak (1983) proposed $n^{0.5}$ as a bandwidth number, it is evident that all estimators performed worst for this m.

Last but not least, by comparing the bias of each estimator for $d = 0.75$ and $d = 1.75$ the inefficiency of Robinson (1995a)'s suggestion to difference a time series until it is stationary to estimate its memory parameter becomes evident for the untapered case as the difference in bias is largely negative. Notwithstanding, tapering data produces estimates which are more nearly invariant to first-differencing as seen by the difference in bias near to zero.

5.2 Influence of Trends

Additionally, for $d = 0.75, e = 0.7$ and $n = 2048$ the shift 0.5 as well as the linear polynomial trend $0.5 \cdot t$ were respectively added to the data and estimated by the estimators as before with the exception that the method of differentiation was used for the GPH when applied on the time series with trend. The results are represented in the following Table 2. As expected, all of the LW estimators and the parzen and bartlett tapered LPE handle to estimate the memory

	GPH	cbGPH	bGPH	pGPH	LW	cbLW	bLW	pLW	hucLW	ELW
Bias shift	0.4752	-0.4659	0.0459	0.0571	0.0157	0.0219	-0.0031	-0.0066	-0.0161	0.0069
Sd shift	0.0858	0.1794	0.0952	0.0930	0.0519	0.0673	0.0733	0.1090	0.0583	0.0480
Bias trend	-0.0881	-0.1518	0.0459	0.0223	0.2483	0.8149	-0.0031	-0.0066	-0.0267	0.2618
Sd trend	0.0694	0.1869	0.0952	0.1116	0.0016	0.0221	0.0733	0.1090	0.0449	0.0034

Table 2: Bias and standard deviation of memory estimates of an $ARFIMA(0,0.75,0)$ with $n = 2048$ and $m = 208 \approx n^{0.7}$

parameter invariantly to shifts, but with a loss in standard deviation which is evidently greater for the GPH estimators than for the LW ones. While the increase in bias for the GPH estimate could be expected, the high bias of the cosine bell taper is surprising as the cosine bell tapered DFTs could be presented as a linear combination of untapered DFTs. However, as its asymptotic properties were derived by Velasco (1999b) using $\mu = 0$, the now positive mean of the process could explain the cbGPH's deteriorated behaviour.

In contrast to that, the asymptotic properties of the estimators deter significantly if a trend is added. Corresponding to the theoretical results of Velasco (1999a,b) concerning the LW and the cosine bell taper the bias of the cbGPH, LW and cbLW are significantly different to zero. This holds also true for the ELW as criticised by Abadir et al. (2007). However, the Kolmogorov tapers with $p = 2$ and $p = 4$ should estimate d invariantly to trends of order one according to Velasco (1999a,b) which also seems to hold true. Further, if the observations are differenced twice and tapered by the taper of order 2 of Hurvich and Chen (2000), the bias of the LW deceeds that of almost every other estimator. As Velasco (1999b) mentions, the cosine bell taper behaves like $p = 1$ concerning its treatment of trends and thus both estimators building upon this taper show up an increase in bias. Finally, there is no uniform change in the behaviour of the standard deviation through adding a trend.

5.3 Influence of Short Memory Components

In the end, 1000 replications of two different $ARFIMA(1,d,0)$ and $ARFIMA(0,d,1)$ were simulated, setting the included short memory component equal to both 0.4 and 0.8 according to the set up of Boubaker et al. (2017). Table 3 presents the results. Apparently, the standard deviation and thus the convergence rate of the estimators does not depend on the existence of short memory components. However, short memory components introduce noise into the estimation of the memory parameter. Hence, the greater the AR component gets, the more does the bias increase

	GPH	cbGPH	bGPH	pGPH	LW	cbLW	bLW	pLW	hucLW	ELW
a=0.4	0.15(0.05)	0.11(0.10)	0.09(0.07)	0.11(0.10)	0.06(0.04)	0.06(0.05)	0.05(0.06)	0.04(0.08)	0.04(0.05)	0.06(0.04)
a=0.8	0.45(0.05)	0.44(0.10)	0.39(0.07)	0.40(0.10)	0.33(0.05)	0.36(0.06)	0.36(0.06)	0.37(0.09)	0.35(0.05)	0.35(0.04)
b=0.4	0.03(0.06)	0.05(0.10)	-0.02(0.07)	0.01(0.10)	-0.03(0.05)	-0.03(0.05)	-0.05(0.06)	-0.06(0.08)	-0.06(0.05)	-0.04(0.04)
b=0.8	-0.31(0.06)	-0.27(0.10)	-0.32(0.07)	-0.27(0.10)	-0.28(0.05)	-0.25(0.07)	-0.31(0.06)	-0.34(0.08)	-0.31(0.05)	-0.28(0.05)

Table 3: Bias and corresponding standard deviation of memory estimates of an ARFIMA model with $n = 2048$, $m = 208 \approx n^{0.7}$ and $d = 0.75$ and AR component a or MA component b

likewise to the results of Boubaker et al. (2017) and Hurvich and Chen (2000), whereby this augmentation does not depend on the type of estimator. Through inclusion of an MA component, negative bias is introduced depending on the magnitude of the included parameter. As the GPH estimators attain a positive bias without short memory components, the bias decreases absolutely against expectation when an MA component with $b = 0.4$ is added. Anyhow, for $b = 0.8$, the triggered negative bias strongly exceeds the positive bias destroying the pleasing properties of the estimators.

6 Conclusion

Once and for all, the higher d is and accordingly the steeper the spectral density gets for $\lambda \to 0+$, the poorer are the finite sample approximations conferring to Velasco and Robinson (2000). However, as Hurvich and Chen (2000) noted, it is sufficient for many empirical applications to have an efficient estimator for $d < \frac{3}{2}$. For $d < -\frac{1}{2}$ tapering can also be used for non-invertible processes as Hurvich and Chen (2000) and Velasco (1999b) argue. Anyhow, Zhurbenko and Trush (1979) remark, tapering data only benefits the estimation for certain data windows and certain ds. Thus, it is necessary to know an approximate range of d in order to choose the appropriate estimator.

