L^p-wavelet regression with correlated errors and inverse problems

Rafał Kulik and Marc Raimondo

University of Ottawa and University of Sydney

Abstract: We investigate global performances of non-linear wavelet estimation in regression models with correlated errors. Convergence properties are studied over a wide range of Besov classes $\mathcal{B}^s_{\pi,r}$ and for a variety of L^p error measures. We consider error distributions with Long-Range-Dependence parameter $\alpha, 0 < \alpha \leq 1$. In this setting we present a single adaptive wavelet thresholding estimator which achieves near-optimal properties simultaneously over a class of spaces and error measures. Our method reveals an elbow feature in the rate of convergence at $s = \frac{\alpha}{2}(\frac{p}{\pi} - 1)$ when $p > \frac{2}{\alpha} + \pi$. Using a vaguelette decomposition of fractional Gaussian noise we draw a parallel with certain inverse problems where similar rate results occur.

Key words and phrases: Adaptation, correlated data, deconvolution, degree of ill posedness, fractional Brownian Motion, fractional differentiation, fractional integration, inverse problems, linear processes, long range dependence, L^p loss, non-parametric regression, maxisets, Meyer wavelet, vaguelettes, WaveD.

1 Introduction

There is a vast literature on regression models with correlated errors. Motivations for such studies are both theoretical and practical. For example, modern nonparametric regression techniques are sensitive to the presence of correlated errors and data-driven methods such as cross-validation are affected by long-range-dependence. We refer to Opsomer, Wang and Yang (2001) for an up-to-date review of the topic and extensive references.

Existing methods for dealing with correlated errors in regression models include kernel and wavelet estimation. Oracle kernel estimators are found in Hall and Hart (1990), and Csorgo and Mielniczuk (1995). In these papers optimal rates have been derived using a bandwidth which depends on the smoothness of the unknown function f.

On the other hand, wavelet thresholding methods achieve near optimal properties in an adaptive fashion i.e. in wavelet regression the fine tuning parameters do not depend on the regression function smoothness. Oracle and adaptive wavelet estimation in regression with correlated errors have been derived in Wang (1996), Wang (1997), Johnstone and Silverman (1997), Johnstone (1999) and Li and Xiao (2007). Most of these results have been established with respect to the mean integrated square error (L^2-loss).

Another key feature of wavelet methods is to provide adaptiveness with respect to various L^p-metrics. Among other approaches the maxiset paradigm has been used successfully to study wavelet thresholding algorithms in various settings, see e.g. Kerkyacharian and Picard (2000), Kerkyacharian and Picard (2002), Autin (2006), Rivoirard (2004) and Johnstone, Kerkyacharian, Picard and Raimondo (2004). In the case where the regression errors are independent, optimal rates and adaptive near-optimal L^p-estimation is now well understood. For example when the regression function belongs to a Besov class $\mathcal{B}^s_{\pi,r}$, it is known that there is an elbow in the rates at $s = \frac{1}{2}(\frac{p}{\pi} - 1)$ when $p > 2 + \pi$, see e.g. Kerkyacharian and Picard (2000). In this paper we follow the maxiset approach to study the regression model when the errors are correlated. Our main result shows that there is an elbow in rate results at $s = \frac{\alpha}{2}(\frac{p}{\pi} - 1)$ when $p > \frac{2}{\alpha} + \pi$. For general p it agrees with the literature in the case of independent errors ($\alpha = 1$) and for $p = 2$ with the correlated noise setting of Wang (1996).

The paper is organised as follows. We present our rate results in the continuous and discrete regression model scenarios, and draw a parallel with certain inverse problems where similar elbow phenomenom arise. The mathematical appendix is available online `http://www.stat.sinica.edu.tw/statistica`.

2 Preliminaries

2.1 Non-parametric regression with LRD errors

Let $\{X_i, i \geq 1\}$ be a stationary Gaussian, zero mean and unit variance sequence satisfying

$$\rho_m := \mathrm{E} X_0 X_m \sim L m^{-\alpha}, \tag{2.1}$$

where $f_m \sim g_m$ means $\lim_m f_m/g_m = 1$, $\alpha \in (0, 1]$ and L is a finite and positive constant.

The aim of this paper is to estimate a function f arising in a nonparametric regression model with fixed-design:

$$Y_i = f(u_i) + X_i, \qquad u_i = i/n, \quad i = 1, \dots, n. \tag{2.2}$$

2.2 Asymptotic model

Long-Range-Dependence can also be described in the continuous setting. As in Wang (1996), we consider the Fractional Gaussian Noise (FGN) model,

$$dY_t = f(t)dt + \varepsilon^\alpha dB_H(t), \quad t \in \mathcal{I} = [0,1], \tag{2.3}$$

where $B_H(t)$ is a standard Fractional Brownian Motion (FBM). That is $B_H(t)$ is a zero mean Gaussian process with covariance,

$$r(s,t) = \frac{1}{2}(|s|^H + |t|^H - |t-s|^{2H}), \quad s,t \in \mathcal{I}. \tag{2.4}$$

The parameter $H = 1 - \alpha/2$ belongs to $[1/2, 1)$. The noise level in (2.3) is ε^α with $\varepsilon = n^{-1/2}$.

2.3 Besov classes

The regression function f is defined on the unit interval $\mathcal{I} = [0,1]$. To avoid edge problems and unnecessary technicalities arising in defining wavelet basis on the interval, we will further assume that f is periodic on $\mathcal{I} = [0,1]$ and present our results using the periodised Meyer wavelet basis. We believe that similar rate results may be achieved without the periodic assumption using other wavelet families provided that the wavelet function has enough regularity. Let ϕ, ψ denote the periodised Meyer scaling and wavelet function, see e.g. Mallat (1998), Meyer (1990). In the periodic setting, we recall that Besov spaces are characterised by the behaviour of the wavelet coefficients

Definition 2.1 *For $f \in L^\pi(\mathcal{I})$,*

$$f = \sum_{j,k} \beta_{j,k}\psi_{j,k} \in \mathcal{B}^s_{\pi,r}(\mathcal{I}) \iff \sum_{j \geq 0} 2^{j(s+1/2-1/\pi)r}[\sum_{0 \leq k \leq 2^j} |\beta_{j,k}|^\pi]^{r/\pi} < \infty. \tag{2.5}$$

As usual $\psi_{j,k}(x) = 2^{j/2}\psi(2^j - k)$ denotes the dilated and translated version of ψ, and $\beta_{j,k} = \int_0^1 f(u)\psi_{j,k}(u)du$ is the associated wavelet coefficient. The parameter

s can be thought of as related to the number of derivatives of f. With different values of π and r, the Besov spaces capture a variety of smoothness features in a function including spatially inhomogeneous behaviour.

