

Estimation of long memory parameter in nonparametric regression

Yeouyoung Cho^a, Changryong Baek^{1,b}

^aSchool of Management Engineering, KAIST, Korea;

^bDepartment of Statistics, Sungkyunkwan University, Korea

Abstract

This paper considers the estimation of the long memory parameter in nonparametric regression with strongly correlated errors. The key idea is to minimize a unified mean squared error of long memory parameter to select both kernel bandwidth and the number of frequencies used in exact local Whittle estimation. A unified mean squared error framework is more natural because it provides both goodness of fit and measure of strong dependence. The block bootstrap is applied to evaluate the mean squared error. Finite sample performance using Monte Carlo simulations shows the closest performance to the oracle. The proposed method outperforms existing methods especially when dependency and sample size increase. The proposed method is also illustrated to the volatility of exchange rate between Korean Won for US dollar.

Keywords: nonparametric regression, kernel bandwidth selection, long memory, exact local Whittle estimator, block bootstrapping

1. Introduction

Nonparametric regression is a statistical method estimating the mean function when the true signal is masked by some level of noise. Many methods such as kernel smoothing, splines, Fourier, and wavelet methods have been developed at an exponential rate over the last two decades. However, most of these methods are focused on cases when the true trend function is contaminated by uncorrelated or weakly correlated errors. In this paper, nonparametric regression is considered when the errors are strongly correlated in the sense of long memory. For related works, see Beran and Feng (2002), Hall and Hart (1990), Künsch (1986), Masry and Mielniczuk (1999), Opsomer *et al.* (2001) and references therein.

To concrete our discussion, consider the following statistical model

$$Y_t = f\left(\frac{t}{n}\right) + \epsilon_t, \quad t = 1, 2, \dots, n,$$

where unobservable error $\{\epsilon_t\}$ is assumed to be a long memory, also known as long-range dependence (LRD). Long memory errors $\{\epsilon_t\}$ are formally defined as the weakly stationary time series with spectral density diverging at zero-frequency,

$$g(\omega) \sim c|\omega|^{-2d}, \quad \text{as } \omega \rightarrow 0, \quad (1.1)$$

¹ Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail: crbaek@skku.edu

and $d \in (0, 1/2)$ is called the long memory or LRD parameter. A popular kernel estimator of mean function is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_i K\left(\frac{x - i/n}{h}\right) Y_i \quad (1.2)$$

with bandwidth h and a kernel function K that integrates to one. A crucial point in nonparametric regression is the selection of bandwidth parameter h . Hall and Hart (1990) showed under some mild assumptions that the optimal bandwidth minimizing the mean integrated squared error (MISE)

$$\text{MISE}(h) = \int_I E\left(\hat{f}_h(x) - f(x)\right)^2 dx, \quad I \subset (0, 1), \quad (1.3)$$

is asymptotically given by

$$h \sim Cn^{-\frac{1-2d}{5-2d}},$$

for some positive constant C . Thus, it is important to estimate the long memory parameter to find the optimal bandwidth.

The long memory parameter estimation is also necessary for data adaptive bandwidth selection methods, such as the cross validation or bootstrap methods. The usual leave-one-out cross validation suffers from severe bias for strongly correlated errors. See Opsomer *et al.* (2001) for an excellent review of this phenomenon. This bias can be reduced by considering the modified cross-validation (MCV) method of Chu and Marron (1991), also known as leave- k -out cross validation. The main idea is to delete k observations forward and backward in estimating $f(t/n)$ so that strong dependence is alleviated. More formally, it is given by minimizing the residual sum of squares given by

$$\hat{h} = \underset{h}{\operatorname{argmin}} n^{-1} \sum_{t=1}^n \left(\hat{f}_h^{(-k)}\left(\frac{t}{n}\right) - Y_t \right)^2, \quad (1.4)$$

where $\hat{f}_h^{(-k)}(t/n)$ is a kernel estimator of $f(t/n)$ with bandwidth h after leaving out Y_{t+j} , $-k \leq j \leq k$ in the estimation. A continuous analog of MCV is proposed by Kim *et al.* (2009) using a bimodal kernel. Still, the block size (or leave-out number) also depends on the LRD parameter that is critical to determine the optimal bandwidth. Hall *et al.* (1995) proposed a bandwidth selection method based on block bootstrap where the MISE is estimated from a block bootstrap sample of block length $(2k + 1)$. The long memory parameter plays an additional key role in estimating the true smooth trend function masked by strongly correlated errors.

This paper considers a more precise estimation of the long memory parameter in the presence of a smooth trend. Somewhat surprisingly, Robinson (1997) showed that the memory parameter can be estimated $\log^{1/2}(n)$ -consistently from the raw data $\{Y_t\}$, even in the presence of a trend. However, the iterative method is typically used, as in Ray and Tsay (1997), because the long memory parameter estimated by Robinson (1997) performs poorly in practice. That is, starting with initial bandwidth h_0 , estimate the long memory parameter from the residual series. Then, update the kernel bandwidth from the estimated long memory parameter. A more efficient method of estimating the long memory parameter was proposed by Hurvich *et al.* (2005), where the trend function estimation is bypassed by trimming, tapering and differencing.