What is more, all presented estimators build upon the periodogram as an estimator of the spectral density. Even though, Tschernig (2013) notes that the leakage effect concerning the estimation of the spectral density for ending time series leads to worse results than theoretically assumed. Thus, other estimation methods could offer an attractive alternative as stated in Velasco (1999a); Faÿ et al. (2009) and Boubaker et al. (2017).

A Appendix

There is a wide range of modified estimators for the non-stationary spectrum, some of which shall be presented now.

A.1 Exact Log-Periodogram Regression

Hurvich et al. (1998) show that the periodogram suffers from asymptotic relative bias which decreases for j at a given Fourier frequency w_j. To eliminate the bias in the non-stationary case, Kim and Phillips (2006) suggest to add a correction term to any DFT which leads to a procedure equivalent to first differencing the series and than applying the LP-regression.
Phillips et al. (1999) introduced an exact log periodogram regression (ELP) procedure building upon the exact form of the DFT. As this procedure resembles the ELW estimator and both estimators rely on nonlinear optimisation, the ELP procedure loses the advantage of linear regression, but is still asymptotically less efficient according to Kim and Phillips (2006).

A.2 ELW Estimator for Unknown Mean

The model the ELW is based on assumes the initial value μ_0 is known. If this is not the case and μ_0 is replaced by the sample average as in Shimotsu et al. (2005); Shimotsu (2010a), the ELW is only consistent for $d \in (-\frac{1}{2}, 1)$ and asymptotically normal for $d \in (-\frac{1}{2}, \frac{3}{4})$. If μ_0 is replaced by X_1, then the estimator is consistent for $d \geq \frac{1}{2}$ and asymptotically normal for $d \in [\frac{1}{2}, 2)$. For this reason, according to Shimotsu (2010b) μ_0 can be estimated as in $\hat{\mu}(d) = w(d)\bar{X} + (1 - w(d))X_1$ with $w(d)$ being a smooth weight function, such that $w(d) = 1 \quad \text{for } d \leq \frac{1}{2} \quad \text{and } w(d) \in [0,1] \quad \text{for } \frac{1}{2} \leq d \leq \frac{3}{4} \quad \text{and } w(d) = 0 \quad \text{for } d \geq \frac{3}{4}$ and X_t corrected by $\hat{\mu}(d)$ in the periodogram ordinates in the objective function. Shimotsu (2010a) shows the consistency and asymptotic normality of this estimator for $d \in (-\frac{1}{2}, 2)$ excluding arbitrary small intervals around 0 and 1. The last suggested possibility is to directly replace X_t by $X_t - \mu$ in the periodogram ordinates and minimise the objective function with respect to (d, G, μ).
The fully extended local Whittle estimator by Abadir et al. (2007) is based on a fully extended discrete Fourier transform including correction terms and can be regarded as complementary to the two-step ELW of Shimotsu (2010b). Nonetheless, they show the consistency and asymptotic $N(0, \frac{1}{4})$ distribution of their estimator for $d \in (-\frac{3}{2}, \infty)$. Further, the mean has not to be estimated and the estimator is robust to presence of polynomial trends of an order as high as $\lfloor d + 1/2 \rfloor$.

On the other hand, memory parameters d of the form $\frac{1+2i}{2}$ with $i \in \mathbb{N}_0$ are not covered by this estimator (for more see Faÿ et al. (2009) p. 17).

A.3 Further Simulation Results

For the simulation set up of 1000 replications of an $ARFIMA(0,d,0)$ with $n = 2048$ and $m = 208 \approx n^{0.7}$ the box plots of each estimator were plotted for each analysed d. The red dots represent the true value of the memory parameter. While in Table 1 no uniform behaviour of the standard deviation could be observed, it gets evident that the distribution of the estimates for $d \in \{0.35, 0.55, 0.75\}$ is independent of d and symmetric. However, for $d \in \{1.75, 2.75, 3.75\}$ many great extreme values mark the distribution of the GPH and LW. The higher the order of the taper gets, the greater does the resemblance of the distribution of the estimator becomes to that of the lower values of d. Obviously, the ELW fits the true memory parameter the best, although requiring the greatest computation time:

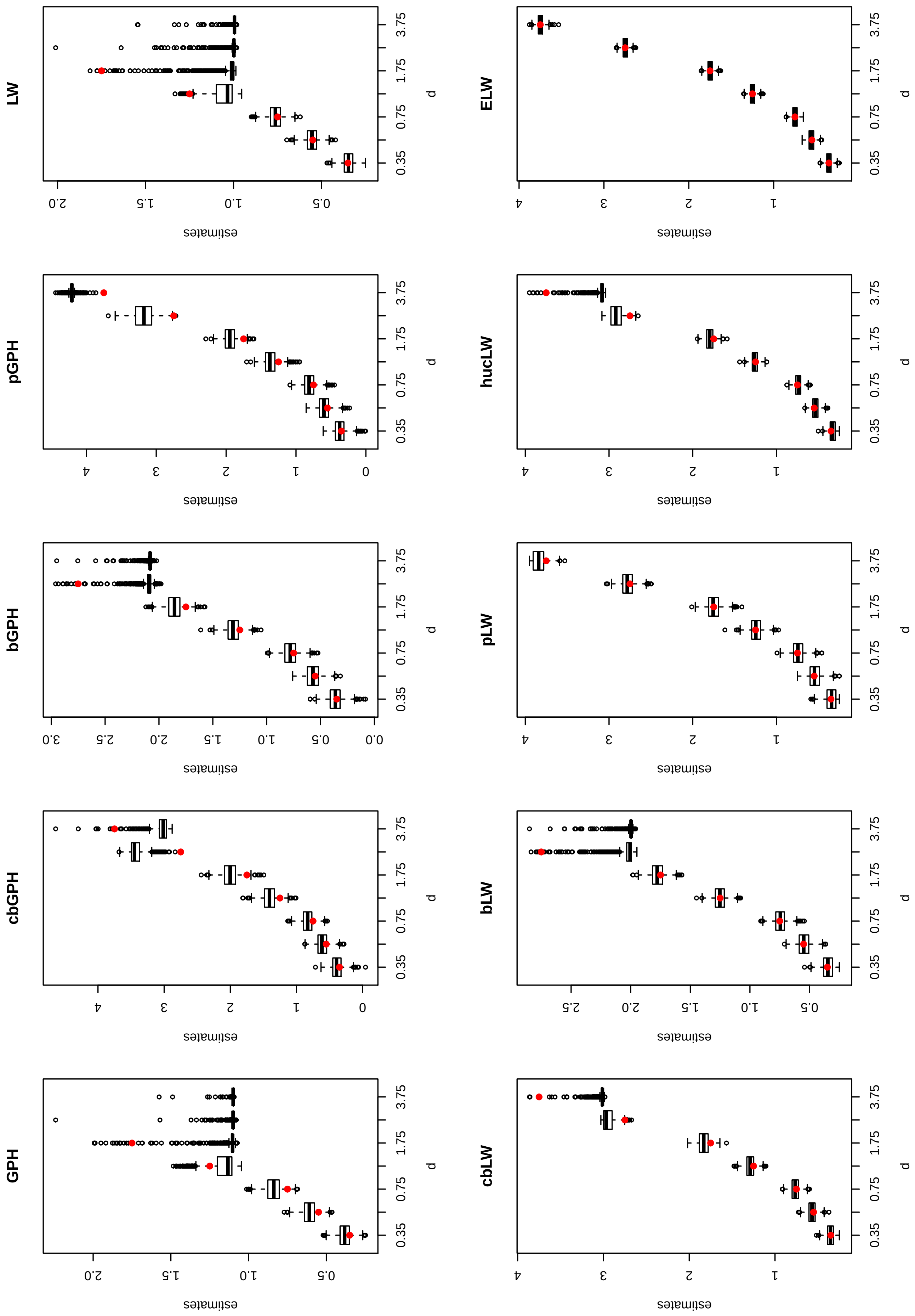

Figure 3: Boxplots of the memory estimates of $ARFIMA(0,d,0)$ for $n = 2048$ and $m = 208 \approx n^{0.7}$