2.4 Non-linear wavelet estimation

Our estimator is based on hard thresholding of a wavelet expansion as follows (here and in the sequel κ will denote the multiple index (j,k) and $\psi_{-1} = \phi$),

$$\hat{f}_n = \sum_{\kappa \in \Lambda} \hat{\beta}_\kappa \, \psi_\kappa \, \mathbb{I}\{|\hat{\beta}_\kappa| \geq \lambda\} \tag{2.6}$$

where the threshold parameter λ and index range Λ will be specified later, and $\hat{\beta}_\kappa$ is an estimator of the wavelet coefficient β_κ. In the discrete model (2.2), we set

$$\hat{\beta}^D_\kappa := \frac{1}{n} \sum_{i=1}^n \psi_\kappa(u_i) Y_i, \tag{2.7}$$

in the continuous model (2.3), we set

$$\hat{\beta}^C_\kappa := \int \psi_\kappa(t) dY_t. \tag{2.8}$$

2.5 Sequence FGN model

Applying the Meyer Wavelet transform to the data (2.3),

$$\int \psi_\kappa(t) dY_t = \int f(t) \psi_\kappa(t) dt + \varepsilon^\alpha \int \psi_\kappa(t) dB_H(t)$$

which we write as

$$\hat{\beta}^C_\kappa = \beta_\kappa + \varepsilon^\alpha \sigma_j z_\kappa, \tag{2.9}$$

where, as in Wang (1996), $\sigma_j^2 = Var(\int \psi_\kappa(t) dB_H(t))$ and z_κ are (weakly) correlated Gaussian random variables with variance 1 and

$$\sigma_j = \tau 2^{-j(1-\alpha)/2},$$

where τ is a scaling parameter which depends on ψ and α,

$$\tau^2 = \tau^2_{\mathrm{A}} = (1 - \frac{\alpha}{2})(1-\alpha) \int_0^1 \int_0^1 \psi(u)\psi(v)|u-v|^{-\alpha} \, \mathrm{d}u \, \mathrm{d}v. \tag{2.10}$$

A similar model was used in Johnstone (1999).

3 Wavelet regression with correlated errors

3.1 A maxiset approach

The connection between regression with LRD-errors and certain inverse problems has been made in Johnstone (1999) and is further discussed in section 3.4. The sequence space representation (2.9) illustrates both the similarities: such as a level-dependent variance and the differences: such as a LRD index dependent noise level. Here we tune the non-linear wavelet approximation (2.6) for regression with LRD-errors in a fashion similar to that of the WaveD method of Johnstone, Kerkyacharian, Picard and Raimondo (2004). In particular we follow a maxiset approach and use a level-dependent threshold together with a fine resolution level which depends on the LRD index α.

Fine resolution level. The range of resolution levels (frequencies) where the approximation (2.6) is used:

$$\Lambda_n = \{(j,k),\ -1 \leq j \leq j_1, 0 \leq k \leq 2^j\}, \tag{3.1}$$

here j_1 is the finest resolution level which we set to be

$$2^{j_1} = \Big(\frac{n}{\log n}\Big)^{\alpha}. \tag{3.2}$$

Threshold. The threshold value $\lambda = \lambda_j$ has three input parameters:

$$\lambda_j = \eta\, \sigma_j\, c_n \tag{3.3}$$

- η: $\eta > \sqrt{8\alpha}\sqrt{2 \vee p}$.
- σ_j: a level-dependent scaling factor, as in section 2.5 we set

$$\sigma_j = \tau\, 2^{-j(1-\alpha)/2}, \tag{3.4}$$

where $\tau^2 = \tau_{\mathrm{A}}^2$, see (2.10), in the asymptotic model and

$$\tau^2 = \tau_{\mathrm{D}}^2 = L \int_0^1\!\!\int_0^1 \psi(u)\psi(v)|u-v|^{-\alpha}\, \mathrm{d}u\, \mathrm{d}v, \tag{3.5}$$

in the discrete model, L depends on the error distribution, see (2.1).

- c_n: a sample size-dependent scaling factor:

$$c_n = (\log n)^{\frac{1}{2}}\, n^{-\frac{\alpha}{2}}. \tag{3.6}$$

3.2 Rate results in the asymptotic model

First we present our rate results in the asymptotic model (2.3).

Theorem 3.1 *Consider (2.3) with $\varepsilon = n^{-1/2}$, and wavelet estimator (2.6) with (2.10), (3.2), (3.3), (3.4), and (3.6). Assume that $p > 1$, $f \in \mathcal{B}^s_{\pi,r}$ with $s \geq \frac{1}{\pi}$, then there exists a constant $C > 0$ such that for all $n \geq 0$,*

$$\mathrm{E}\left\| f - \hat{f}_n \right\|_p^p \leq C\Big(\frac{(\log n)^{\frac{1}{\alpha}}}{n}\Big)^{\gamma},$$

with

$$\gamma = \frac{\alpha s p}{2(s + \frac{\alpha}{2})}, \qquad \text{if } s \geq \frac{\alpha}{2}(\frac{p}{\pi} - 1), \tag{3.7}$$

$$s - \left(\frac{1}{\pi} - \frac{1}{p}\right)_+ > \frac{s}{2s+\alpha}, \tag{3.8}$$

and

$$\gamma = \frac{\alpha p (s - \frac{1}{\pi} + \frac{1}{p})}{2(s - \frac{1}{\pi} + \frac{\alpha}{2})}, \qquad \text{if } \frac{1}{\pi} < s < \frac{\alpha}{2}(\frac{p}{\pi} - 1). \tag{3.9}$$

Remark 3.1 The two rate regimes (3.7) and (3.9) are usually referred as the 'dense' and 'sparse' phase, see e.g. Kerkyacharian and Picard (2000) in the iid case or Kerkyacharian, Picard and Tribouley (1997) in the density case. Our results show that the boundary region $s = \frac{\alpha}{2}(\frac{p}{\pi} - 1)$ depends on the LRD index α, and the sparse region is smaller for dependent data.

