MS Creative Component: A simulation study of bandwidth selection methods for non-parametric regression with long and semi-long range dependent errors

Zirou Zhou

May 2021

1 Introduction

Assume we observe random variables $(X_i, Y_i)_{i=1}^n$, where Y_i takes values in $\mathbb{R}$ and X_i takes values in $\Omega \subseteq \mathbb{R}$ with a common density f . The density f is compactly supported, bounded and continuous with $f(x) > 0$ for all x. Consider the model

$$Y_i = m(X_i) + e_i \ (i = 1, ..., n),$$

where m is an unknown smooth regression function and the e_i are unobserved random variables such that

$$\mathrm{E}(e_i|X_i) = 0, \ \mathrm{Cov}(e_i, e_j|X_i, X_j) = \sigma_e^2 \rho_n(X_i - X_j) \ (i, j = 1, ..., n), \quad (1)$$

with $\sigma^2 < \infty$ and ρ_n a stationary correlation function satisfying $\rho_n(0) = 1$, $\rho_n(-x) = \rho_n(x)$ and $|\rho_n(x)| \leq 1$ for all x. The subscript n allows the correlation function ρ_n to shrink as $n \to \infty$.

Many non-parametric kernel based regression methods require the tuning or selection of a smoothing parameter, often referred to as the bandwidth. This introduces a certain degree of arbitrariness in the estimation procedure. Therefore, a vast number of bandwidth selection methods have been developed Fan and Gijbels (1996); Konishi and Kitagawa (2008). Diggle and Hutchinson (1989) and Hart (1991) have shown that these methods perform rather poorly under (1), since they typically require the assumption that the errors are independent and identically distributed (i.i.d.) .

Several modifications have been proposed to account for short-range correlation Opsomer et al. (2001); Francisco-Fernández et al. (2005). The plug-in method proposed by Francisco-Fernández et al. (2005) requires estimating the correlation structure of the model of order 1, and can only be applied when we have equally spaced design points. Kim et al. (2009) showed that bimodal kernels are quite effective even when the errors are heavily correlated. However, three shortcomings remain: prior information about the correlation structure is required, bandwidth selection methods need to be modified to work in this context, and most theoretical results have been established for fixed equispaced designs. De Brabanter et al. (2018) introduced a whole new method by using a bimodal kernel function K such that $K(0) = 0$. They showed that by using this kernel and minimizing the Residual sum of squareS (RSS), the bandwidth selection procedure is asymptomatically optimal for short range dependent errors. De Brabanter & Sabzikar (2021) expended this theory to errors with fractional local to unity root (FLUR) model structure. FLUR models are stationary time series with semi-long range dependence property in the sense that their covariance function resembles that of a long memory model for moderate lags but eventually diminishes exponentially fast according to the presence of a decay factor governed by an exponential tempering parameter. When this parameter is sample size dependent, the asymptotic theory for these regression models admit a wide range of stochastic processes with behavior that includes long, semi-long, and short memory processes.

As the tempering parameter is sample size dependent, in this CC, we will empirically testify how the sample size effects the bandwidth choice and rate of convergence, with long, semi-long, and short memory processes. We will provide simulation results, and explanations for our conjectures.

1.1 Introductory about Fractional Local Unity Root and Tempered Linear Process

The Local unity root (LUR) model time series {X(t)} is defined as

$$X(t) = \rho_N X(t-1) + \xi(t), t = 1, ..., N, X(0) = 0$$

Where $\rho_N = 1 - \frac{c}{N}$ for $c > 0$, $\xi(t)$ are i.i.d. innovations and N is the sample size. By using the operator $BX(t) = X(t-1)$, the model is expressed as

$$X_{d,\rho_N}(t) = \sum_{k=0}^{\infty} \rho_N^k \omega_{-d} \xi(t-k), \ t \in \mathbb{Z}$$

where

$$\omega_{-d}(k) = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}, d \in \mathbb{R} \setminus \mathbb{N}_{-}$$

As $N \to \infty$, $\rho_N^k \to 1$,so that $X_{d,1}$ forms a fractional time series including the well known ARFIMA model with long-range dependence.

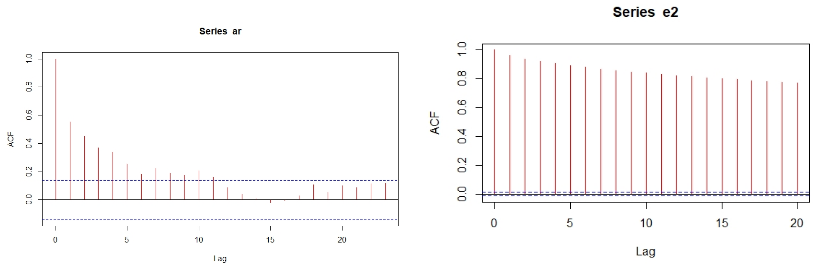


Figure 1: The difference between an AR (short-range dependence) and ARFIMA (long-range dependence)

The tempered linear process (TLP) with moving averages is

$$X_{d,\lambda}(j) = \sum_{k=0}^{\infty} e^{-\lambda k} b_d(k)\xi(j-k), j \in \mathbb{Z}$$

where $\xi(j)$ are i.i.d. innovations, $b_d(k)$ regularly varying at infinity as k^{d-1}.

Since $\rho_N \sim e^{-\lambda N}$ as $N \to \infty$, we can replace X_{d,ρ_N} with X_{d,λ_N} as the tempered linear process.

1.2 Tempering parameters for long, semi-long, and short memory processes

The main difference between short memory process (short range dependent, SRD) and long memory process (long range dependent, LRD), semi-long memory process (semi-long range dependent, SLRD) is mainly dependent on the memory parameter d & tempering parameter λ_*. The SRD has $\lambda_* = \infty$, the LRD and SLRD have $\lambda_* < \infty$. And whether $d < 1/2$ is the the main different between LRD and SLRD.