	GPH	cbGPH	bGPH	pGPH	LW	cbLW	bLW	pLW	hucLW	ELW
e=0.5, d=0.35	0.04(0.15)	-0.04(0.29)	0.04(0.19)	0.07(0.25)	0.00(0.08)	0.03(0.10)	0.06(0.12)	0.09(0.18)	-0.04(0.08)	-0.01(0.09)
e=0.5, d=0.55	0.08(0.14)	-0.22(0.28)	0.06(0.19)	0.13(0.24)	0.01(0.09)	0.05(0.12)	0.04(0.14)	0.06(0.21)	-0.05(0.10)	0.00(0.09)
e=0.5, d=0.75	0.12(0.14)	-0.41(0.29)	0.09(0.19)	0.18(0.24)	0.02(0.09)	0.05(0.12)	0.05(0.15)	0.05(0.24)	-0.03(0.11)	-0.01(0.09)
e=0.5, d=1.25	-0.04(0.13)	-0.86(0.30)	0.18(0.19)	0.38(0.25)	-0.17(0.09)	0.11(0.13)	0.07(0.14)	0.06(0.22)	0.02(0.11)	-0.01(0.09)
e=0.5, d=1.75	-0.57(0.15)	-1.26(0.30)	0.30(0.20)	0.63(0.27)	-0.67(0.15)	0.21(0.15)	0.13(0.15)	0.10(0.23)	0.12(0.12)	-0.01(0.09)
e=0.5, d=2.75	-1.61(0.07)	-2.27(0.26)	-0.44(0.19)	1.18(0.34)	-1.72(0.10)	0.27(0.08)	-0.60(0.18)	0.17(0.24)	0.29(0.12)	-0.01(0.09)
e=0.5, d=3.75	-2.62(0.05)	-5.53(0.43)	-1.49(0.11)	0.81(0.13)	-2.74(0.06)	-0.69(0.12)	-1.65(0.11)	0.16(0.09)	-0.54(0.14)	-0.01(0.10)
e=0.6, d=0.35	0.03(0.08)	0.04(0.15)	0.03(0.11)	0.03(0.15)	-0.00(0.05)	0.01(0.07)	0.01(0.08)	0.02(0.10)	-0.03(0.06)	-0.00(0.06)
e=0.6, d=0.55	0.06(0.08)	0.04(0.15)	0.04(0.12)	0.08(0.15)	0.01(0.06)	0.03(0.08)	0.01(0.09)	0.01(0.13)	-0.03(0.07)	0.01(0.06)
e=0.6, d=0.75	0.08(0.08)	0.04(0.16)	0.06(0.11)	0.10(0.15)	0.02(0.06)	0.03(0.08)	0.01(0.09)	0.01(0.13)	-0.02(0.07)	-0.00(0.06)
e=0.6, d=1.25	-0.10(0.09)	0.03(0.16)	0.11(0.12)	0.22(0.16)	-0.18(0.08)	0.07(0.09)	0.03(0.09)	0.02(0.13)	0.02(0.07)	-0.00(0.06)
e=0.6, d=1.75	-0.63(0.13)	0.05(0.17)	0.18(0.12)	0.35(0.16)	-0.69(0.13)	0.14(0.10)	0.06(0.09)	0.03(0.14)	0.08(0.08)	-0.00(0.06)
e=0.6, d=2.75	-1.66(0.05)	0.22(0.13)	-0.56(0.14)	0.73(0.25)	-1.73(0.08)	0.23(0.07)	-0.66(0.14)	0.08(0.14)	0.22(0.09)	-0.00(0.06)
e=0.6, d=3.75	-2.67(0.03)	-1.10(0.12)	-1.59(0.08)	0.60(0.07)	-2.75(0.05)	-0.71(0.09)	-1.71(0.08)	0.13(0.09)	-0.60(0.12)	-0.00(0.06)
e=0.7, d=0.35	0.03(0.05)	0.04(0.10)	0.01(0.07)	0.02(0.09)	-0.00(0.04)	0.00(0.05)	-0.00(0.05)	-0.00(0.07)	-0.02(0.04)	0.00(0.04)
e=0.7, d=0.55	0.06(0.05)	0.06(0.10)	0.02(0.07)	0.05(0.10)	0.00(0.04)	0.02(0.05)	-0.00(0.06)	-0.01(0.08)	-0.01(0.05)	0.00(0.04)
e=0.7, d=0.75	0.09(0.09)	0.08(0.10)	0.03(0.07)	0.06(0.10)	0.01(0.04)	0.01(0.05)	-0.00(0.06)	-0.01(0.08)	-0.01(0.05)	-0.00(0.04)
e=0.7, d=1.25	-0.08(0.09)	0.16(0.11)	0.06(0.07)	0.12(0.10)	-0.19(0.08)	0.04(0.06)	0.00(0.06)	-0.01(0.08)	0.01(0.04)	-0.00(0.08)
e=0.7, d=1.75	-0.62(0.11)	0.25(0.13)	0.11(0.08)	0.19(0.10)	-0.71(0.11)	0.08(0.07)	0.03(0.06)	0.00(0.09)	0.05(0.05)	-0.00(0.12)
e=0.7, d=2.75	-1.64(0.04)	0.67(0.13)	-0.64(0.10)	0.43(0.16)	-1.74(0.06)	0.19(0.07)	-0.71(0.11)	0.03(0.08)	0.16(0.08)	-0.00(0.08)
e=0.7, d=3.75	-2.65(0.02)	-0.71(0.14)	-1.66(0.05)	0.46(0.04)	-2.75(0.03)	-0.72(0.07)	-1.74(0.06)	0.09(0.08)	-0.64(0.10)	-0.00(0.09)
e=0.8, d=0.35	0.00(0.03)	0.01(0.06)	0.00(0.05)	0.01(0.06)	-0.01(0.03)	-0.00(0.03)	-0.01(0.04)	-0.01(0.05)	-0.00(0.03)	0.01(0.02)
e=0.8, d=0.55	0.00(0.03)	0.01(0.06)	0.00(0.05)	0.02(0.07)	-0.01(0.03)	0.00(0.04)	-0.01(0.04)	-0.01(0.05)	-0.00(0.03)	0.02(0.02)
e=0.8, d=0.75	0.02(0.04)	0.01(0.06)	0.01(0.05)	0.03(0.06)	0.00(0.03)	-0.00(0.04)	-0.01(0.04)	-0.01(0.05)	-0.00(0.03)	0.01(0.02)
e=0.8, d=1.25	-0.21(0.07)	0.00(0.06)	0.02(0.05)	0.05(0.07)	-0.21(0.07)	0.01(0.04)	-0.02(0.04)	-0.02(0.05)	0.00(0.03)	0.01(0.02)
e=0.8, d=1.75	-0.74(0.08)	-0.00(0.06)	0.04(0.05)	0.08(0.07)	-0.74(0.09)	0.03(0.05)	-0.01(0.04)	-0.03(0.05)	0.02(0.04)	0.01(0.02)
e=0.8, d=2.75	-1.75(0.03)	0.20(0.11)	-0.71(0.08)	0.20(0.10)	-1.76(0.05)	0.15(0.08)	-0.75(0.10)	-0.03(0.06)	0.11(0.07)	0.01(0.02)
e=0.8, d=3.75	-2.75(0.01)	-0.75(0.03)	-1.73(0.03)	0.36(0.05)	-2.77(0.03)	-0.73(0.05)	-1.77(0.05)	0.01(0.06)	-0.66(0.08)	0.01(0.02)
e=0.868, d=0.35	-0.01(0.02)	-0.00(0.04)	-0.01(0.04)	-0.01(0.05)	-0.02(0.02)	-0.02(0.03)	-0.02(0.03)	-0.02(0.04)	0.03(0.02)	0.05(0.02)
e=0.868, d=0.55	-0.02(0.02)	-0.01(0.04)	-0.02(0.04)	-0.01(0.05)	-0.02(0.02)	-0.02(0.03)	-0.03(0.03)	-0.03(0.04)	0.02(0.02)	0.05(0.02)
e=0.868, d=0.75	-0.01(0.03)	-0.01(0.04)	-0.02(0.04)	-0.01(0.05)	-0.03(0.02)	-0.03(0.03)	-0.04(0.03)	-0.04(0.04)	0.01(0.02)	0.05(0.02)
e=0.868, d=1.25	-0.24(0.07)	-0.02(0.04)	-0.04(0.04)	-0.02(0.05)	-0.25(0.07)	-0.04(0.03)	-0.06(0.03)	-0.07(0.04)	-0.01(0.02)	0.05(0.02)
e=0.868, d=1.75	-0.77(0.07)	-0.03(0.04)	-0.04(0.04)	-0.01(0.05)	-0.78(0.08)	-0.03(0.05)	-0.07(0.03)	-0.09(0.04)	-0.01(0.03)	0.05(0.02)
e=0.868, d=2.75	-1.78(0.03)	0.18(0.14)	-0.81(0.07)	0.03(0.07)	-1.79(0.05)	0.09(0.11)	-0.82(0.09)	-0.13(0.04)	0.04(0.08)	0.05(0.02)
e=0.868, d=3.75	-2.78(0.01)	-0.68(0.03)	-1.82(0.02)	0.29(0.08)	-2.80(0.02)	-0.71(0.04)	-1.84(0.04)	-0.12(0.06)	-0.63(0.07)	0.05(0.02)