Remark 3.2 For $\alpha = 1$ our rate results agree with results obtained in the regression model with independent errors (cf. Theorem 6.1 in Kerkyacharian and Picard (2000)). For $\alpha < 1$ our rate results in the sparse phase seems to be new in the regression setting. From (3.9) we see that the condition $p > \frac{2}{\alpha} + \pi$ is required for the sparse regime to be visible.

Remark 3.3 Our estimator is adaptive with respect to the smoothness class as our tuning paradigm does not depend on s. At present the method is not adaptive with respect to the LRD parameter α as both the fine resolution parameter (3.2) and threshold scaling value (3.4) depend on α. We believe that it is possible to develop a tuning paradigm which does not involve α, using a fourier domain stoping rule and random thresholds as in Cavalier and Raimondo (2007).

3.3 Rate results in the discrete model

We now state our results in the discrete model (2.2).

Theorem 3.2 *Consider (2.2), and wavelet estimator (2.6) with (3.2), (3.3), (3.4), (3.6) and scaling factor*

$$\tau^2 = \tau_{\mathrm{D}}^2 = L \int_0^1\!\!\int_0^1 \psi(u)\psi(v)|u-v|^{-\alpha}\,\mathrm{d}u\,\mathrm{d}v,$$

then rate results of theorem 3.1 hold.

Remark 3.4 The proofs of theorems 3.1 and 3.2 are based on the maxiset theorem from Kerkyacharian and Picard (2000) with details given in the online Appendix. The technical novelties appear in moment bounds and large deviation results for wavelet coefficients which we establish under LRD. The maxiset approach allows model-specific choice of the thresholding constant. This clarifies the effect of LRD on the threshold. About (3.5) and (2.10), note that the FGN model is obtained by taking the limit of *standardized* partial sums

$$\frac{1}{\sqrt{\mathrm{Var}(\sum_{i=1}^n X_i)}} \sum_{i=1}^n X_i$$

and then thresholding is applied. On the other hand, in the discrete model, thresholding is applied to original data directly. Of course the constants τ_{A} and τ_{D} agree in the special case where X_i is defined as increments of $B_H(\cdot)$, i.e. $X_i = B_H(i) - B_H(i-1)$. Alternative approaches show that if one starts from the discrete model the resulting noise variables in the continuous model (2.9) are asymptotically weakly dependent, either via fast decaying covariances (Wang (1996)), or mixing properties (Johnstone (1999)). In this fashion Wang (1996) has established the asymptotic equivalence of the discrete model (2.2) with the continuous model (2.3) for the quadratic minimax risk. Whether such equivalence extends to general L^p error measures, as indicated by our rate results, is an issue for future work.

Remark 3.5 The algorithm of Donoho and Raimondo (2004) (see `WaveD` software package, Raimondo and Stewart (2007)) may be used to study the numerical performances of the estimator (2.6). This is under investigation by the authors.

3.4 Connection with certain inverse problems in white noise

We give some heuristic arguments which suggest that the rate results (3.7) and (3.9) are near optimal. As in Wang (1996), Johnstone (1999), Cavalier (2004) we make a connection between regression in LRD noise and certain inverse problems. We use the wavelet decomposition of FGN by Meyer, Sellan and Taqqu (1999), referring to this paper for details. In an attempt to stay as close as possible to their original construction we shall, from now on, consider functions $f \in L^2(\mathbb{R})$ and denote (ϕ, ψ) a Meyer wavelet basis of $L^2(\mathbb{R})$. We conjecture that most arguments below extend to the periodic setting, as in Johnstone, Kerkyacharian, Picard and Raimondo (2004). For any $d \in \mathbb{R}$, let $D^d_{.}$ denotes the operator, which in Fourier domain is defined by

$$\widehat{D}^d_f(\omega) := (i\,\omega)^d \,\widehat{f}(\omega). \tag{3.10}$$

For $d > 0$ this corresponds to fractional differentiation and for $d < 0$ to fractional integration. We set $\psi^{(d)} := D^d_\psi$ i.e.

$$\widehat{\psi}^{(d)}(\omega) = (i\,\omega)^d \,\widehat{\psi}(\omega), \tag{3.11}$$

in the Fourier domain. In the time domain, for any $\kappa = (j, k)$, we set

$$\psi^{(d)}_\kappa(x) = \psi^{(d)}_{j,k}(x) := 2^{j/2}\psi^{(d)}(2^j x - k). \tag{3.12}$$

We recall that the $(\psi^{(d)}_\kappa)$'s are biorthogonal i.e.

$$\int \psi^{(d)}_\kappa(x)\psi^{(-d)}_{\kappa'}(x)dx = 1 \text{ only if } \kappa = \kappa' \text{ and } \ 0 \text{ otherwise.} \tag{3.13}$$

With a similar definition for $\phi^{(d)}$. Let $d = H - 1/2 = (1-\alpha)/2$,

$$dB_H(t) = \sum_k z_k\, \phi^{(-d)}(t-k) + \sum_j \sum_k 2^{-jd}\, z_{j,k}\, \psi^{(-d)}_{j,k}(t), \tag{3.14}$$

where $z_{j,k}$ are iid standard normal and z_k is a discrete-time fractional ARIMA. Introducing vaguelettes:

$$\mathcal{U}_\kappa(t) = \mathcal{U}_{j,k}(t) := 2^{jd}\, \psi^{(d)}_{j,k}(t), \tag{3.15}$$

which *standardise* noise contributions in sequence domain, using (3.13) and (3.14),

$$\langle dB_H, \mathcal{U}_\kappa\rangle = 2^{jd}\, 2^{-jd}\langle \psi^{(d)}_\kappa, \psi^{(-d)}_\kappa\ \rangle\, z_\kappa = z_\kappa \sim \mathcal{N}(0,1), \tag{3.16}$$

moreover the $z_\kappa = z_{j,k}$'s are independent. Applying this in (2.3),

$$y_\kappa := \langle dY, \mathcal{U}_\kappa \rangle = \theta_\kappa + \varepsilon^\alpha z_\kappa, \tag{3.17}$$

where by Plancherel, Hermitien symmetry, (3.11) and (3.15)

$$\theta_\kappa = \int f(t)\mathcal{U}_\kappa(t)dt = \int \widehat{f}(\omega)[\widehat{\mathcal{U}}_\kappa(\omega)]^* d\omega = \int \widehat{f}(\omega)(-i\omega)^d \, \widehat{\psi}^*_\kappa(\omega)d\omega.$$