The autocovariogram of FLUR and TLP resemble long range dependent series out to moderate lag lengths but eventually decays exponentially fast. Giraitis et al. (2000) named this behavior semi-long memory which is analogous to the semi-heavy tail property in Barndorff-Nielsen (1998). Giraitis et al. (2003) introduced semi-long memory ARFIMA(0, d, 0), semi-long memory long memory linear Autoregressive conditional heteroskedasticity (LARCH), and semi-long memory Autoregressive conditional heteroskedasticity (ARCH) processes and used these models to investigate the power and robustness of the R/S type tests under contiguous and semi-long memory alternatives. Dacorogna et al. (1993) and Granger et al. (1996) argued that, for some economic time series, such as the magnitude of certain powers of financial returns, the covariance functions degrade slowly at first but ultimately go off much faster.

Assumptions to obtain the main results:

- The tempering parameter λ_N satisfies $N\lambda_N \to \lambda_* \in [0, \infty]$, as $N \to \infty$.
- The bandwidth h satisfies $h = h_N \to 0$ and $Nh \to \infty$ as $N \to \infty$.
- The kernel K is a symmetric density function with support on [-1, 1] with bounded first derivative K'.

Further, we will have:

(a) For SRD, $d \in \mathbb{R} \setminus \mathbb{N}_-$, especially in $[0, 1]$ and $\lambda_* = \infty$

(b) For LRD, $-1/2 < d < 1/2$, especially in $(0, 1/2)$ and $\lambda_* = 0$

(c) For SLRD, $d \in (0, \infty)$, especially in $(0, 1)$ and $0 < \lambda_* < \infty$

2 Convergence of SRD

This simulation is based on Theorem 2 of De Brabanter et Sabzikar (2021).

$$\frac{\lambda_N^d}{\sqrt{Nh}} \sum_{j=1}^{N} K(\frac{Nx - j}{Nh}) X_{d,\lambda_n}(j) \xrightarrow{d} N(0, \sigma^2_{d,\infty})$$

Thus,

$$\frac{\lambda_N^d}{\sigma_{d,\infty}\sqrt{Nh}} \sum_{j=1}^{N} K(\frac{Nx - j}{Nh}) X_{d,\lambda_n}(j) \xrightarrow{d} N(0, 1)$$

Where σ^2 is the variance of the innovations

$$\sigma^2_{d,\infty} = \sigma^2 \int_{-1}^{1} K^2(u) du$$

By using the Epanechnikov kernel, $\sigma^2_{d,\infty} = 3/5\ \sigma^2$, by using Monte Carlo simulation (500), and $h = 2N^{-1/5}$, the $\lambda_N = \infty$ the assumptions of the theorem are meet.

The range for the memory parameter d is behaving 0.1 and 0.7, and with sample size N = 500 and 20000. The variance of the innovations is 1.

Figure 2 illustrates the normalized densities of the simulated X_{d,λ_n} with short range dependent errors for two different values of the parameter d and sample size N, attached with the stand normal density.

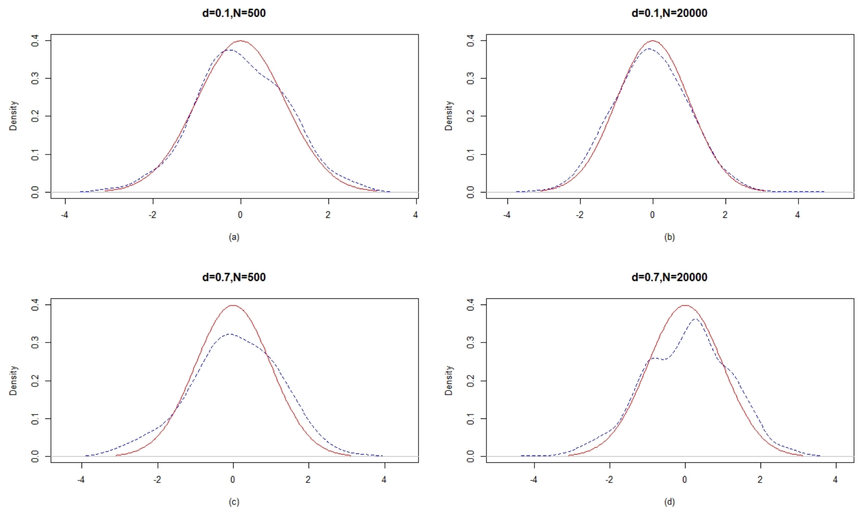


Figure 2: Standard normal density and Kernel density estimations of simulations with $\lambda_N = N^{-1/5}$

With different finite sample sizes, and different d, under the same SRD condition, the densities of X_{d,λ_n} are all converge to the same normal distribution with the $\sigma^2_{d,\infty}$. The convergence of TLP with SRD does not seem to deponent on d and N.

3 Bandwidth choice

3.1 Local polynomial Regression model

Consider the non-parametric equispaced fixed design regression model

$$Y(j) = m(\frac{j}{N}) + X_{d,\lambda_N}(j), j = 1, ..., N$$

where $m(x)$is the unknown regression function and X_{d,λ_N} is a TLP model.

By using the R package artfima (McLeod et al., 2016) to simulate the TLP X_{d,λ_N}, Figure 3 illustrates the relationship between true regression $m(x)$ and the simulated observed data Y.

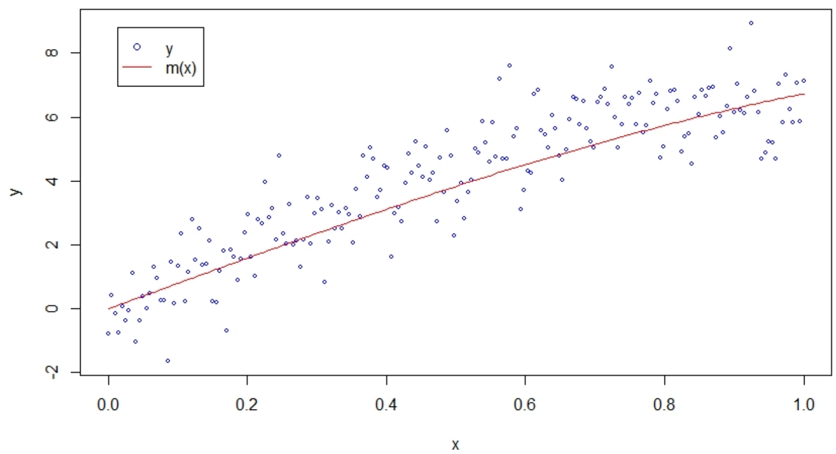


Figure 3: the nonparametric equispaced fixed design regression model

We will try to use local Polynomial regression to estimate the m(x) without knowing the structure of the TLP error. And figure out the how the bandwidth choice changes, especially in these types of stochastic processes.