However, an additional tuning parameter is also required in estimating the long memory parameter. For example, the number of frequencies used in the semiparametric estimation of long memory such

as exact local Whittle estimation (ELW) plays a central role. One of the pioneering methods suggested by Henry (2001) is to minimize the mean squared error (MSE)

$$\hat{m} = \operatorname{argmin}_m \mathbb{E} \left(\hat{d}(m) - d \right)^2, \tag{1.5}$$

where $\hat{d}(m)$ is the LW estimator based on m number of low frequencies.

This paper starts with the observation that kernel bandwidth selection can be unified by minimizing the MSE of the long memory parameter for a more precise estimation of the LRD parameter. The available methods iteratively find kernel bandwidth h by minimizing (empirical) MISE in (1.3), but the LRD parameter is estimated from the residuals by minimizing the MSE in (1.5). Therefore, this kernel bandwidth is potentially sub-optimal for the estimation of LRD parameter since it minimizes errors from the true trend function. Instead, it is proposed to find the tuning parameters simultaneously by minimizing a single loss function

$$\left(\hat{h}_{lw}, \hat{m}_{lw} \right) = \operatorname{argmin}_{h,m} \mathbb{E} \left(\hat{d}(m, h) - d \right)^2, \tag{1.6}$$

where $\hat{d}(m, h)$ is the ELW estimation of the LRD parameter from the residual series with kernel bandwidth h and ELW bandwidth m . In the newly suggested framework, the LRD parameter encapsulates both goodness of fit and the measure of strong dependence. This is also intuitive in the sense that residuals are good proxies for model checking. The MSE is estimated by block bootstrapping methods. A more detailed description of the proposed method is elaborated in Section 2. The finite sample performance is examined in Section 3 and illustrated with KRW-USD exchange rate in Section 4. Section 5 contains the conclusion.

2. Description of the proposed method

In this section, the tuning parameters selection method based on the MSE of the LRD parameter estimation is detailed. Illustrated here with the Nadaraya-Watson estimator and the ELW estimator for their simplicity and superior performance in practice. However, it can be applied to other types of Kernel estimators such as the local polynomial estimator and/or a variety of LRD parameter estimation methods such as the log-periodogram estimator also known as GPH estimator (Geweke and Porter-Hudak, 1983; Robinson, 1995).

For given observations Y_1, \dots, Y_n , consider a kernel estimator

$$\hat{f}_h(x) = \sum_i w_i(h) Y_i, \quad w_i(h) = K \left(\frac{x - i/n}{h} \right) \Bigg/ \sum_{j=1}^n K \left(\frac{x - j/n}{h} \right), \tag{2.1}$$

where K is a kernel function and h is a kernel bandwidth. Then, the residuals can be written as

$$e_t(h) = Y_t - \hat{f}_h(t), \quad t = 1, \dots, n. \tag{2.2}$$

The LRD parameter is now estimated based on the residual series $\{e_t, t = 1, \dots, n\}$. The ELW estimator is a semiparametric estimation of the long memory parameter proposed by Shimotsu and Phillips (2005) shown to be consistent and asymptotically normal outside the stationary region if the optimization is searched within the length of a $9/2$ interval. It is formally defined by

$$\hat{d}(m, h) = \operatorname{argmin}_{d \in \Theta} R_{m,h}(d),$$

where $\Theta = [\Delta_1, \Delta_2]$ for $-\infty < \Delta_1 < \Delta_2 < \infty$ with $\Delta_2 - \Delta_1 \leq 9/2$ and

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

Here $I_{\Delta^d e(h)}(\lambda_j)$ is the periodogram of a fractionally differenced series at the Fourier frequencies $\lambda_j = 2\pi j/n$, $j = 1, \dots, [n/2]$, where $[x]$ represent the nearest integer less than or equal to x . Precisely, it is given as

$$I_{\Delta^d e}(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n \Delta^d e_t(h) \exp(-it\lambda_j) \right|^2,$$

where the fractional differencing is given by

$$\Delta^d e_t(h) = (1 - B)^d e_t(h) = \sum_{k=0}^{t-1} \frac{\Gamma(-d+k)}{\Gamma(-d)\Gamma(k+1)} (Y_{t-k} - \hat{f}_h(t-k)).$$

The ELW bandwidth refers to the number of low frequencies m used in the estimation.