Hence θ_κ can be interpreted as $\theta_\kappa = \int K_f(t)\psi_\kappa(t)dt$ where K_f is the operator defined in Fourier domain by $\widehat{K^d_f}(\omega) = (-i\,\omega)^d \, \widehat{f}(\omega)$. This shows that, in wavelet sequence space (3.17), the model (2.3) is equivalent to

$$dY_t = K_f(t)dt + \xi dB(t), \; t \in [0,1], \tag{3.18}$$

where K_f is a fractional differentiation operator, $B(t)$ is a Gaussian white noise and the noise level $\xi = \varepsilon^\alpha$. Introducing

$$\mathcal{V}_\kappa := 2^{-jd} \; \psi^{(-d)}_{j,k}(t), \tag{3.19}$$

using (3.13), (3.15) and (3.19) we see that the system $\mathcal{U}_\kappa, \mathcal{V}_\kappa$ is biorthogonal. Further, each of the systems $\mathcal{U}_\kappa$, $\mathcal{V}_\kappa$ forms a Riesz basis, see Meyer, Sellan and Taqqu (1999). Thus, the system $\mathcal{U}_\kappa, \mathcal{V}_\kappa$ yields the wavelet vaguelette decomposition (WVD) of K_f, see Donoho (1995). For such operators or certain smooth convolutions, where the kernel $k(t)$ satisfies $|\widehat{k}(\omega)| \sim |\omega|^d$, it is customary to define the Degree of Ill-posedness (DIP) as $\nu = -d$ so that $|\widehat{k}(\omega)| \sim |\omega|^{-\nu}$ agrees with the standard WVD representation where the optimal rate over Besov balls, is $r(\xi) = \xi^{2\gamma}$,

$$\gamma = \frac{sp}{1 + 2(s+\nu)}, \quad \text{if } s \geq (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}) \tag{3.20}$$

and

$$\gamma = \frac{(s - 1/\pi + 1/p)p}{1 + 2(s + \nu - 1/\pi)}, \quad \text{if } \frac{1}{\pi} < s < (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}). \tag{3.21}$$

See Donoho (1995), Johnstone, Kerkyacharian, Picard and Raimondo (2004), Cavalier and Raimondo (2007), Hoffmann and Reiss (2008).

Remark 3.6 The parallel between (2.3) and (3.18) holds with $d = (1-\alpha)/2$. For $\alpha \in (0,1]$, the DIP $\nu = -d = (\alpha-1)/2$ is in the negative range $-\frac{1}{2} < \nu \leq 0$, whereas, typically, inverse problems are considered for $\nu > 0$ ($\nu = 0$ representing the direct case). We note, for example, that rate results (3.20) and (3.21) of Johnstone, Kerkyacharian, Picard and Raimondo (2004) can be extented to cover $\nu > -\frac{1}{2}$ since, in this case, inverting the operator reduces the noise variance. As a result, for $-1/2 < \nu \leq 0$ the rates (3.20) and (3.21) are faster than in the direct case ($\nu = 0$). We shall refer to the region where $-1/2 < \nu \leq 0$ as the *fast zone* and $\nu > 0$ as the *slow zone*. A similar rate phenomenom, with negative DIP, arises in certain Berkson errors-in-variables models, see e.g. Delaigle (2007).

Remark 3.7 Donoho (1995) has established optimality properties of WVD for inverse problems (3.18) where K_f is a fractional integration operator. In this scenario, the WVD "inverses" K_f by applying a differentiation operator which, in Fourier domain representation (3.10) corresponds to $d > 0$, and leads to vaguelette (3.15). Unlike the case of FGN (2.3), applying the vaguelette transform (3.15) to (3.18) inflates the noise by a multiplicative scale of order $2^{j\nu}$ where $\nu = d > 0$ is, for obvious reason, called the Degree of Ill Posedness. Our approach is similar, albeit twofold and reverse. First, we apply vaguelette transform (3.15) to FGN (2.3) which leads to inverse model (3.18) where K_f is a differentiation operator. In a second step, we apply the WVD paradigm to the differential operator K_f using biorthogonal vaguelette transform (3.19) in (3.18); this affects the noise scale by a multiplicative factor of order $2^{j\nu}$, but this time $\nu = -d < 0$ corresponds to a negative DIP which reduces the noise variance.

Epilogue. Recall that the parallel between (2.3) and (3.18) holds with $\nu = -d = (\alpha - 1)/2$, so that we are always in the *fast zone* $-1/2 < \nu \leq 0$. At first this may seem surprising since for $\alpha > 0$ the rates, at noise level ξ, are faster than in the independent case $\alpha = 1$. However, in the parallel between (2.3) and (3.18) one has to adjust the noise level to $\xi = \varepsilon^\alpha$, so that the detrimental effect of LRD is concentrated on the noise level ε^α which tends to zero at a much slower rate when $\alpha \in (0, 1)$ than in the independent errors scenario ($\alpha = 1$). Finally, we note that with $\nu = (\alpha - 1)/2$ and $\xi = \varepsilon^\alpha = n^{-\alpha/2}$, the rates (3.20) and (3.21) agree, up to log terms, with our rate results (3.7) and (3.9).

References

Autin, F. (2006). Maxisets for density estimation on R, *Math. Methods of Statist.*, **15**(2), 123–145.

Cavalier, L. (2004). Estimation in a problem of fractional integration. *Inverse Problems*, **20**, 1445-1454.

Cavalier, L. and Raimondo, M. (2007). Wavelet deconvolution with noisy eigenvalues, *IEEE Trans. Signal Proc.* **55**(6), 2414–2424.

Csorgo, S. and Mielniczuk, J. (1995). Nonparametric Regression Under Long-Range Dependent Normal Errors. *Ann. Statist.* **23**, 1000–1014.