3.2 LRD Bandwidth choice

According to the remark 2 of De Brabanter et Sabzikar (2021), the optimal bandwidth h_{opt} for Long Range Dependence errors ($d \in (0, 1/2)$) is

$$\begin{aligned} h_{opt} &= \left[\frac{(1-2d)\sigma^2\Gamma(1-2d) \int_{-1}^{1} \int_{-1}^{1} K(u)K(v)|u-v|^{2d-1} du dv}{\Gamma(d)\Gamma(1-d) \int_{-1}^{1} u^2 K(u) du^2 \int_0^1 m''(x)^2 dx} \right]^{\frac{1}{5-2d}} n^{\frac{2d-1}{5-2d}} \\ &= C(d)\ n^{\frac{2d-1}{5-2d}} \end{aligned}$$

Based on the above theoretical results, the order of bandwidth h_{opt} will be $\frac{2d-1}{5-2d}$.

To verify this conjecture, we simulated an data base of FLUR based on the R package artfima (McLeod et al., 2016). The innovations have mean zero and unit variance. And set the $m(x)$ has larger variance to guarantee the signal of noise ratio (SNR).

To estimated the local polynomial regression of degree 3, we need to use the following kernel:

$$K(u) = \frac{u^2}{\sqrt{2\pi}}\ exp(-u^2)$$

The plot of a estimated regression and the true regression with the sample size 500 and d= 0.45:

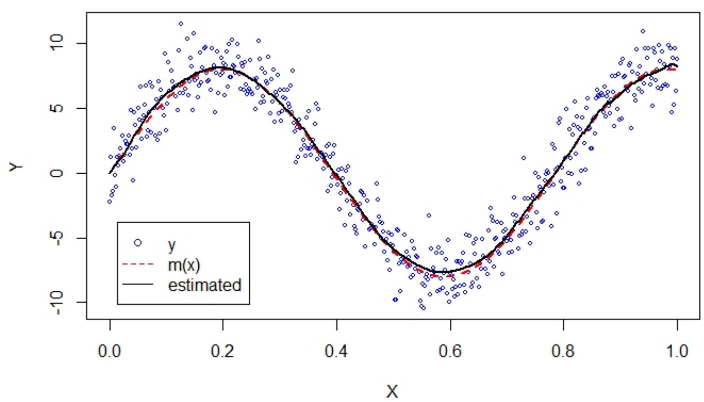
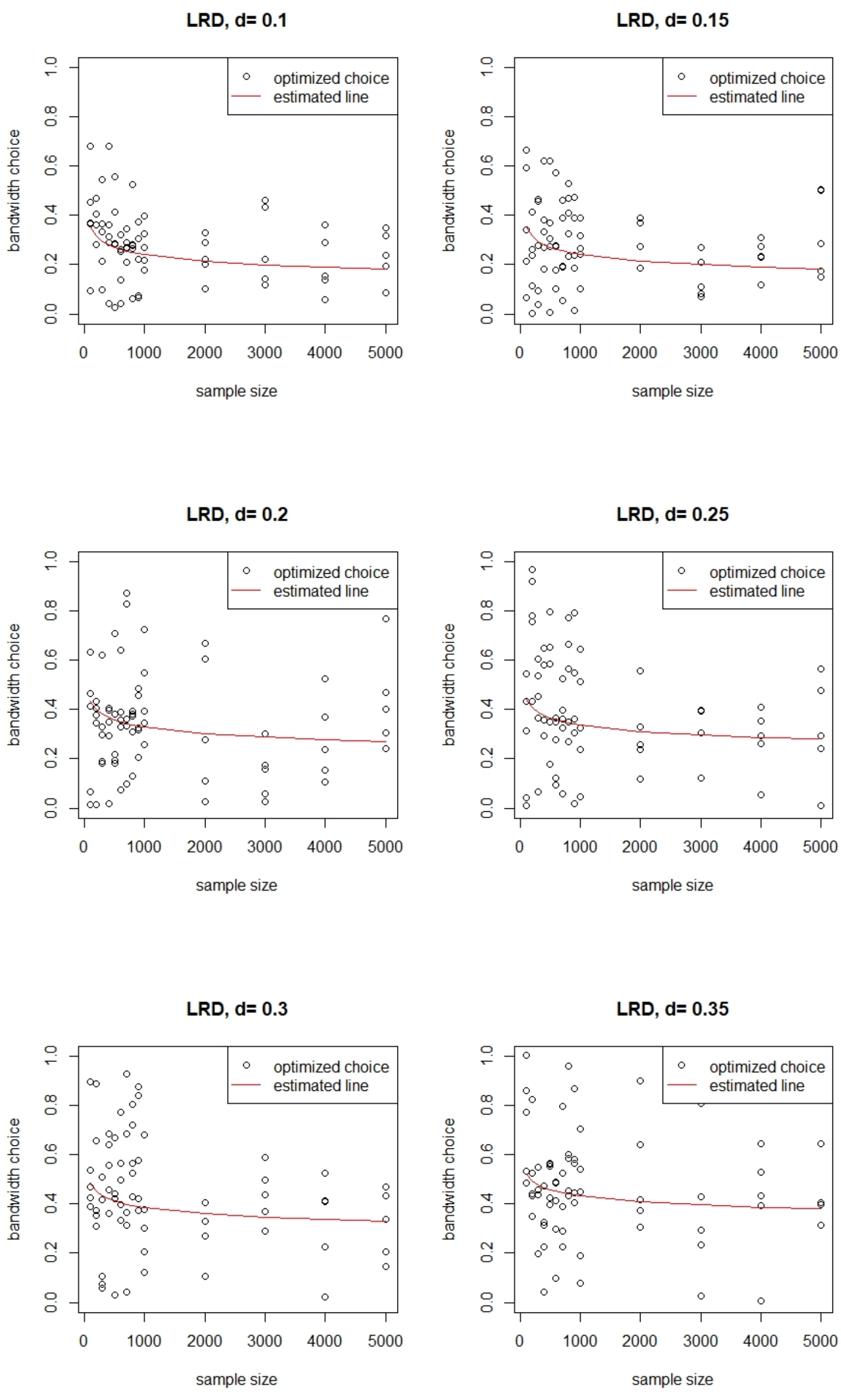


Figure 4: example for estimated Non-parametric regression model with LRD

This plot shows the estimated local polynomial regression has reach the purposes and don't have a obviously overfit problem. The bandwidth choice is rational to be viewed as optimized.