Recall that the proposed bandwidths selector is given by

$$(\hat{h}_{lw}, \hat{m}_{lw}) = \underset{h,m}{\operatorname{argmin}} \mathbb{E} \left(\hat{d}(m, h) - d \right)^2, \quad (2.3)$$

where $\hat{d}(m, h)$ is the ELW estimation of the LRD parameter from the residual series with kernel bandwidth h and ELW bandwidth m . Therefore, the MSE of the LRD parameter estimation is regarded as a function of the kernel bandwidth h and ELW bandwidth m so that the best bandwidths can be selected by minimizing a single quadratic loss. However, the true LRD parameter d is unknown; and need to be estimated from the data. The so-called block bootstrap method in Hall *et al.* (1995) and Zhou and Taqqu (2007) is used here, but other bootstrapping such as the Sieve bootstrap studied in Poskitt (2008) and the frequency domain methods used by Kim and Nordman (2013) can also be used accordingly.

The block bootstrap sample of residual series $\{e_t(h)\}$, for given bandwidth h , is obtained as follows. First, center the residuals

$$\hat{e}_t(h) := e_t(h) - \bar{e}(h), \quad \bar{e}(h) = n^{-1} \sum_t e_t(h),$$

and draw the starting point of the new block i_m uniformly from $\{1, \dots, n - \ell\}$ where ℓ is the block size. Then, sample a block of observations $(\hat{e}_{i_m}(h), \dots, \hat{e}_{i_m+\ell-1}(h))$ with replacement b times until it has more than n observations. A final bootstrap pseudo-series is obtained by taking only the first n observations. A block size is critical in finite sample performance, and following the suggestion of Politis and White (2004), this paper uses the block size determined by

$$\ell = 2 \min_k \left\{ k \geq 1 \text{ such that } |\hat{\rho}(k)| \leq \frac{2}{\sqrt{n}} \right\}, \quad (2.4)$$

where $\hat{\rho}(k)$ is a sample autocorrelation function at lag k based on $e_t(h)$. This rule selects the block size data adaptively by considering lags with negligible autocorrelations. The constant 2 in (2.4) is

to accommodate negative lags. Politis and White (2004) proved the validity of rule (2.4) for weakly dependent series including polynomial decaying autocorrelations. However, the rigorous proof for strongly dependent series remains open.

Finally, the bootstrap strategy for selecting optimal tuning parameters is described as follows:

- Step 1. Obtain an initial estimate of $\hat{d}^{(0)}$ from Hurvich *et al.* (2005) and set $m^{(0)} = \lceil n^{0.8} \rceil$.
- Step 2. Iterate the following procedures until the relative ratio is within the error bound,

$$\left| \frac{\hat{d}^{(i+1)} - \hat{d}^{(i)}}{\hat{d}^{(i)}} \right| \leq \varepsilon.$$

1. Update a kernel bandwidth from

$$h^{(i+1)} = \operatorname{argmin}_h \mathbb{E}^* \left(\hat{d}^*(h, m^{(i)}) - \hat{d}^{(i)} \right)^2,$$

where \mathbb{E}^* represents empirical average over n_B bootstrap replications and $\hat{d}^*(h, m^{(i)})$ is an ELW estimator from the block bootstrap sample of residuals $\hat{\varepsilon}_t^*(h)$ with block length selection rule (2.4) and $m^{(i)}$ number of frequencies.

2. Update an ELW number of frequency by

$$m^{(i+1)} = \operatorname{argmin}_m \mathbb{E}^* \left(\hat{d}^*(h^{(i+1)}, m) - \hat{d}^{(i)} \right)^2,$$

where the ELW estimator $\hat{d}^*(h^{(i+1)}, m)$ is calculated from the bootstrap sample of residuals $\hat{\varepsilon}_t^*(h^{(i+1)})$ with block length selection rule (2.4).

3. Update an ELW estimator

$$\hat{d}^{(i+1)} = \hat{d}(h^{(i+1)}, m^{(i+1)}).$$

Remark 1. Hurvich *et al.* (2005) is used for initial estimator $\hat{d}^{(0)}$ for its consistency, but any consistent estimator suffices.

3. Monte Carlo simulations study

This section reports the finite sample performance of the proposed method through extensive Monte Carlo simulations. Four data generating processes (DGPs) are considered as follows:

- (DGP1) $f_1(x) = 1$,
- (DGP2) $f_2(x) = \sin(2\pi x)$,
- (DGP3) $f_3(x) = 300x^3(1-x)^3$,
- (DGP4) $f_4(x) = 10x^4$.

with Gaussian FARIMA(0, d , 0) processes for long memory errors given as

$$(1 - B)^d \varepsilon_t = u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2).$$