Delaigle, A. (2007). Nonparametric density estimation from data with a mixture of Berkson and classical errors. *Canadian Journal of Statistics*, **35**, 89–104.

Donoho, D.L. (1995). Nonlinear Solution of Linear Inverse Problems by Wavelet-Vaguelette Decomposition. *Applied and Computational Harmonic Analysis* **2**, 101–126.

Donoho, D. and Raimondo, M. (2004). Translation invariant deconvolution in a periodic setting, *The International Journal of Wavelets, Multiresolution and Information Processing* **14**(1), 415–423.

Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. *Stoch. Proc. Appl.* **36**, 339–351.

Hoffmann, M. and Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator, *Ann. Statist.* **36**, 310-336.

Johnstone, I.M. (1999). Wavelet threshold estimators for correlated data and inverse problems: Adaptivity results. *Statistica Sinica* **9**, 51–83.

Johnstone, I.M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B.* **59**, 319–351.

Johnstone, I.M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004). Wavelet Deconvolution in a periodic setting. *J. Roy. Statist. Soc. Ser. B.* **66**, 547-573.

Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well-concentrated basis. *Test* **9**, 283–344.

Kerkyacharian, G. and Picard, D. (2002). Minimax or maxisets?, *Bernoulli* **8** (2), 219–253.

Kerkyacharian, G., Picard, D. and Tribouley, K. (1997). L^p adaptive density estimation. *Bernoulli* **2**, 229–247.

Li, L. and Xiao, Y. (2007). On the minimax optimality of block thresholded wavelet estimators with long memory data. *Journal of Statistical Planning and Inference* **137**, 2850–2869.

Mallat, S. (1998) , *A wavelet tour of signal processing (2nd Edition)*, Academic Press Inc., San Diego, CA.

Meyer, Y., Sellan, F. and Taqqu, M.S. (1999). Wavelets, generalized white noise and fractional integration: The synthesis of fractional Brownian motion, *Journal of Fourier Analysis and Applications*, **5**, 465–494.

Meyer, Y. (1990) , *Ondelettes et Opérateurs-I*, Hermann.

Opsomer, J., Wang, Y. and Yang, Y. (2001). Nonparametric Regression with Correlated Errors. *Statistical Science* **16** (2), 134–153.

Raimondo, M. and Stewart, M.(2007). The waved transform in R. *Journal of Statistical Software*, **21**, 1–27.

Rivoirard, V. (2004). Maxisets for linear procedures, *Statist. Probab. Lett.* **67**(3), 267–275.

Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24**, 466–484.

Wang, Y. (1997). Minimax estimation via wavelets for indirect long-memory data. *Journal of Statistical Planning and Inference* **64**, 45–55.

Rafal Kulik
Dept. of Mathematics and Statistics
University of Ottawa
585 King Edward Av.
Ottawa ON K1N 6N5, Canada
E-mail: (rkulik@uottawa.ca)

Marc Raimondo
School of Mathematics and Statistics
University of Sydney
NSW 2006
Australia
E-mail: (marcr@maths.usyd.edu.au)

4 Online appendix: proof of theorems 3.1 and 3.2

The proofs of theorems 3.1 and 3.2 are based on the maxiset theorem from Kerkyacharian and Picard (2000). The steps are similar to those of Johnstone, Kerkyacharian, Picard and Raimondo (2004). The technical novelties appear in moment bounds and large deviation results for wavelet coefficients (2.7), (2.8) which we establish under LRD assumption.

4.1 Maxiset Theorem

The following theorem is borrowed from Kerkyacharian and Picard (2000). We refer to section 4.4 for the definition of Temlyakov property. First, we introduce some notation: μ will denote the measure such that for $j \in \mathbb{N},\ k \in \mathbb{N}$,

$$\mu\{(j,k)\} = \|\sigma_j \psi_{j,k}\|_p^p = \sigma_j^p 2^{j(\frac{p}{2}-1)} \|\psi\|_p^p \tag{6.1}$$

$$l_{q,\infty}(\mu) = \left\{ f,\ \sup_{\lambda>0} \lambda^q \mu\{(j,k):\ |\beta_{j,k}| > \sigma_j \lambda\} < \infty \right\} \tag{6.2}$$

Theorem 4.1 *Let $p > 1,\ 0 < q < p,\ \{\ \psi_{j,k}, j \geq -1, k = 0,1,...,2^j\}$ be a periodised wavelet basis of $L^2(\mathcal{I})$ and σ_j be a positive sequence such that the heteroscedastic basis $\sigma_j \psi_{j,k}$ satisfies Temlyakov property. Suppose that Λ_n is a set of pairs (j,k) and c_n is a deterministic sequence tending to zero with*

$$\sup_n \mu\{\Lambda_n\}\, c_n^p < \infty. \tag{6.3}$$

If for any n and any pair $\kappa = (j,k) \in \Lambda_n$, we have

$$\mathrm{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \quad \leq \quad C\,(\sigma_j\, c_n)^{2p} \tag{6.4}$$

$$P\Big(|\hat{\beta}_\kappa - \beta_\kappa| \geq \eta\, \sigma_j\, c_n/2\Big) \quad \leq \quad C\,(c_n^{2p} \wedge c_n^4) \tag{6.5}$$

for some positive constants η and C then, the wavelet based estimator

$$\hat{f}_n = \sum_{\kappa \in \Lambda_n} \hat{\beta}_\kappa\, \psi_\kappa\, \mathbb{I}\{|\hat{\beta}_\kappa| \geq\ \eta\, \sigma_j\, c_n\} \tag{6.6}$$

is such that, for all positive integers n,

$$\mathrm{E}\|\hat{f}_n - f\|_p^p \leq C\, c_n^{p-q},$$

if and only if :

$$
\begin{aligned}
f \quad &\in \quad l_{q,\infty}(\mu), \quad \textit{and}, &(6.7)\\
\sup_n c_n^{q-p} \quad &\| \quad f - \sum_{\kappa\in\Lambda_n} \beta_\kappa\psi_\kappa\|_p^p < \infty. &(6.8)
\end{aligned}
$$