3.2.1 The optimal bandwidth choices from the local polynomial regression for the estimated error

Based on the method showed above, by choosing different d and repeating 5 times, the optimized bandwidth choice for the FLUR with LRD is showed below. and by taking the power regression in the form of $h_{opt} = aN^b$, we can testify relationship between the order of bandwidth choice h_{opt} and the sample size N.


LRD, d= 0.1
LRD, d= 0.15
LRD, d= 0.2
LRD, d= 0.25
LRD, d= 0.3
LRD, d= 0.35
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000

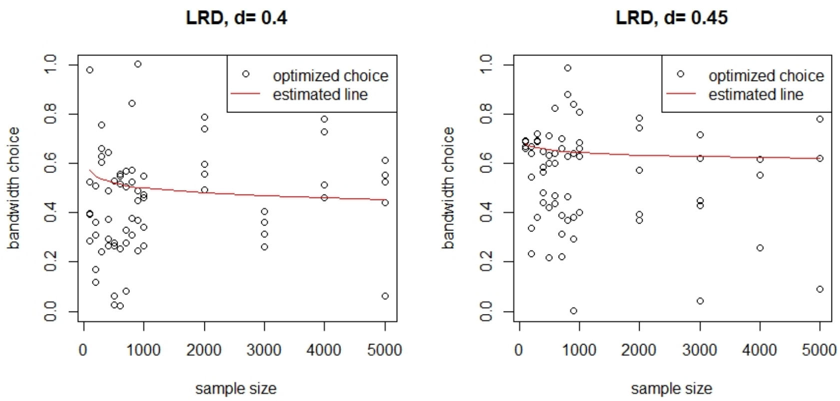


Figure 5: the optimized bandwidth choice for the FLUR with LRD

Table 1: the optimized bandwidth choice for the FLUR with LRD

	sample size	optimized bandwidth choice			
	d=	0.1	0.15	0.2	0.25
1	100	0.3643387	0.3701004	0.4264339	0.4704011
2	200	0.3252611	0.3324544	0.3908325	0.4358066
3	300	0.3043751	0.312234	0.3714013	0.4167618
4	400	0.2903748	0.2986377	0.3582034	0.4037562
5	500	0.2799603	0.2885007	0.3482901	0.3939484
6	600	0.271729	0.2804741	0.3403944	0.386112
7	700	0.2649585	0.2738621	0.3338584	0.3796081
8	800	0.2592303	0.2682608	0.3282983	0.3740628
9	900	0.2542804	0.2634152	0.3234708	0.3692388
10	1000	0.2499328	0.2591548	0.3192126	0.3649764
11	2000	0.2231259	0.232794	0.2925628	0.338135
12	3000	0.2087984	0.2186351	0.2780173	0.3233585
13	4000	0.1991943	0.2091146	0.2681378	0.3132677
14	5000	0.1920501	0.2020164	0.2607171	0.305658

	d=	0.3	0.35	0.4	0.45
1	100	0.5177697	0.602546	0.6199531	0.6578664
2	200	0.4874524	0.5819867	0.6020554	0.6427573
3	300	0.4705477	0.5702871	0.5918264	0.6340804
4	400	0.4589104	0.5621289	0.5846744	0.6279951
5	500	0.4500824	0.5558815	0.5791864	0.6233153
6	600	0.4429955	0.5508285	0.5747407	0.6195175
7	700	0.4370908	0.5465921	0.5710085	0.6163246
8	800	0.4320396	0.5429487	0.5677952	0.613572
9	900	0.4276326	0.5397552	0.5649758	0.6111543
10	1000	0.4237285	0.5369144	0.5624656	0.6089997
11	2000	0.3989177	0.5185945	0.5462276	0.5950128
12	3000	0.3850833	0.5081692	0.5369471	0.5869804
13	4000	0.3755597	0.5008997	0.5304583	0.5813472
14	5000	0.368335	0.4953327	0.5254793	0.577015

3.2.2 Estimation results

By fitting the optimized bandwidth choice from the fitted local polynomial regression with the function of $\hat{h}_{opt} = a \times N^b$

The estimated $\hat{b}$ and the memory parameter d has relationship

Table 2: Order of bandwidth for LRD

d	$\hat{b}$	(2d-1)/(5-2d)
0.1	-0.16368201	-0.16666667
0.15	-0.15476028	-0.14893617
0.2	-0.12577166	-0.13043478
0.25	-0.12020361	-0.11111111
0.3	-0.09704896	-0.09090909
0.35	-0.05008519	-0.06976744
0.4	-0.04826280	-0.04761905
0.45	-0.03352068	-0.02439024

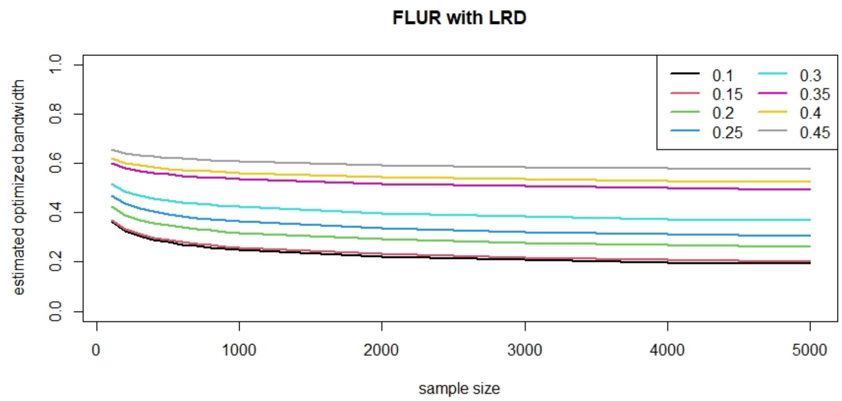


Figure 6: LRD bandwidth choice

The order of the optimized bandwidth choice do have a positive relationship with the memory parameter d, which is basically conform to the theory.