Table 4: Bias and corresponding standard deviation in brackets of estimators with bandwidth $m = n^e$ and $n = 2048$

	GPH	cbGPH	bGPH	pGPH	LW	cbLW	bLW	pLW	hucLW	ELW
e=0.5, d=0.35	0.04(0.21)	0.06(0.41)	0.05(0.25)	0.08(0.32)	0.01(0.09)	0.04(0.12)	0.02(0.13)	0.06(0.19)	-0.03(0.09)	-0.01(0.11)
e=0.5, d=0.55	0.06(0.20)	0.05(0.43)	0.09(0.26)	0.17(0.32)	0.00(0.11)	0.07(0.15)	-0.01(0.17)	0.01(0.26)	-0.06(0.12)	0.01(0.11)
e=0.5, d=0.75	0.11(0.20)	0.02(0.40)	0.13(0.26)	0.22(0.32)	0.02(0.11)	0.07(0.15)	0.00(0.18)	0.01(0.29)	-0.04(0.13)	-0.01(0.10)
e=0.5, d=1.25	-0.08(0.17)	-0.04(0.43)	0.23(0.25)	0.49(0.31)	-0.17(0.10)	0.15(0.16)	0.03(0.18)	0.00(0.29)	0.03(0.13)	-0.02(0.11)
e=0.5, d=1.75	-0.61(0.15)	-0.10(0.41)	0.37(0.25)	0.77(0.34)	-0.66(0.15)	0.25(0.18)	0.09(0.18)	0.04(0.30)	0.14(0.15)	-0.02(0.11)
e=0.5, d=2.75	-1.64(0.07)	-0.17(0.27)	-0.36(0.24)	1.41(0.37)	-1.72(0.10)	0.29(0.08)	-0.63(0.21)	0.14(0.31)	0.34(0.13)	-0.02(0.11)
e=0.5, d=3.75	-2.64(0.08)	-1.69(0.20)	-1.42(0.16)	0.93(0.15)	-2.74(0.09)	-0.67(0.15)	-1.69(0.14)	0.13(0.14)	-0.50(0.16)	-0.02(0.11)
e=0.6, d=0.35	0.03(0.12)	0.06(0.22)	0.03(0.15)	0.05(0.20)	0.00(0.06)	0.02(0.08)	0.01(0.09)	0.02(0.13)	-0.03(0.07)	-0.01(0.07)
e=0.6, d=0.55	0.04(0.12)	0.05(0.22)	0.06(0.15)	0.11(0.20)	0.00(0.07)	0.04(0.10)	-0.01(0.11)	-0.01(0.17)	-0.04(0.09)	0.01(0.07)
e=0.6, d=0.75	0.08(0.12)	0.04(0.22)	0.08(0.15)	0.13(0.20)	0.02(0.07)	0.04(0.10)	-0.00(0.11)	-0.00(0.17)	-0.02(0.09)	-0.00(0.07)
e=0.6, d=1.25	-0.13(0.11)	-0.00(0.22)	0.14(0.15)	0.30(0.20)	-0.17(0.09)	0.09(0.11)	0.02(0.11)	0.00(0.17)	0.02(0.09)	-0.01(0.07)
e=0.6, d=1.75	-0.65(0.12)	-0.02(0.22)	0.24(0.16)	0.47(0.21)	-0.68(0.13)	0.17(0.12)	0.06(0.12)	0.02(0.18)	0.09(0.10)	-0.01(0.07)
e=0.6, d=2.75	-1.68(0.04)	0.00(0.15)	-0.50(0.17)	0.94(0.30)	-1.73(0.08)	0.24(0.07)	-0.66(0.16)	0.09(0.19)	0.26(0.11)	-0.01(0.07)
e=0.6, d=3.75	-2.68(0.06)	-1.36(0.16)	-1.54(0.12)	0.70(0.09)	-2.74(0.07)	-0.69(0.12)	-1.71(0.12)	0.13(0.11)	-0.56(0.14)	-0.01(0.07)
e=0.7, d=0.35	0.04(0.07)	0.03(0.14)	0.02(0.10)	0.03(0.13)	-0.00(0.05)	0.01(0.06)	-0.00(0.07)	0.00(0.09)	-0.02(0.05)	0.00(0.05)
e=0.7, d=0.55	0.08(0.07)	0.03(0.10)	0.03(0.13)	0.06(0.14)	0.00(0.05)	0.02(0.07)	-0.01(0.07)	-0.01(0.11)	-0.02(0.06)	0.01(0.05)
e=0.7, d=0.75	0.12(0.07)	0.04(0.15)	0.05(0.10)	0.08(0.13)	0.01(0.05)	0.02(0.07)	-0.01(0.07)	-0.01(0.11)	-0.02(0.06)	0.00(0.05)