This theorem identifies the 'Maxiset' of a general wavelet estimator of the form (6.6). This is done by using conditions (6.7) and (6.8) for an appropriate choice of q. In the proof of the theorems we will choose q according to the dense or sparse regime by setting:

$$q = q_d := \frac{\alpha p}{2s+\alpha}, \quad \text{when} \quad s \geq \frac{\alpha}{2}\Big(\frac{p}{\pi} - 1\Big) \tag{6.9}$$

$$q = q_s := \frac{\frac{\alpha p}{2} - 1}{s - \frac{1}{\pi} + \frac{\alpha}{2}}, \qquad \text{when} \quad s < \frac{\alpha}{2}\Big(\frac{p}{\pi} - 1\Big). \tag{6.10}$$

4.2 Moment bounds and large deviation estimates in fBM model

Here $\hat\beta_\kappa = \hat\beta_\kappa^C$, see (2.8). In what follows C denotes a generic constant which does not depends on n but may change from line to line. From (2.9) with $\varepsilon = n^{-1/2}$, (3.4) and (3.6) it follows that $\mathrm{E}\hat\beta_\kappa = \beta_\kappa$ and

$$\mathrm{Var}\hat\beta_\kappa = \mathrm{Var}(\varepsilon^\alpha \int \psi_\kappa(t)dB_H(t)) = n^{-\alpha}2^{-j(1-\alpha)}\tau^2 \leq C\sigma_j^2\, c_n^2.$$

Since the rv's $\hat\beta_\kappa - \beta_\kappa$ are Gaussian higher moments bound (6.4) follows from the previous inequality. Similarly,

$$\Pr\Big(|\hat\beta_\kappa - \beta_\kappa| > \eta\sigma_j\, c_n/2\Big) \leq \exp\left(-\log n\frac{\eta^2}{8}\right) \leq C\,(c_n^{2p} \wedge c_n^4) \tag{6.11}$$

provided $\eta > \sqrt{8\alpha}\sqrt{p\vee 2}$. Which proves (6.5).

4.3 Moment bounds and large deviation estimates in the discrete model

Here $\hat\beta_\kappa = \hat\beta_\kappa^D$, see (2.7). Write

$$
\begin{aligned}
\hat\beta_\kappa - \beta_\kappa &= \hat\beta_\kappa - \mathrm{E}\hat\beta_\kappa + \mathrm{E}\hat\beta_\kappa - \beta_\kappa\\
&= \frac{1}{n}\sum_{i=1}^n X_i\psi_\kappa(u_i) + \left(\frac{1}{n}\sum_{i=1}^n f(u_i)\psi_\kappa(u_i) - \beta_\kappa\right).
\end{aligned}
$$

The main tool to derive rates of convergence is the following lemma. To establish moments bounds we do not assume that X_i's are Gaussian. These estimates may be of independent interest.

Lemma 4.2 *For each fixed j and k, and $p > 1$,*

$$\mathrm{E}(\hat{\beta}_\kappa - \beta_\kappa)^2 \sim 2^{-j(1-\alpha)} n^{-\alpha} \tau_D^2, \tag{6.12}$$

$$\mathrm{E}\left|\hat{\beta}_\kappa - \beta_\kappa\right|^p = O\left(n^{-\alpha p/2} 2^{-jp(1-\alpha)/2}\right). \tag{6.13}$$

If moreover X_i's are Gaussian, then for all $\lambda > n^{-1}$,

$$\Pr\left(|\hat{\beta}_\kappa - \beta_\kappa| > \lambda\right) \le \frac{n^{-\alpha/2} 2^{-j(1-\alpha)/2}}{\lambda} \exp\left(-\frac{\lambda^2}{2(n^{-\alpha} 2^{-j(1-\alpha)} \tau_D^2)}\right). \tag{6.14}$$

To prove this lemma we will replace β_κ with $\hat{\beta}_\kappa$ and use $|\mathrm{E}\hat{\beta}_\kappa - \beta_\kappa| = O(n^{-1})$. (Note that this just the distance between the integral $\int f(x)\psi_\kappa(x)\,\mathrm{d}x$ and the Riemann-Stjeltjes sum.

Proof:

Note that

$$\sum_{i=1}^n \psi_\kappa^2(u_i) = 2^j \sum_{i=1}^n \psi^2\left(2^j \frac{i}{n} - k\right) = 2^j n \int_0^1 \psi^2(2^j x) dx + o(n) = n + o(n). \tag{6.15}$$

Bearing in mind that $\mathrm{Var}(X_i) = \mathrm{E}(X_1^2) = 1$ we have:

$$\begin{aligned}
\mathrm{E}(\hat{\beta}_\kappa - \mathrm{E}\hat{\beta}_\kappa)^2 &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i \psi_\kappa(u_i)\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^n \psi_\kappa^2(u_i) + \sum_{i \ne l} \psi_\kappa(u_i)\psi_\kappa(z_l)\mathrm{Cov}(X_i, X_l)\right).
\end{aligned}$$

By (6.15) above, the first part is of order $n^{-1} + o(n^{-1})$. For the second part we have

$$\begin{aligned}
&\sum_{i \ne l} \psi_\kappa(u_i)\psi_\kappa(z_l)\mathrm{Cov}(X_i, X_l) \\
&= \sum_{i \ne l} 2^j |i - l|^{-\alpha} \psi\left(2^j \frac{i}{n} - k\right) \psi\left(2^j \frac{l}{n} - k\right) \\
&= L 2^j n^{-\alpha} \sum_{i \ne l} \left|\frac{i}{n} - \frac{l}{n}\right|^{-\alpha} \psi\left(2^j \frac{i}{n} - k\right) \psi\left(2^j \frac{l}{n} - k\right),
\end{aligned}$$

which behaves asymptotically as $2^{-j(1-\alpha)}n^{2-\alpha}\tau_D^2$.

Further, the first part dominates the second one if and only if $2^j > n$, which is not possible. Thus (6.12) follows.

To prove (6.13), let

$$b_r = \sum_{i=r}^{n} a_{i-r}\psi_\kappa(u_i), \qquad r = 1, \ldots, n,$$

$$b_r = \sum_{i=1}^{n} a_{i-r}\psi_\kappa(u_i), \qquad r = -\infty, \ldots, 0.$$

Also, note that by (6.12),

$$v_n^2 := \mathrm{Var}\left(\sum_{r=-\infty}^{n} \epsilon_r b_r\right) = \sum_{r=-\infty}^{n} b_r^2 = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\psi_\kappa(u_i)\right)$$

and thus

$$v_n^2/(n^{2-\alpha}2^{-j(1-\alpha)}\tau_D^2) \to 1 \tag{6.16}$$

as $n \to \infty$.