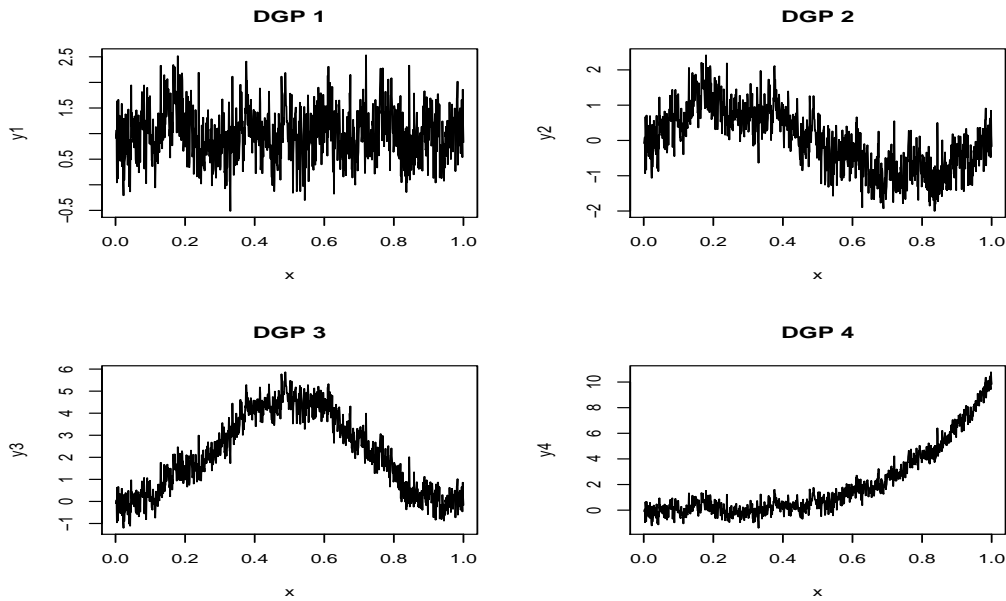


Figure 1: Time plots of DGPs considered in the simulations.

DGP1 is a constant function so that it essentially the same as to estimate the LRD parameter. It is included to see whether the proposed method works well even without any obvious trend. DGP2 considers a cyclic trend; and this makes the long memory parameter estimation harder (e.g. Baek and Pipiras (2014)). DGP3 is used in Chu and Marron (1991) and DGP4 appears in Hurvich *et al.* (2005). Figure 1 depicts a realization of each DGP with FARIMA(0, 0.3, 0) errors with sample size $n = 1,000$.

To compare a new tuning parameters selection rule based on ELW criterion, it is compared to four other methods. The first method uses a bimodal kernel described in Kim *et al.* (2009) to estimate the optimal bandwidth for the mean function, say h_b , and block bootstrap residuals are obtained by

$$e_t(h_b) = Y_t - \hat{f}_{h_b}(t), \quad t = 1, \dots, n,$$

to estimate the LRD parameter. Then, the optimal ELW bandwidth is obtained by iterating

$$m_b^{(i+1)} = \operatorname{argmin}_m \mathbb{E}^* \left(\hat{d}^*(h_b, m) - \hat{d}(h_b, m_b^{(i)}) \right)^2 \quad (3.1)$$

until convergence with $m^{(0)} = \lceil n^{0.8} \rceil$.

The second method is the MCV of Chu and Marron (1991) with the adaptive choice of block length. As detailed in Kim *et al.* (2009), the block length can be determined by finding

$$\hat{k} = \min_k \left\{ k \geq 1 \text{ such that } |\hat{\rho}(k)| \leq \frac{2}{\sqrt{n}} \right\},$$

where $\rho(k)$ is the sample autocorrelation from the residual series. After finding the optimal bandwidth, say h_m , the remaining procedure is the same as the above. That is, similar to (3.1), an ELW estimator is

Table 1: MSE and ASE($\times 1000$) for DGP1 and DGP2 with sample size $n = 1,000$

d	DGP1					DGP2					
		ELW	Bimodal	MCV	Oracle	HLS	ELW	Bimodal	MCV	Oracle	HLS
0.10	MSE	0.350	0.327	0.357	0.320	0.368	0.650	0.436	0.613	0.325	0.358
	ASE	0.520	0.992	2.616	0.673		15.271	7.067	7.109	5.281	
0.20	MSE	0.215	0.225	0.353	0.210	0.462	0.379	0.461	1.040	0.206	0.443
	ASE	0.976	2.725	6.618	1.781		25.283	13.546	14.747	10.064	
0.30	MSE	0.168	0.199	0.323	0.163	0.331	0.216	0.526	1.152	0.173	0.366
	ASE	3.202	7.256	14.563	4.078		34.038	24.792	27.558	17.759	
0.35	MSE	0.152	0.183	0.346	0.156	0.335	0.215	0.577	1.710	0.176	0.317
	ASE	5.953	11.907	21.398	6.404		42.610	36.941	39.845	25.105	
0.40	MSE	0.208	0.291	0.540	0.206	0.347	0.218	0.433	2.096	0.184	0.357
	ASE	10.293	19.291	30.338	9.638		57.916	48.812	51.644	32.486	
0.45	MSE	0.167	0.249	0.796	0.157	0.248	0.187	0.462	2.864	0.166	0.265
	ASE	14.314	28.314	43.282	13.078		77.455	62.660	67.712	40.986	

MSE = mean squared error; ASE = average sum of squares; DGP = data generating processes; ELW = exact local Whittle estimation; MCV = modified cross-validation; HLS = Hurvich *et al.* (2005) method.

calculated with the optimal bandwidth minimizing quadratic error estimated from the block bootstrap samples of residuals $e_t(h_m)$ with $m^{(0)} = [n^{0.8}]$.