e=0.7, d=1.25	-0.04(0.09)	0.08(0.16)	0.08(0.10)	0.17(0.13)	-0.18(0.08)	0.05(0.07)	0.00(0.07)	-0.01(0.11)	0.01(0.06)	0.00(0.05)
e=0.7, d=1.75	-0.58(0.12)	0.16(0.18)	0.15(0.10)	0.27(0.14)	-0.71(0.11)	0.10(0.09)	0.03(0.08)	0.00(0.11)	0.06(0.07)	-0.63(0.05)
e=0.7, d=2.75	-1.61(0.04)	0.45(0.15)	-0.60(0.12)	0.59(0.21)	-1.74(0.06)	0.21(0.07)	-0.70(0.13)	0.03(0.12)	0.20(0.09)	-0.00(0.05)
e=0.7, d=3.75	-2.61(0.04)	-1.60(0.33)	-1.62(0.10)	0.54(0.06)	-2.75(0.06)	-0.71(0.10)	-1.73(0.10)	0.10(0.10)	-0.61(0.12)	-0.00(0.05)
e=0.8, d=0.35	0.05(0.04)	0.04(0.10)	0.00(0.06)	0.01(0.09)	-0.01(0.03)	-0.00(0.05)	-0.01(0.05)	-0.01(0.07)	-0.00(0.04)	0.02(0.04)
e=0.8, d=0.55	0.08(0.05)	0.10(0.11)	0.01(0.07)	0.03(0.09)	-0.01(0.03)	0.00(0.05)	-0.02(0.05)	-0.02(0.07)	-0.00(0.04)	0.03(0.04)
e=0.8, d=0.75	0.13(0.05)	0.18(0.11)	0.01(0.06)	0.03(0.09)	-0.00(0.04)	-0.00(0.05)	-0.02(0.05)	-0.02(0.07)	-0.00(0.04)	0.02(0.04)
e=0.8, d=1.25	-0.02(0.09)	0.42(0.14)	0.02(0.06)	0.07(0.09)	-0.21(0.07)	0.01(0.05)	-0.03(0.05)	-0.03(0.07)	0.00(0.04)	0.02(0.04)
e=0.8, d=1.75	-0.56(0.11)	0.73(0.17)	0.06(0.07)	0.13(0.09)	-0.74(0.10)	0.04(0.07)	-0.01(0.05)	-0.04(0.07)	0.02(0.05)	0.02(0.04)
e=0.8, d=2.75	-1.59(0.03)	1.67(0.19)	-0.70(0.10)	0.31(0.14)	-1.76(0.04)	0.16(0.09)	-0.75(0.11)	-0.04(0.08)	0.13(0.09)	0.02(0.04)
e=0.8, d=3.75	-2.59(0.03)	0.03(0.27)	-1.71(0.09)	0.42(0.05)	-2.77(0.04)	-0.72(0.08)	-1.78(0.09)	0.00(0.09)	-0.63(0.10)	0.02(0.04)
e=0.868, d=0.35	0.01(0.03)	0.01(0.07)	-0.01(0.05)	-0.01(0.07)	-0.02(0.03)	-0.02(0.04)	-0.02(0.04)	-0.02(0.05)	0.03(0.03)	0.06(0.03)
e=0.868, d=0.55	0.03(0.03)	0.05(0.07)	-0.02(0.05)	-0.01(0.07)	-0.03(0.03)	-0.02(0.04)	-0.04(0.04)	-0.04(0.05)	0.02(0.03)	0.06(0.03)
e=0.868, d=0.75	0.05(0.04)	0.10(0.07)	-0.03(0.05)	-0.01(0.07)	-0.03(0.03)	-0.04(0.04)	-0.05(0.04)	-0.05(0.05)	0.01(0.03)	0.06(0.03)
e=0.868, d=1.25	-0.13(0.08)	0.25(0.09)	-0.04(0.05)	-0.01(0.07)	-0.25(0.07)	-0.04(0.05)	-0.08(0.04)	-0.08(0.06)	-0.01(0.03)	0.06(0.03)
e=0.868, d=1.75	-0.67(0.09)	0.45(0.11)	-0.04(0.05)	0.01(0.07)	-0.78(0.09)	-0.03(0.06)	-0.09(0.04)	-0.11(0.06)	-0.01(0.04)	0.06(0.03)
e=0.868, d=2.75	-1.69(0.02)	1.18(0.19)	-0.81(0.08)	0.09(0.10)	-1.80(0.04)	0.09(0.12)	-0.83(0.10)	-0.15(0.07)	0.06(0.09)	0.06(0.03)
e=0.868, d=3.75	-2.69(0.02)	-0.20(0.22)	-1.82(0.08)	0.33(0.08)	-2.81(0.04)	-0.68(0.06)	-1.86(0.09)	-0.15(0.08)	-0.58(0.08)	0.06(0.03)