3.3 SLRD Bandwidth choice

For FLUR with SLRD error, based on De Brabanter at al (2018), the conjecture of the h_{opt} is $O(n^{-1/5})$, which is of the same order as short range dependence and i.i.d. cases.

To verify this conjecture, we simulated an data based on R package artfima (McLeod et al., 2016). The innovations have mean zero and SLRD unit variance. And set the $m(x)$ has larger variance to get ride of the overfit problem.

To estimate the local polynomial regression, we need to use the kernel:

$$K(u) = \frac{u^2}{\sqrt{2\pi}} exp(-u^2)$$

And the degree of local polynomials we choose is 3

The plot of a estimated regression and the true regression with the sample size 500 and d= 0.8:

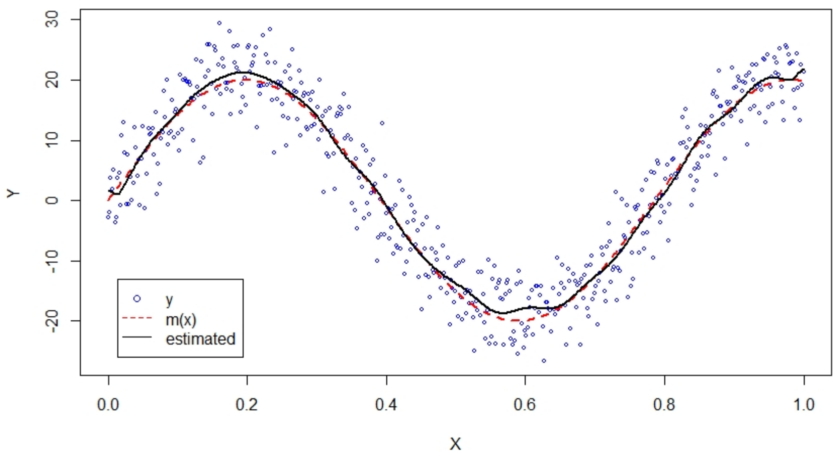
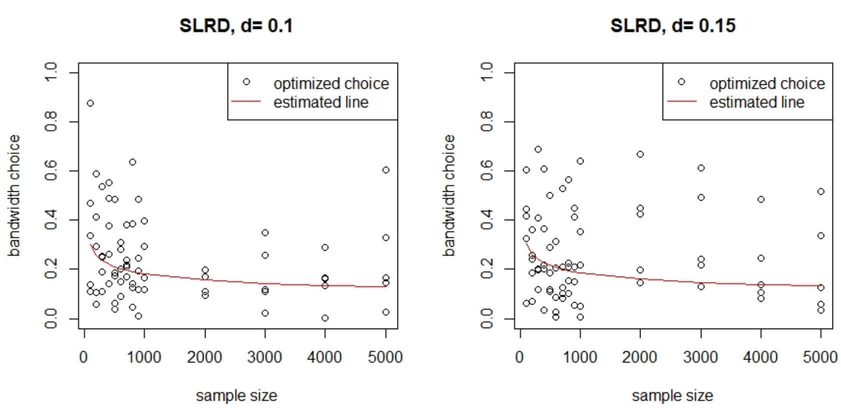


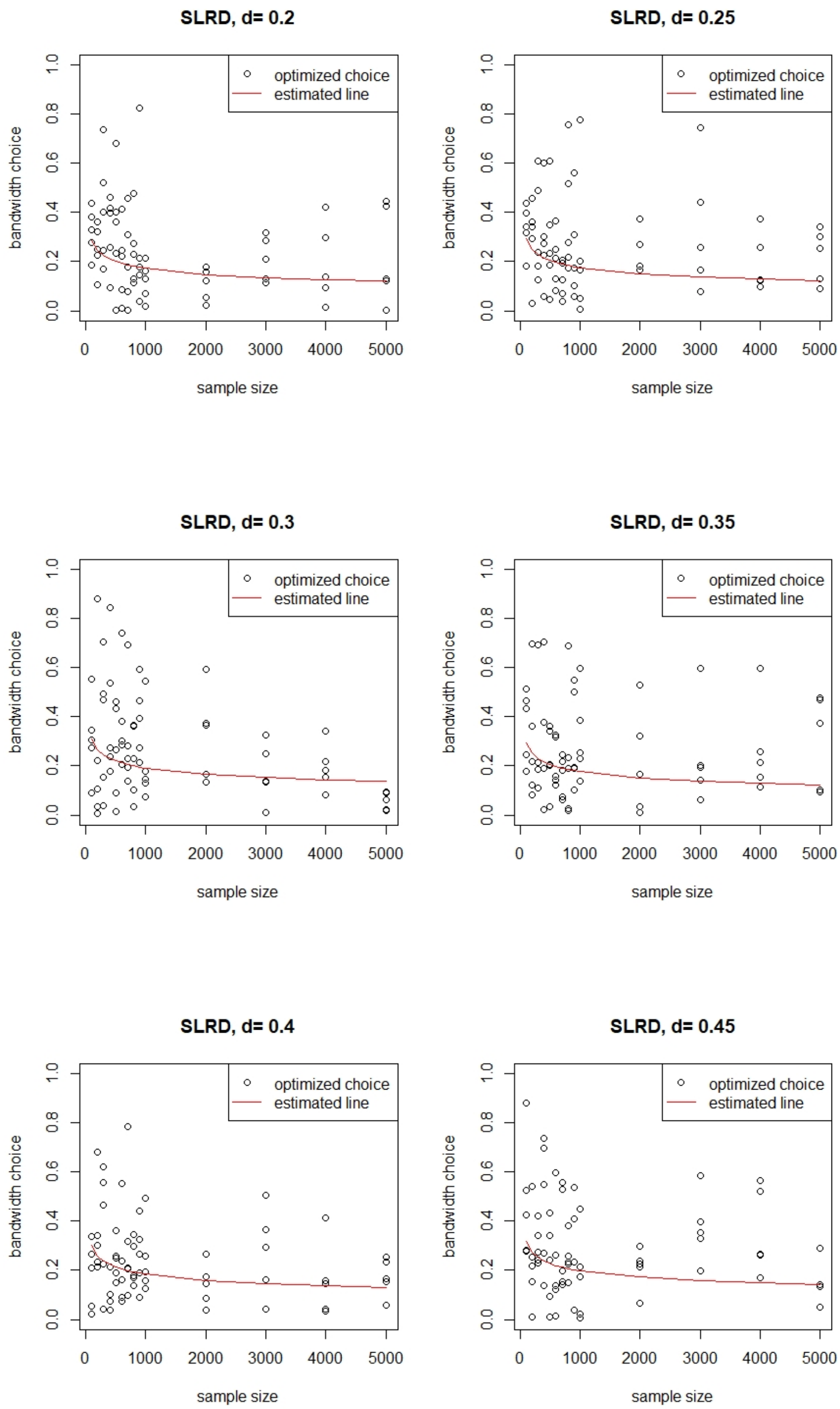
Figure 7: example for estimated Non-parametric regression model with SLRD