The third method is the oracle bandwidth assuming that the true mean function is known. It is defined as

$$h_{\text{opt}} = \underset{h}{\operatorname{argmin}} n^{-1} \sum_{t=1}^n \left(\hat{f}_h(t) - f\left(\frac{t}{n}\right) \right)^2$$

and the LRD parameter is estimated similarly as described above with block bootstrapping residuals $e_t(h_{\text{opt}})$. In addition to three LRD parameter estimators described above, Hurvich *et al.* (2005) is also considered for comparison where no iterative procedure is applied. The detailed tuning parameters for Hurvich *et al.* (2005) are the same as in their paper with $n^{0.15}$ trimming; and denoted as HLS.

All results are based on $N = 500$ replications and the nonparametric regression estimator $\hat{f}_h(t)$ uses the kernel function in Robinson (1997), $K(x) = 0.5(1 + \cos(\pi x))$, $|x| \leq 1$, for consistency. The FARIMA(0, d , 0) errors ϵ_t are generated with standard deviation 0.5 regardless of LRD parameter d by setting $\sigma_u^2 = 0.25\Gamma^2(1 - d)/\Gamma(1 - 2d)$. Bootstrap replication n_B is also important for the performance and efficiency of the algorithm, and it is found that $n_B = 50$ is quite successful. The performance measures are the (empirical) MSE

$$\text{MSE} = \mathbb{E}^* \left(\hat{d}(h, m) - d \right)^2$$

and the (empirical) average sum of squares (ASE)

$$\text{ASE} = \mathbb{E}^* \left(n^{-1} \sum_{t=1}^n \left(\hat{f}_h(t) - f\left(\frac{t}{n}\right) \right)^2 \right).$$

Table 1 shows the MSE for the LRD parameter and the ASE($\times 1000$) for all five methods when DGP1 and DGP2 are considered. For DGP1, note that our proposed method based on ELW performs nicely in all cases considered for both MSE and ASE. The MCV method performs the worst for both MSE and ASE, but this is consistent with the simulations results in Hall *et al.* (1995), where they also reported that MCV tends to find a smaller bandwidth due to a flatter MISE curve leading to

Table 2: MSE and ASE($\times 1000$) for DGP3 and DGP4 with sample size $n = 1,000$

d		DGP3					DGP4				
		ELW	Bimodal	MCV	Oracle	HLS	ELW	Bimodal	MCV	Oracle	HLS
0.10	MSE	0.548	0.422	0.674	0.362	0.381	0.733	0.633	0.906	0.691	0.334
	ASE	11.683	7.748	7.319	5.492		15.727	11.860	12.027	10.345	
0.20	MSE	0.303	0.465	1.088	0.173	0.397	0.545	1.539	2.874	0.454	0.412
	ASE	23.227	14.242	14.847	10.529		34.601	20.970	23.510	18.306	
0.30	MSE	0.156	0.534	0.489	0.141	0.366	0.190	1.583	4.603	0.171	0.378
	ASE	32.797	26.374	29.880	18.795		50.370	36.818	41.957	31.055	
0.35	MSE	0.146	0.523	2.046	0.144	0.324	0.193	1.310	6.464	0.175	0.304
	ASE	43.288	38.894	42.613	26.879		61.549	49.826	56.411	40.531	
0.40	MSE	0.185	0.567	2.879	0.178	0.313	0.190	1.016	7.661	0.181	0.309
	ASE	72.083	52.771	57.891	35.278		72.420	65.500	73.650	51.750	
0.45	MSE	0.165	0.499	4.04	0.168	0.267	0.157	0.721	8.483	0.163	0.262
	ASE	121.465	67.802	75.012	46.236		87.726	84.667	93.999	67.795	

MSE = mean squared error; ASE = average sum of squares; DGP = data generating processes; ELW = exact local Whittle estimation; MCV = modified cross-validation; HLS = Hurvich *et al.* (2005) method.

Table 3: MSE and ASE($\times 1000$) for DGP4 with sample size $n = 2,000$ and $n = 5,000$

d		$n = 2,000$					$n = 5,000$				
		ELW	Bimodal	MCV	Oracle	HLS	ELW	Bimodal	MCV	Oracle	HLS
0.40	MSE	0.096	0.592	5.298	0.115	0.155	0.050	5.350	3.523	0.059	0.062
	ASE	58.497	67.460	72.098	46.592		51.439	63.874	62.942	37.671	
0.45	MSE	0.077	0.506	7.416	0.102	0.149	0.046	3.375	6.721	0.060	0.065
	ASE	81.769	91.124	95.648	61.356		78.854	95.686	89.942	54.041	

MSE = mean squared error; ASE = average sum of squares; DGP = data generating processes; ELW = exact local Whittle estimation; MCV = modified cross-validation; HLS = Hurvich *et al.* (2005) method.

numerically unstable values. Instead, the bimodal kernel method, which is a continuous analogue of MCV, seems to be more numerically stable than MCV in this case. ELW performs even better than the oracle method in some cases, but the true trend is only a constant function so it may do so due to sampling fluctuations.