Table 5: Bias and corresponding standard deviation in brackets of estimators with bandwidth $m = n^e$ and $n = 1024$

	GPH	cbGPH	bGPH	pGPH	LW	cbLW	bLW	pLW	hucLW	ELW
e=0.5, d=0.35	0.07(0.26)	0.05(0.61)	0.06(0.31)	0.10(0.38)	0.01(0.10)	0.05(0.14)	0.11(0.19)	0.61(0.54)	-0.03(0.09)	-0.01(0.14)
e=0.5, d=0.55	0.12(0.27)	-0.25(0.61)	0.08(0.31)	0.19(0.38)	0.00(0.13)	0.08(0.18)	0.09(0.22)	0.61(0.55)	-0.07(0.15)	-0.00(0.13)
e=0.5, d=0.75	0.18(0.27)	-0.52(0.65)	0.14(0.34)	0.25(0.41)	0.01(0.14)	0.07(0.19)	0.09(0.25)	0.63(0.55)	-0.06(0.17)	-0.00(0.12)
e=0.5, d=1.25	0.03(0.21)	-1.24(0.65)	0.27(0.31)	0.59(0.38)	-0.16(0.12)	0.18(0.19)	0.13(0.24)	0.67(0.56)	0.04(0.16)	-0.02(0.14)
e=0.5, d=1.75	-0.49(0.20)	-1.93(0.67)	0.43(0.33)	0.95(0.42)	-0.65(0.17)	0.31(0.22)	0.20(0.25)	0.68(0.56)	0.18(0.18)	-0.02(0.14)
e=0.5, d=2.75	-1.55(0.09)	-3.98(0.68)	-0.29(0.28)	1.63(0.40)	-1.72(0.11)	0.32(0.10)	-0.52(0.23)	0.73(0.40)	0.38(0.15)	-0.02(0.14)
e=0.5, d=3.75	-2.55(0.06)	-8.93(0.90)	-1.36(0.16)	1.05(0.18)	-2.74(0.09)	-0.66(0.17)	-1.59(0.14)	0.19(0.08)	-0.47(0.15)	-0.02(0.14)
e=0.6, d=0.35	0.07(0.15)	0.05(0.34)	0.05(0.20)	0.08(0.26)	0.00(0.08)	0.03(0.10)	0.01(0.11)	0.10(0.20)	-0.03(0.08)	-0.00(0.09)
e=0.6, d=0.55	0.11(0.16)	-0.06(0.36)	0.07(0.20)	0.14(0.26)	0.01(0.10)	0.05(0.13)	-0.01(0.15)	0.07(0.25)	-0.05(0.12)	0.01(0.09)
e=0.6, d=0.75	0.16(0.16)	-0.18(0.36)	0.10(0.20)	0.18(0.27)	0.01(0.10)	0.04(0.13)	-0.01(0.15)	0.06(0.25)	-0.04(0.11)	-0.00(0.08)
e=0.6, d=1.25	0.02(0.13)	-0.44(0.38)	0.18(0.21)	0.40(0.27)	-0.17(0.09)	0.11(0.14)	0.01(0.15)	0.08(0.26)	0.02(0.11)	-0.01(0.09)
e=0.6, d=1.75	-0.52(0.16)	-0.61(0.41)	0.29(0.21)	0.65(0.14)	-0.68(0.16)	0.21(0.16)	0.06(0.16)	0.10(0.26)	0.12(0.13)	-0.01(0.09)
e=0.6, d=2.75	-1.56(0.07)	-1.24(0.37)	-0.43(0.22)	1.24(0.36)	-1.73(0.08)	0.27(0.08)	-0.65(0.19)	0.19(0.25)	0.30(0.12)	-0.01(0.09)
e=0.6, d=3.75	-2.56(0.05)	-4.68(0.62)	-1.48(0.12)	0.85(0.13)	-2.74(0.07)	-0.68(0.13)	-1.71(0.12)	0.15(0.12)	-0.53(0.12)	-0.01(0.09)
e=0.7, d=0.35	0.03(0.09)	0.07(0.18)	0.03(0.13)	0.05(0.17)	-0.00(0.06)	0.01(0.08)	0.00(0.08)	0.03(0.12)	-0.03(0.06)	0.06(0.06)
e=0.7, d=0.55	0.05(0.10)	0.09(0.18)	0.04(0.13)	0.08(0.17)	0.00(0.07)	0.03(0.09)	-0.00(0.10)	0.01(0.15)	-0.03(0.08)	0.02(0.07)
e=0.7, d=0.75	0.09(0.10)	0.10(0.18)	0.06(0.13)	0.11(0.18)	0.01(0.07)	0.02(0.09)	0.00(0.10)	0.02(0.15)	-0.02(0.08)	0.01(0.06)
e=0.7, d=1.25	-0.10(0.10)	0.12(0.19)	0.11(0.13)	0.23(0.17)	-0.18(0.08)	0.07(0.10)	0.01(0.10)	0.01(0.15)	0.01(0.08)	0.00(0.07)
e=0.7, d=1.75	-0.64(0.12)	0.18(0.19)	0.18(0.14)	0.38(0.19)	-0.70(0.12)	0.14(0.12)	0.04(0.11)	0.02(0.16)	0.08(0.09)	0.00(0.07)
e=0.7, d=2.75	-1.67(0.05)	0.45(0.16)	-0.55(0.16)	0.80(0.28)	-1.74(0.07)	0.23(0.08)	-0.68(0.16)	0.07(0.16)	0.23(0.10)	0.00(0.07)
e=0.7, d=3.75	-2.67(0.04)	-0.83(0.15)	-1.59(0.09)	0.64(0.07)	-2.75(0.05)	-0.70(0.11)	-1.73(0.09)	0.12(0.11)	-0.58(0.10)	0.00(0.07)
e=0.8, d=0.35	0.01(0.06)	0.07(0.11)	0.01(0.09)	0.02(0.12)	-0.01(0.04)	-0.00(0.06)	-0.01(0.06)	-0.01(0.08)	-0.00(0.05)	0.03(0.04)
e=0.8, d=0.55	0.02(0.06)	0.07(0.11)	0.01(0.09)	0.04(0.11)	-0.01(0.05)	0.00(0.06)	-0.02(0.07)	-0.03(0.10)	-0.01(0.06)	0.03(0.04)
e=0.8, d=0.75	0.04(0.06)	0.08(0.11)	0.02(0.09)	0.05(0.12)	-0.01(0.05)	-0.00(0.06)	-0.02(0.07)	-0.03(0.10)	-0.00(0.06)	0.03(0.04)
e=0.8, d=1.25	-0.17(0.08)	0.08(0.11)	0.03(0.09)	0.11(0.12)	-0.21(0.07)	0.02(0.07)	-0.03(0.07)	-0.05(0.10)	0.00(0.06)	0.03(0.04)
e=0.8, d=1.75	-0.70(0.10)	0.10(0.12)	0.07(0.10)	0.19(0.13)	-0.74(0.11)	0.06(0.09)	-0.02(0.08)	-0.06(0.10)	0.04(0.07)	0.03(0.04)
e=0.8, d=2.75	-1.73(0.03)	0.36(0.13)	-0.68(0.12)	0.45(0.20)	-1.77(0.05)	0.17(0.09)	-0.76(0.13)	-0.05(0.11)	0.16(0.10)	0.03(0.04)
e=0.8, d=3.75	-2.73(0.03)	-0.89(0.13)	-1.70(0.07)	0.50(0.05)	-2.78(0.04)	-0.70(0.08)	-1.79(0.08)	0.01(0.12)	-0.59(0.08)	0.08(0.04)
e=0.868, d=0.35	-0.01(0.05)	0.03(0.08)	-0.01(0.07)	-0.00(0.09)	-0.03(0.04)	-0.02(0.05)	-0.03(0.05)	-0.03(0.06)	0.04(0.04)	0.09(0.04)
e=0.868, d=0.55	-0.02(0.05)	0.02(0.08)	-0.03(0.07)	0.00(0.09)	-0.04(0.04)	-0.03(0.05)	-0.05(0.05)	-0.06(0.08)	0.02(0.04)	0.08(0.04)
e=0.868, d=0.75	-0.02(0.05)	0.02(0.09)	-0.03(0.07)	-0.00(0.10)	-0.04(0.04)	-0.04(0.05)	-0.06(0.06)	-0.07(0.08)	0.01(0.04)	0.08(0.04)
e=0.868, d=1.25	-0.24(0.07)	0.00(0.09)	-0.05(0.07)	0.02(0.10)	-0.26(0.07)	-0.05(0.06)	-0.09(0.06)	-0.11(0.08)	-0.01(0.05)	-0.06(0.04)
e=0.868, d=1.75	-0.78(0.09)	-0.00(0.09)	-0.04(0.08)	0.05(0.11)	-0.79(0.10)	-0.03(0.08)	-0.10(0.06)	-0.15(0.08)	-0.00(0.05)	0.08(0.04)
e=0.868, d=2.75	-1.79(0.02)	0.31(0.19)	-0.81(0.10)	0.19(0.16)	-1.82(0.05)	0.10(0.13)	-0.86(0.12)	-0.20(0.09)	0.08(0.11)	0.08(0.04)
e=0.868, d=3.75	-2.80(0.02)	-0.62(0.09)	-1.83(0.05)	0.38(0.08)	-2.82(0.04)	-0.65(0.07)	-1.89(0.07)	-0.19(0.11)	-0.52(0.06)	0.07(0.04)