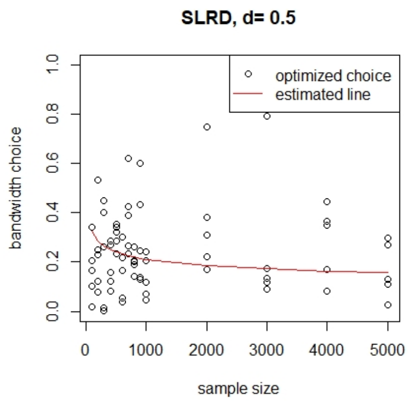
This plot shows the estimated local polynomial regression has reach the purposes and don't have a obviously overfit problem. The bandwidth choice is rational to be viewed as optimized.

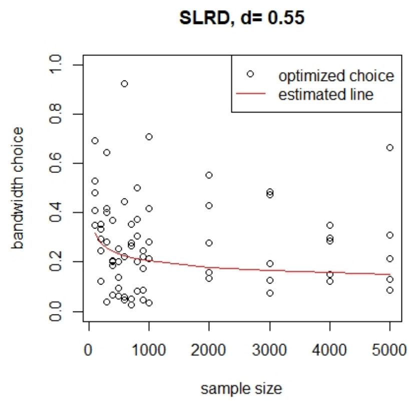
3.3.1 The optimal bandwidth choices from the local polynomial regression for the estimated error

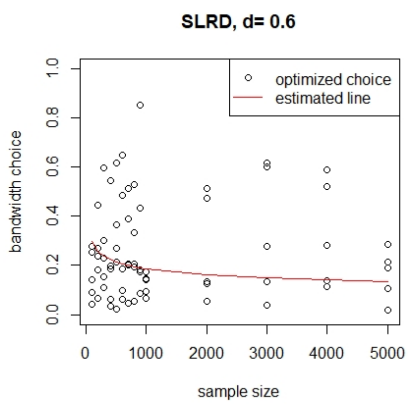
Based on the method showed above, by choosing different d and repeating 5 times, the optimized bandwidth choice for the FLUR with SLRD is shown below. and by taking the power regression in the form of $h_{opt} = aN^b$, we can testify relationship between the order of bandwidth choice h_{opt} and the sample size N.

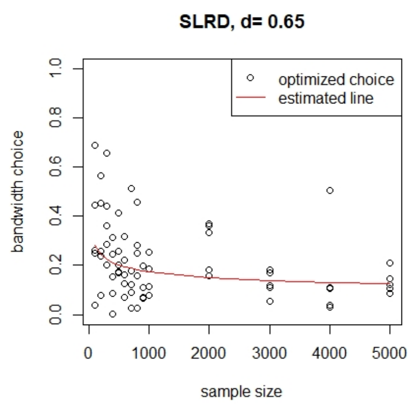



SLRD, d= 0.2
SLRD, d= 0.25
SLRD, d= 0.3
SLRD, d= 0.35
SLRD, d= 0.4
SLRD, d= 0.45
optimized choice
estimated line
bandwidth choice
sample size
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000

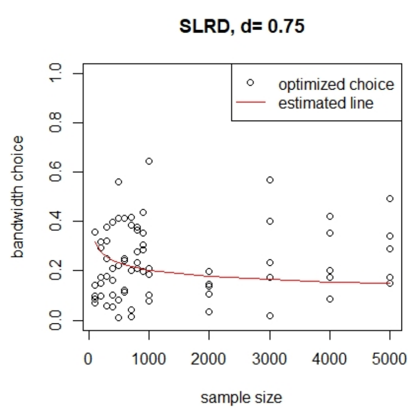

SLRD, d= 0.5
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000


SLRD, d= 0.55
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000


SLRD, d= 0.6
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000


SLRD, d= 0.65
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000


SLRD, d= 0.7
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000


SLRD, d= 0.75
optimized choice
estimated line
bandwidth choice
sample size
0.0 0.2 0.4 0.6 0.8 1.0
0 1000 2000 3000 4000 5000