When the cyclic trend is considered as in DGP2, however, the oracle method shows the smallest MSE and ASE as expected. However, observe that our ELW method is closest to the oracle in terms of MSE for moderate to large LRD parameters. Indeed, the proposed ELW method works well for estimating the LRD parameter in the presence of a smooth trend. However, the ASE is closest to the oracle when the bimodal kernel method is used. Observe also that the ASE increases as the LRD parameter d increases, and coincides with a strong correlation that masks the true trend function.

Table 2 shows the results for DGP3 and DGP4. The overall interpretations are similar to DGP2 with an emphasis that the proposed ELW method is closest to the oracle as LRD parameter d increases. The ASE is closest to the oracle when the bimodal kernel method is used. However, the proposed ELW method outperform existing methods if the sample size is increased and the persistency becomes stronger. For example, Table 3 shows the results for DGP4 with a quadratic trend when the sample size increases to $n = 2,000$ and $5,000$ with LRD parameters $d = 0.4$ and $d = 0.45$, respectively. Now, the proposed ELW method outperforms MCV and bimodal bandwidth selectors for both LRD parameter estimation and smooth function estimation. Figure 2 shows the boxplots of ASE and the estimated LRD parameter for DGP4 with $d = 0.4$ when the sample size $n = 5,000$. Bimodal or MCV methods sometimes find the LRD estimator close to zero, but this may be because the trend function is oversmoothed due to the smaller bandwidth as already reported in Hall *et al.* (1995). In summary,

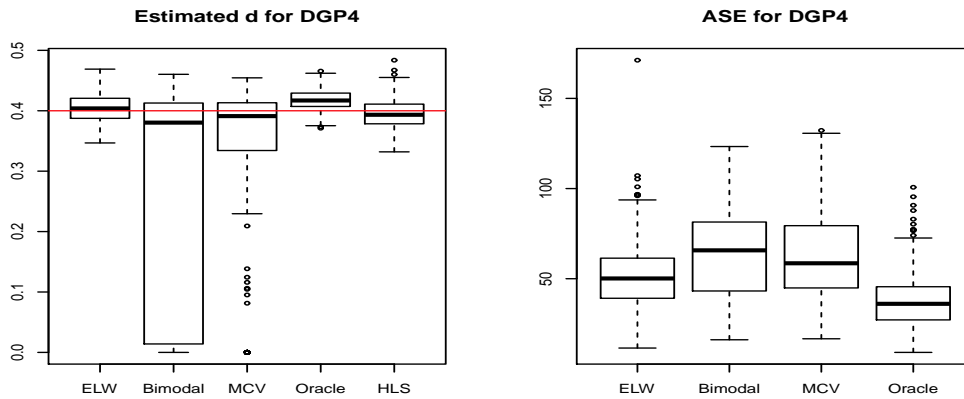


Figure 2: Time plots of estimated d and ASE for DGP4 with $d = 0.4$ and sample size $n = 5,000$. DGP = data generating processes; ASE = average sum of squares; ELW = exact local Whittle estimation; MCV = modified cross-validation; HLS = Hurvich *et al.* (2005) method.

this simulation study shows that the proposed ELW method successfully estimates the long memory parameter in the presence of a trend. The proposed ELW method particularly outperforms the other methods as dependency and sample size increase.

4. Real data application

To illustrate our proposed method based on ELW estimation of LRD parameter, we have considered the volatility of exchange rate between Korean Won (KRW) for one US dollar (USD). The index is expressed in local currency, that is, in exchange of 1 US dollar to KRW, and we have considered exchange rate from Jan 1, 2002 to Dec 31, 20013. It is widely recognized that the volatility exhibits both non-stationary and LRD properties as it is nicely documented in Stărică and Granger (2005).

We study the power-transformed absolute differences,

$$Y_i = |I_i - I_{i-1}|^{0.25},$$

where I denotes the original exchange rate. The total number of observations is 3,027. The reason for taking a quarter transformation is to make the series close to Gaussian, and studied similarly in the literature (e.g. Ding *et al.* (1993)).

Figure 3 represents observations $\{Y_i\}$ (top left) and its corresponding normal QQ-plot (top right), the sample autocorrelations (SACF) plot (bottom left) and ELW estimator according to frequencies selected (bottom right). From the visual inspection of plots, first, there is a smooth concave-like trend in the middle of observations, and also a quarter transformation seems to make the marginal distribution close to normal. Observe also that the sample autocorrelations decay very slowly and stays very high even for lag 100. This clearly indicates that our observations are very strongly correlated. ELW parameter estimates are positive and stay away from zero for wide range of frequencies used, so we observe both non-stationarity and LRD in the volatility of exchange rate between KRW and USD.