Table 6: Bias and corresponding standard deviation in brackets of estimators with bandwidth $m = n^e$ and $n = 512$

References

Abadir, K. M., Distaso, W., and Giraitis, L. (2007). Nonstationarity-extended local whittle estimation. *Journal of Econometrics*, 141(2):1353–1384.

Arteche, J. and Velasco, C. (2005). Trimming and tapering semi-parametric estimates in asymmetric long memory time series. *Journal of Time Series Analysis*, 26(4):581–611.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1):5–59.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2016). *Long-Memory Processes.* Springer.

Bloomfield, P. (2004). *Fourier analysis of time series: an introduction.* John Wiley & Sons.

Boubaker, H., Boutahar, M., and Khalfaoui, R. (2017). Wavelets and estimation of long memory in nonstationary models: Does anything beat the exact local whittle estimator? *Communications in Statistics-Simulation and Computation*, 46(2):1189–1218.

Boutahar, M. and Khalfaoui, R. (2011). Estimation of the long memory parameter in non stationary models: A simulation study.

Cheung, Y.-W. and Lai, K. S. (1993). A fractional cointegration analysis of purchasing power parity. *Journal of Business & Economic Statistics*, 11(1):103–112.

Dahlhaus, R. (1983). Spectral analysis with tapered data. *Journal of Time Series Analysis*, 4(3):163–175.

Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *The Annals of Statistics*, pages 808–841.

Dahlhaus, R. et al. (1997). Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37.

Faÿ, G., Moulines, E., Roueff, F., and Taqqu, M. S. (2009). Estimators of long-memory: Fourier versus wavelets. *Journal of econometrics*, 151(2):159–177.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4):221–238.

Hurvich, C. M. and Chen, W. W. (2000). An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis*, 21(2):155–180.

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. *Journal of Time Series Analysis*, 19(1):19–46.

Hurvich, C. M. and Ray, B. K. (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of time series analysis*, 16(1):17–41.

Kim, C. S. and Phillips, P. C. (2006). Log periodogram regression: the nonstationary case.

Nouira, L., Boutahar, M., and Marimoutou, V. (2009). The effect of tapering on the semiparametric estimators for nonstationary long memory processes. *Statistical Papers*, 50(2):225.

Percival, D. B. and Walden, A. T. (1993). *Spectral analysis for physical applications*. cambridge university press.

Phillips, P. C. et al. (1999). *Unit root log periodogram regression*. Cowles Foundation for Research in Economics at Yale University.

Robinson, P. (1995a). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*.

Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, pages 1630–1661.

Shimotsu, K. (2010a). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(2):501–540.

Shimotsu, K. (2010b). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(2):501–540.

Shimotsu, K., Phillips, P. C., et al. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933.

Sibbertsen, P. (2004). Long memory in volatilities of german stock returns. *Empirical Economics*, 29(3):477–488.

Taqqu, M. S., Teverovsky, V., and Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals*, 3(04):785–798.

Tschernig, R. (2013). *Wechselkurse, Unsicherheit und long memory*. Springer-Verlag.

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. *Spectral analysis of time series*, 25(6):2546.

Velasco, C. (1999a). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127.

Velasco, C. (1999b). Non-stationary log-periodogram regression. *Journal of Econometrics*, 91(2):325–371.

Velasco, C. (2000). Non-gaussian log-periodogram regression. *Econometric Theory*, 16(1):44–79.

Velasco, C. and Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, 95(452):1229–1243.

Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för matematik*, 2(5):423–434.

Zhurbenko, I. and Trush, N. (1979). Estimation of spectral densities of stationary processes. *Lithuanian Mathematical Journal*, 19(1):47–60.