Note now that each Gaussian sequence (2.1) can be represented as

$$X_i = \sum_{m=0}^{\infty} a_m \epsilon_{i-m}, \quad i \geq 1, \tag{6.17}$$

where a_m is a regularly varying sequence with index $-(\alpha+1)/2$ and $\{\epsilon_i, i \geq 1\}$ is a centered sequence of i.i.d. random variables. Via Rosenthal inequality, for $p \geq 2$

$$\begin{aligned}
\mathrm{E}\left|\hat\beta_\kappa - \mathrm{E}\hat\beta_\kappa\right|^p &= \mathrm{E}\left|\frac{1}{n}\sum_{i=1}^{n} X_i\psi_\kappa(u_i)\right|^p \\
&= n^{-p}\mathrm{E}\left|\sum_{m=0}^{\infty} a_m \sum_{i=1}^{n} \epsilon_{i-m}\psi_\kappa(u_i)\right|^p = n^{-p}\mathrm{E}\left|\sum_{r=-\infty}^{n} \epsilon_r b_r\right|^p \\
&\leq n^{-p}\left(\sum_{r=-\infty}^{n} b_r^2\right)^{p/2} + n^{-p}\sum_{r=-\infty}^{n} |b_r|^p \\
&\leq n^{-p}\left(\sum_{r=-\infty}^{n} b_r^2\right)^{p/2} + n^{-p}\sup_r |b_r|^{p-2}\sum_{r=-\infty}^{n} b_r^2 \\
&= n^{-p}O\left(\left(n^{2-\alpha}2^{-j(1-\alpha)}\right)^{p/2}\right) + n^{-p}n^{p/2-1}O\left(n^{2-\alpha}2^{-j(1-\alpha)}\right) \\
&= O\left(n^{-\alpha p/2}2^{-jp/2(1-\alpha)} + n^{1-\alpha-p/2}2^{-j(1-\alpha)}\right).
\end{aligned}$$

The second term is negligible for all j such that $2^j \le n$.

To prove (6.14) note that $\sum_{i=1}^n X_i \psi_\kappa(u_i) \sim \mathcal{N}(0, v_n^2)$. Thus,

$$\Pr\left(|\hat\beta_\kappa - \mathrm{E}\hat\beta_\kappa| > \lambda\right) \le C \frac{v_n}{n\lambda} \exp\left(-\frac{n^2\lambda^2}{2v_n^2}\right).$$

and the result follows by (6.16). $\odot$

Consequently,

$$\mathrm{E}\left|\hat\beta_\kappa - \beta_\kappa\right|^p = O\left(n^{-\alpha p/2} 2^{-jp/2(1-\alpha)}\right) = O(\sigma_j^p c^p(n)) \tag{6.18}$$

and taking $\lambda = \lambda_j = \eta \sigma_j c_n$,

$$\Pr\left(|\hat\beta_\kappa - \beta_\kappa| > \eta\sigma_j c_n/2\right) \le \exp\left(-\log n \frac{\eta^2}{8}\right) = O(c_n^{2p}) \tag{6.19}$$

provided $\eta > \sqrt{8p\alpha}$. The similar argument applies to $1 < p < 2$. In this case we require $\eta > \sqrt{16\alpha}$.

4.4 Temlyakov property

As seen in Johnstone, Kerkyacharian, Picard and Raimondo (2004, appendix A), the basis $(\sigma_j \psi_{j,k}(.))$ satisfies Temlyakov property as soon as

$$\sum_{\Lambda_n} 2^j\, \sigma_j^2 \le C \sup_{\Lambda_n} \left(2^j \sigma_j^2\right),$$

and

$$\sum_{\Lambda_n} 2^{jp/2}\, \sigma_j^p \le C \sup_{\Lambda_n} \left(2^{jp/2} \sigma_j^p\right), \quad 1 \le p < 2,$$

which is clearly satisfied when $\sigma_j^2 = \tau^2\, 2^{-j(1-\alpha)}$, as prescribed in (3.4).

4.5 Fine resolution tuning

Here we check that condition (6.3) is satisfied. Using (3.4) and (6.1),

$$\mu(\Lambda_n) = \sum_{j \le j_1} \sum_{k=0}^{2^j-1} \mu(j,k) = \sum_{j \le j_1} 2^j \mu(j,k) = \tau^p \sum_{j \le j_1} 2^j\, 2^{j(\frac{p}{2} - 1 - \frac{p(1-\alpha)}{2})} = O(2^{\frac{j_1}{2}})$$

from (3.2) and (3.6), with $p > 1$,

$$\mu(\Lambda_n)c_n^p = \Big(\frac{n}{\log n}\Big)^{\frac{\alpha}{2}} \Big(\frac{(\log n)^{\frac{1}{2}}}{n^{\frac{\alpha p}{2}}}\Big) = O\Big(c_n^{p-1}\big(\frac{\log n}{(\log n)^{\alpha}}\big)^{\frac{1}{2}}\Big) = o(1),$$

which shows that condition (6.3) is satisfied.

4.6 Besov embedding and Maxiset condition (Part I)

For both the dense (6.9) and sparse (6.10) regime, we look for a Besov scale δ such that

$$\mathcal{B}^{\delta}_{\pi,r} \subseteq l_{q,\infty}. \tag{6.20}$$

As usual we note that it is easier to work with

$$l_q(\mu) = \left\{ f \in L_p : f = \sum_{j,k \in A_j} \frac{|\beta_{jk}|^q}{\sigma_j^q} \, \|\sigma_j \psi_{j,k}\|_p^p < \infty \right\},$$

where A_j is a set of cardinality proportional to 2^j. Using (3.4) and $\|\sigma_j\psi_{j,k}\|_p^p = \sigma_j^p \, 2^{j(\frac{p}{2}-1)} = 2^{j(\frac{\alpha p}{2}-1)}$, we see that $f \in l_q(\mu)$ if

$$\sum_{j\ge 0} 2^{j\frac{(\alpha p - 2 + (1-\alpha)q)}{2}} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^q = \sum_{j \ge 0} 2^{jq\left[\frac{\alpha(p-q)}{2q} + \frac{1}{2} - \frac{1}{q}\right]} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^q < +\infty.$$