We have applied our proposed method to estimate a smooth trend perturbed by strongly correlated errors. Figure 4 shows estimated smooth trend and the SACF of absolute residuals. However, it still shows slowly decaying autocorrealtions that are weaker than the original series. This is also observed from ELW estimator plot, showing smaller LRD parameter values. The resulting ELW estimator is

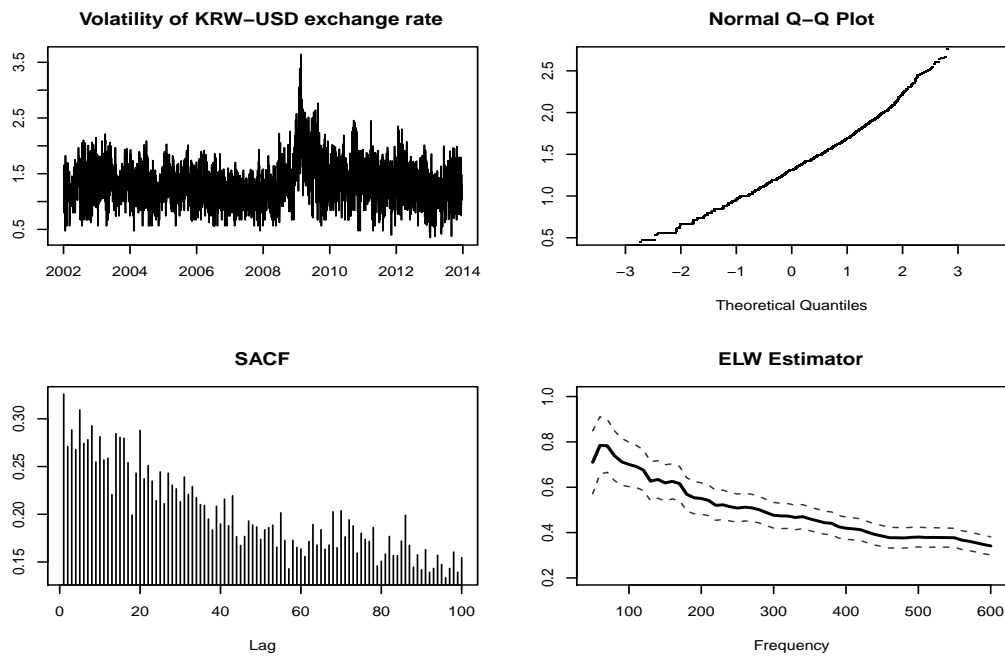


Figure 3: The volatility of KRW and USD during 2002–2007 with the SACFs and ELW LRD parameter estimates. SACF = sample autocorrelations; ELW = exact local Whittle estimation; LRD = long-range dependence.

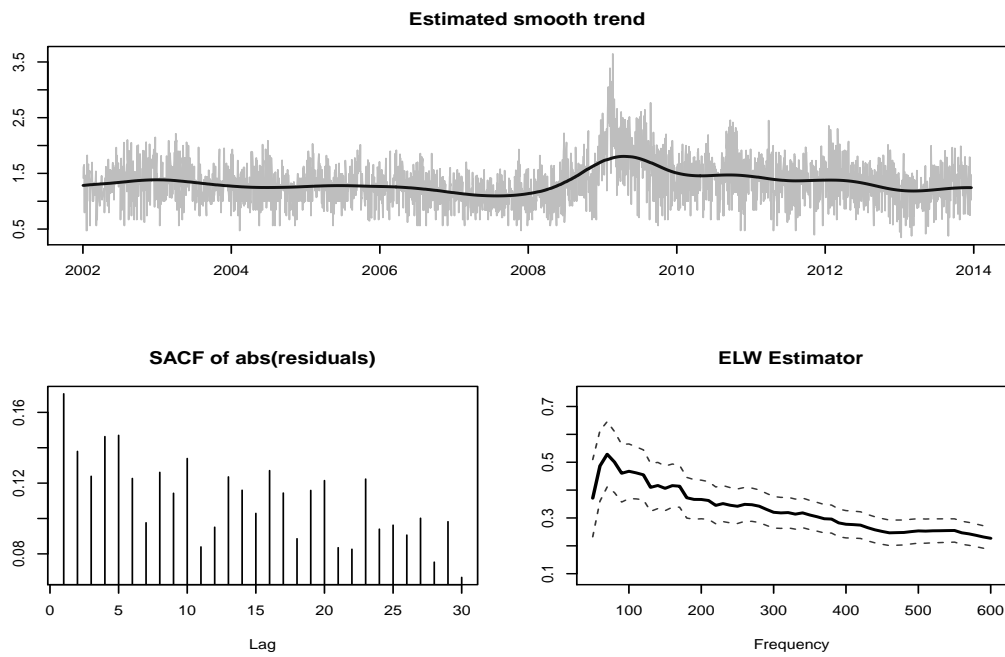


Figure 4: Estimated smooth trend, correlograms on absolute residuals and ELW estimation from residuals. SACF = sample autocorrelations; ELW = exact local Whittle estimation.