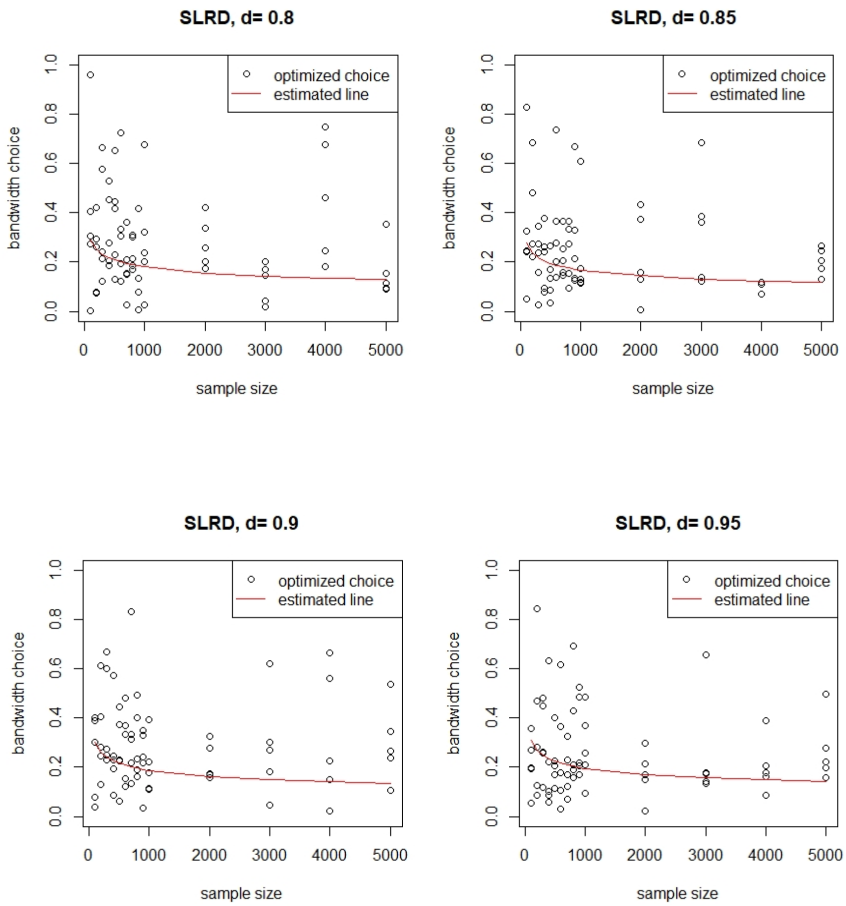


Figure 8: the optimized bandwidth choice for the FLUR with SLRD

Table 3: the optimized bandwidth choice for the FLUR with SLRD

	sample size	optimized bandwidth choice			
	d=	0.1	0.15	0.2	0.25
1	100	0.3245627	0.3311379	0.3107004	0.3169058
2	200	0.2825546	0.2887298	0.2694763	0.2747496
3	300	0.2605492	0.2664871	0.2479463	0.2527397
4	400	0.2459836	0.2517528	0.2337219	0.2382012
5	500	0.2352487	0.2408874	0.2232531	0.2275028
6	600	0.2268264	0.2323588	0.2150485	0.2191192
7	700	0.2199411	0.2253839	0.2083473	0.2122724
8	800	0.214146	0.2195114	0.2027114	0.2065147
9	900	0.2091612	0.2144587	0.197867	0.2015657
10	1000	0.2048005	0.2100375	0.1936316	0.1972394
11	2000	0.1782932	0.1831385	0.1679403	0.1710017
12	3000	0.1644077	0.1690302	0.1545226	0.1573029
13	4000	0.1552167	0.1596844	0.1456578	0.1482543
14	5000	0.148443	0.1527926	0.1391335	0.1415957

	d=	0.3	0.35	0.4	0.45
1	100	0.3341451	0.3176172	0.325934	0.3418206
2	200	0.2926807	0.2757251	0.2843738	0.300477
3	300	0.2708537	0.2538302	0.2625647	0.2786513
4	400	0.2563617	0.2393583	0.248113	0.264134
5	500	0.245657	0.2287037	0.2374534	0.2533965
6	600	0.2372432	0.2203513	0.2290848	0.2449481
7	700	0.2303548	0.2135279	0.2222397	0.2380252
8	800	0.2245496	0.2077881	0.2164758	0.2321867
9	900	0.2195506	0.2028536	0.2115159	0.2271558
10	1000	0.2151732	0.1985388	0.2071755	0.2227479
11	2000	0.1884722	0.1723525	0.1807583	0.1958063
12	3000	0.1744166	0.1586663	0.1668957	0.1815835
13	4000	0.1650845	0.1496201	0.1577096	0.1721233
14	5000	0.1581912	0.14296	0.150934	0.1651262

	d=	0.5	0.55	0.6	0.65	0.7
1	100	0.3014666	0.3163151	0.2781535	0.284452	0.2984069
2	200	0.2619753	0.2771628	0.2391885	0.2453816	0.2591016
3	300	0.2413181	0.256547	0.2189774	0.2250644	0.2385554
4	400	0.2276572	0.2428567	0.2056818	0.2116777	0.2249735
5	500	0.2175958	0.2327428	0.1959275	0.2018449	0.2149734
6	600	0.2097061	0.2247926	0.188302	0.1941511	0.2071335
7	700	0.203259	0.2182831	0.1820867	0.1878752	0.2007286
8	800	0.1978347	0.2127968	0.1768689	0.1826031	0.1953406
9	900	0.1931704	0.2080722	0.1723907	0.1780756	0.1907083
10	1000	0.1890914	0.2039348	0.168481	0.1741209	0.1866577
11	2000	0.1643209	0.1786926	0.1448794	0.1502048	0.1620717
12	3000	0.151364	0.1654011	0.1326373	0.1377681	0.1492197
13	4000	0.1427953	0.1565747	0.124584	0.1295737	0.140724
14	5000	0.1364844	0.1500541	0.1186757	0.1235549	0.1344688

	d=	0.75	0.8	0.85	0.9	0.95
1	100	0.3090833	0.3142239	0.2971913	0.3228173	0.295457
2	200	0.2695766	0.2746637	0.2576227	0.2839412	0.2557421
3	300	0.248851	0.2538734	0.236966	0.2634083	0.2350335
4	400	0.2351197	0.2400841	0.2233224	0.2497469	0.2213656
5	500	0.2249928	0.2299061	0.2132829	0.2396402	0.2113138
6	600	0.2170432	0.2219113	0.2054159	0.2316867	0.2034406
7	700	0.2105414	0.215369	0.1989914	0.2251685	0.1970133
8	800	0.205067	0.209858	0.1935888	0.2196705	0.19161
9	900	0.2003564	0.205114	0.1889453	0.2149325	0.1869671
10	1000	0.1962345	0.2009614	0.184886	0.2107809	0.1829093
11	2000	0.171152	0.1756607	0.1602699	0.1853971	0.1583229
12	3000	0.1579935	0.1623644	0.1474192	0.1719903	0.1455028
13	4000	0.1492756	0.1535454	0.1389313	0.1630702	0.1370414
14	5000	0.1428461	0.1470361	0.1326856	0.1564711	0.1308186