From (2.5), the latter condition holds when $f \in \mathcal{B}^{\delta}_{q,q}$ for

$$\delta = \frac{\alpha}{2}\Big(\frac{p}{q} - 1\Big). \tag{6.21}$$

Now depending on whether we are in the dense (6.9) or sparse phase (6.10) we look for s and π such that

$$\mathcal{B}^{s}_{\pi,r} \subseteq \mathcal{B}^{\delta}_{q,q}. \tag{6.22}$$

The dense phase. By definition (6.9) of $q = q_d$ we have $\pi \ge q_d$. Hence (6.22) follows from (6.26) as long as $s \ge \delta = \frac{\alpha}{2}(\frac{p}{q} - 1)$ which is always true under the dense regime where $q = q_d$. Note that $\delta = \frac{\alpha}{2}(\frac{p}{q_d} - 1) = s$, thus automatically $\delta > 0$.

The sparse phase. Take $q = q_s$ and $\delta = \frac{\alpha}{2}\left(\frac{p}{q_s} - 1\right) = \alpha\frac{sp - \frac{p}{\pi} + 1}{\alpha p - 2}$. We consider two cases. If $\pi > q_s$ we use the embedding (6.26). We have to check that $s > \alpha\frac{sp - \frac{p}{\pi}+1}{\alpha p - 2}$ which is equivalent to $s < \frac{\alpha}{2}\left(\frac{p}{\pi} - 1\right)$, which is true in the sparse case. Further, we must guarantee that $\delta > 0$ which leads to the two conditions i) $p > 2/\alpha$ and $s > \frac{1}{\pi} - \frac{1}{p}$ or ii) $p < 2/\alpha$ and $s < \frac{1}{\pi} - \frac{1}{p}$. However, the last one is not relevant since we have $s > \frac{1}{\pi}$. Thus we established (6.22) for $q_s < \pi < q_d$.

If $\pi < q_s$ we introduce a new Besov scale s' and index $q = q_s$ such that

$$s - \frac{1}{\pi} = s' - \frac{1}{q}, \quad s' = \frac{\alpha}{2}\Big(\frac{p}{q} - 1\Big). \tag{6.23}$$

In this case, (6.27) and (6.21) ensures that

$$\mathcal{B}^s_{\pi,r} \subseteq \mathcal{B}^{s'}_{q,q} \equiv l_q(\mu),$$

as had to be proved. Solving (6.23) yields definition (6.10) of q under the sparse regime.

4.7 Besov embedding and Maxiset condition (Part II)

First we look for a Besov scale δ such that for any $f \in \mathcal{B}^\delta_{p,r}$ the maxiset condition (6.8) is satisfied. Using (3.1), (3.2) it follows

$$c_n^{q-p} \|f - \sum_{\kappa \in \Lambda_n} \beta_\kappa \Psi_\kappa\|_p^p = c_n^{q-p}\, 2^{-j_1 \delta p}\, \|f\|_{\mathcal{B}^\delta_{p,r}} = O\Big(c_n^{q-p+2\delta p}\, (\frac{(\log n)^\alpha}{\log n})^{\delta p}\Big).$$

Thus condition (6.7) holds for any $f \in \mathcal{B}^\delta_{p,r}$ if

$$\delta = \frac{1}{2}(1 - \frac{q}{p}). \tag{6.24}$$

Now we look for s and π such that

$$\mathcal{B}^s_{\pi,r} \subseteq \mathcal{B}^\delta_{p,r}. \tag{6.25}$$

To answer this question, we will use two different types of Besov embedding, depending on whether $\pi \geq p$ or $\pi < p$. We recall that

$$\mathcal{B}^s_{\pi,r} \subseteq \mathcal{B}^{s''}_{p,r}, \quad \text{provided that } \pi \geq p, \text{ and } s \geq s''. \tag{6.26}$$

$$\mathcal{B}^s_{\pi,r} \subseteq \mathcal{B}^{s''}_{p,r}, \quad \text{provided that } \pi < p, \text{ and } s - \frac{1}{\pi} = s'' - \frac{1}{p}. \tag{6.27}$$

The case $\pi \geq p$. We note that in this case we are always in the dense phase since s must be non-negative. Here we use (6.26) with $s'' = \delta$ at (6.24). Hence we see that (6.25) holds as long as $s \geq \frac{1}{2}(1 - \frac{q}{p})$. Using definition (6.9) of $q = q_d$ this will happen when $s \geq \frac{1-\alpha}{2}$.

The dense case when $\pi < p$. Here we introduce a new Besov scale s'' such that $s - \frac{1}{\pi} = s'' - \frac{1}{p}$ and use embedding (6.27). For (6.25) to hold in the dense case we need $s'' \geq \delta$ for $q = q_d$ at (6.9), we obtain the following condition:

$$s \geq \frac{2}{2s + \alpha} + \frac{1}{\pi} - \frac{1}{p}. \tag{6.28}$$

The sparse case when $\pi < p$. Here we introduce a new Besov scale s'' such that $s - \frac{1}{\pi} = s'' - \frac{1}{p}$ and use embedding (6.27). For (6.25) to hold in the sparse case we need $s'' \geq \delta$ for $q = q_s$ at (6.10), we obtain the following condition:

$$s > \frac{1}{\pi} - \frac{\alpha}{2}$$

which is always true since $s > \frac{1}{\pi}$.

4.8 Proof of theorem 3.1 and theorem 3.2

The proof(s) are a direct application of theorem 4.1 with σ_j, c_n and η given in sections 3.1. Combining results of sections 4.4,...,4.7 we see that all the assumptions theorem 4.1 are satisfied. Using the embedding results of section 4.6 we derive rate exponent (3.7) for any $f \in \mathcal{B}^s_{\pi,r}$ from definition (6.9) of q when $s \geq \frac{\alpha}{2}(\frac{p}{\pi} - 1)$. Finally we derive rate exponent (3.9) for any $f \in \mathcal{B}^s_{\pi,r}$ using definition (6.10) of q when $\frac{1}{\pi} - \frac{1}{p} < s < \frac{\alpha}{2}(\frac{p}{\pi} - 1)$.