$\hat{d} = 0.3325$ from bandwidths $\hat{h}_{lw} = 0.146$ and $\hat{m}_{lw} = 233$ with 95% confidence interval (0.267, 0.396). Hence, this analysis shows that volatility of KRW-USD exchange rate has both smooth trend and stationary LRD errors. This provides an alternative modeling of non-stationary-like observations in the framework of everlasting debate between changes-in-mean and long memory. See, for example, Lee *et al.* (2015), Song and Baek (2019) and references for further discussions.

5. Conclusions

A new tuning parameter selection rule is proposed for the long memory parameter estimation in the presence of a smooth trend. Tuning parameters are selected by minimizing the single MSE of the long memory parameter from the residuals. A simulations study shows outstanding performance of the proposed method. It was closest to the oracle, and outperformed other methods as dependency and sample size increase. It also remains an interesting future work on extension to bivariate LRD series as studied in Baek *et al.* (2020).

Acknowledgments

This work was supported by the Basic Science Research Program from the National Research Foundation of Korea (NRF-2017R1A1A1A05000831, NRF-2019R1F1A1057104).

References

- Baek C, Kechagias S, and Pipiras V (2020). Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity. *Journal of Statistical Planning and Inference*, **205**, 245–268.
- Baek C and Pipiras V (2014). On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation. *Electronic Journal of Statistics*, **8**, 931–964.
- Beran J and Feng Y (2002). SEMIFAR models—a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, **40**, 393–419.
- Chu CK and Marron JS (1991). Comparison of two bandwidth selections with dependent errors. *The Annals of Statistics*, **19**, 1906–1918.
- Ding Z, Granger CWJ, and Engle RF (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83–106.
- Geweke J and Porter-Hudak S (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.
- Hall P and Hart JD (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and their Applications*, **36**, 339–351.
- Hall P, Lahiri SN, and Polzehl J (1995). On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *The Annals of Statistics*, **23**, 1921–1936.
- Henry M (2001). Robust automatic bandwidth for long memory. *Journal of Time Series Analysis*, **22**, 293–316.
- Hurvich CM, Gabriel L, and Philippe S (2005). Estimation of long memory in the presence of a smooth nonparametric trend. *Journal of the American Statistical Association*, **100**, 853–871.
- Kim TY, Park BU, Moon MS, and Kim, C (2009). Using bimodal kernel for inference in nonparametric regression with correlated errors. *Journal of Multivariate Analysis*, **100**, 1487–1497.
- Kim Y and Nordman DJ (2013). A frequency domain bootstrap for Whittle estimation under long-range dependence. *Journal of Multivariate Analysis*, **115**, 405–420.

- Künsch H (1986). Discrimination between Monotonic Trends and Long-Range Dependence, *Journal of Applied Probability*, **23**, 1025–1030.
- Lee T, Kim M, and Baek C (2015). Tests for volatility shifts in GARCH against long-range dependence, *Journal of Time Series Analysis*, **36**, 127–153.
- Masry E and Mielniczuk J (1999). Local linear regression estimation for time series with long-range dependence, *Stochastic Processes and their Applications*, **82**, 173–193.
- Opsomer J, Wang Y, and Yang Y (2001). Nonparametric regression with correlated errors, *Statistical Science. A Review Journal of the Institute of Mathematical Statistics*, **16**, 134–153.
- Politis DN and White H (2004). Automatic block-length selection for the dependent bootstrap, *Econometric Reviews*, **23**, 53–70.
- Poskitt DS (2008). Properties of the Sieve bootstrap for fractionally integrated and non-invertible Processes, *Journal of Time Series Analysis*, **29**, 224–250.
- Ray BK and Tsay RS (1997). Bandwidth selection for kernel regression with long-range dependent errors, *Biometrika*, **84**, 791–802.
- Robinson PM (1995). Gaussian semiparametric estimation of long range dependence, *The Annals of Statistics*, **23**, 1630–1661.
- Robinson PM (1997). Large-sample inference for nonparametric regression with dependent errors, *The Annals of Statistics*, **25**, 2054–2083.
- Shimotsu K and Phillips PCB (2005). Exact local Whittle estimation of fractional integration, *The Annals of Statistics*, **33**, 1890–1933.
- Song J and Baek C (2019). Detecting structural breaks in realized volatility, *Computational Statistics & Data Analysis*, **134**, 58–75.
- Stărică C and Granger C (2005). Nonstationarities in stock returns, *The Review of Economics and Statistics*, **87**, 503–522.
- Zhou Y and Taqqu MS (2007). Applying bucket random permutations to stationary sequences with long-range dependence, *Fractals*, **15**, 105–126.

Received August 5, 2019; Revised September 27, 2019; Accepted October 6, 2019