3.3.2 Estimation results

By fitting the optimized bandwidth choice from the fitted local polynomial regression with the function of $\hat{h}_{opt} = a \times N^b$

The estimated $\hat{b}$ and the memory parameter d has the following relationship

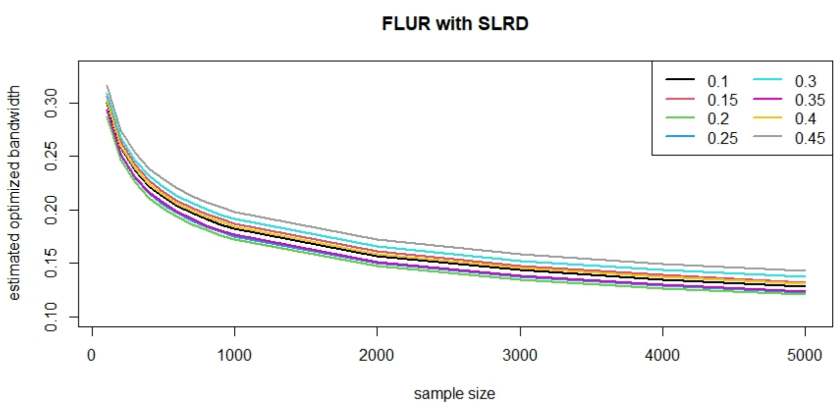
Table 4: Order of bandwidth for SLRD and LRD

d	$\hat{b}$
0.1	-0.1899676
0.15	-0.1977119
0.2	-0.2153656
0.25	-0.2059366
0.3	-0.1911469
0.35	-0.2040586
0.4	-0.1967913
0.45	-0.1859847

d	$\hat{b}$
0.1	-0.16368201
0.15	-0.15476028
0.2	-0.12577166
0.25	-0.12020361
0.3	-0.09704896
0.35	-0.05008519
0.4	-0.04826280
0.45	-0.03352068

Table 5: Order of bandwidth for SLRD

d	$\hat{b}$
0.5	-0.2025674
0.55	-0.1906285
0.6	-0.2177336
0.65	-0.2131580
0.7	-0.2037630
0.75	-0.1973001
0.8	-0.1941265
0.85	-0.2061322
0.9	-0.1851256
0.95	-0.2082584



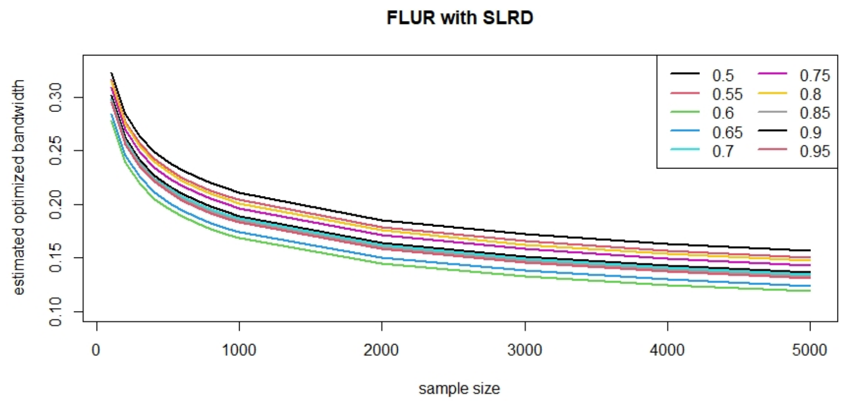


Figure 9: SLRD bandwidth choice

The order of the bandwidth choice do keep close to the {-1/5}, which is basically conform to the original conjecture. And compare to the LRD cases with same memory parameter d, the bandwidth choices keep close to -0.2. This shows the λ_N is the key component to determined the order of optimized bandwidth choice.

4 Conclusions

By simulating the fractional local to unity root (FLUR) model with different memory parameters, we testify the convergence of the short memory process and figure out the relationship between the memory parameter and the order of the bandwidth, which shows how we may use asymptotic theory in parametric and non-parametric regression models in statistics to deal with these stochastic processes. One extension is to verify whether the rate of convergence of the semi-long memory process is related with the tempering parameter. This feature will be extremely important when we want to approximate LRD processes with SLRD ones.

References

[1] FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications.* NewYork: Chapman & Hall/CRC.

[2] KONISHI, S. & KITAGAWA, G. (2008).*Information Criteria and Statistical Modeling.* NewYork: Springer

[3] DIGGLE, P. J. & HUTCHINSON, M. F. (1989). On spline smoothing with correlated errors. *Aust. J. Statist.* 31, 166–82.

[4] CHU, C. K. & MARRON, J. S. (1991). Comparison of two bandwidth selectors with dependent errors.*Ann. Statist.* 4, 1906–18.

[5] OPSOMER, J.,WANG,Y. &YANG,Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.* 16, 134–53.

[6] FRANCISCO-FERNÁNDEZ, M., OPSOMER, J. D. & VILAR-FERNÁNDEZ, J. M. (2004). Plug-in bandwidth selector for local polynomial regression estimator with correlated errors. *J. Nonparam. Statist.* 16, 127–51.

[7] KIM, T.Y., PARK, B. U.,MOON, M. S.&KIM, C. (2009). Using bimodal kernel for inference in nonparametric regression with correlated errors. *J. Mult. Anal.* 100, 1487–97.

[8] DE BRABANTER, K., CAO, F., GIJBELS, I., & OPSOMER, J. (2018). Local polynomial regression with correlated errors in random design and unknown correlation structure. *Biometrika.*, 105(3), 681–690.

[9] DE BRABANTER, K., & SABZIKAR, F. (2021b). Asymptotic theory for regression models with fractional local to unity root errors. *Metrika.* Published. https://doi.org/10.1007/s00184-021-00812-7