Wavelet regression in random design with heteroscedastic dependent errors

Rafał Kulik

Marseille, 16 December 2008

Session in memory of Marc Raimondo

Rafał Kulik
Fixed-design regression under Long range Dependence
Fixed-design regression under Long range Dependence

We consider the fixed-design regression problem

\[ Y_i = f(i/n) + \epsilon_i, \]
Fixed-design regression under Long range Dependence

We consider the fixed-design regression problem

\[ Y_i = f(i/n) + \epsilon_i, \]

where \( \epsilon_i, i \geq 1 \) are LRD, Gaussian, with

\[ \rho_m := \mathbb{E}\epsilon_0\epsilon_m \sim Lm^{-\alpha}, \alpha \in (0, 1]. \]
**Fixed-design regression under Long range Dependence**

We consider the *fixed-design regression* problem

\[ Y_i = f(i/n) + \epsilon_i, \]

where \( \epsilon_i, i \geq 1 \) are LRD, Gaussian, with

\[ \rho_m := E\epsilon_0 \epsilon_m \sim Lm^{-\alpha}, \alpha \in (0, 1]. \]  \hspace{1cm} (1)

Then the equivalent continuous model is \(^1\)

\[ dY(t) = f(t)dt + \sigma n^{-\alpha/2} dB_{1-\alpha/2}(t). \]

\(^1\)Brown, Low, Cai, 1990s
Estimator
Estimator

We expand

\[ f(x) = \sum_{k=0}^{\infty} \alpha_{j_0k} \phi_{j_0k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{\infty} \beta_{jk} \psi_{jk}(x), \]

where \( j_0 \) is a low resolution level, \( \phi, \psi \) are scaling and wavelet functions, respectively; and \( h_{jk}(x) := 2^{j/2} h(2^j x - k) \).
Estimator

We expand

\[ f(x) = \sum_{k=0}^{\infty} \alpha_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{\infty} \beta_{jk} \psi_{jk}(x), \]

where \( j_0 \) is a low resolution level, \( \phi, \psi \) are scaling and wavelet functions, respectively; and \( h_{jk}(x) := 2^{j/2} h(2^j x - k) \). We estimate

\[ \hat{f}_{j_1}(x) = \hat{\alpha}_{00} \phi_{00}(x) + \sum_{j=0}^{j_1} \sum_{k=0}^{2^j - 1} \hat{\beta}_{jk} 1\{|\hat{\beta}_{jk}| > \lambda\} \psi_{jk}(x), \quad (2) \]

where

\[ \lambda = \tau 2^{-j(1-\alpha)/2} \sqrt{\log n} \frac{\sqrt{\log n}}{n^{\alpha/2}}. \]
Rates of convergence
Rates of convergence

**Theorem 1.** Assume that $p > 1$, $2^{j_1} \sim (n / \log(n))^\alpha$, $f \in B^{s}_{\pi, r}$ with $s \geq \frac{1}{\pi} \vee \frac{1}{2}$, then there exists a constant $C > 0$ such that for all $n \geq 0$,

\[
E \left\| f - \hat{f}_{j_1} \right\|_p^p \leq C \left( \frac{\log n}{n} \right)^{\frac{1}{\alpha}}, \\
\gamma = \frac{\alpha sp}{2(s + \frac{\alpha}{2})}, \quad \text{if} \quad s \geq \alpha \left( \frac{p}{\pi} - 1 \right), \quad s - \left( \frac{1}{\pi} - \frac{1}{p} \right) > \frac{s}{2s + \alpha}, \quad (3) \\
\gamma = \frac{\alpha p(s - \frac{1}{\pi} + \frac{1}{p})}{2(s - \frac{1}{\pi} + \frac{\alpha}{2})}, \quad \text{if} \quad \frac{1}{\pi} < s < \frac{\alpha p}{2} \left( \frac{1}{\pi} - 1 \right). \quad (4)
\]
Rates of convergence

**Theorem 1.** Assume that $p > 1$, $2^{j_1} \sim (n/\log(n))^\alpha$, $f \in B_{\pi, r}^s$ with $s \geq \frac{1}{\pi} \vee \frac{1}{2}$, then there exists a constant $C > 0$ such that for all $n \geq 0$,

$$
\mathbb{E} \left\| f - \hat{f}_{j_1} \right\|_p^p \leq C \left( \frac{(\log n)^{\frac{1}{\alpha}}}{n} \right) \gamma,
$$

$$
\gamma = \frac{\alpha sp}{2(s + \frac{\alpha}{2})}, \quad \text{if } s \geq \frac{\alpha}{2} \left( \frac{p}{\pi} - 1 \right), \quad s - \left( \frac{1}{\pi} - \frac{1}{p} \right) + \frac{s}{2s + \alpha}, \quad (3)
$$

$$
\gamma = \frac{\alpha p(s - \frac{1}{\pi} + \frac{1}{p})}{2(s - \frac{1}{\pi} + \frac{\alpha}{2})}, \quad \text{if } \frac{1}{\pi} < s < \frac{\alpha}{2} \left( \frac{p}{\pi} - 1 \right). \quad (4)
$$

---

Random-design regression
Random-design regression

Assume that

\[ Y_i = f(X_i) + \sigma(X_i)\epsilon_i, \]

where \( X_i \sim G \) are random.
Random-design regression

Assume that

\[ Y_i = f(X_i) + \sigma(X_i)\epsilon_i, \]

where \( X_i \sim G \) are random.

If \( \epsilon_i \) are i.i.d., it turns out that this model is not equivalent to the continuous model, unless \( X_i \) are standard uniform.
Random-design regression

Assume that

\[ Y_i = f(X_i) + \sigma(X_i)\epsilon_i, \]

where \( X_i \sim G \) are random.

If \( \epsilon_i \) are i.i.d., it turns out that this model is not equivalent to the continuous model, unless \( X_i \) are standard uniform.

Moreover, if \( \epsilon_i \) are LRD it cannot be equivalent to any continuous model.
Wavelet regression, random design and LRD

Heuristic

---

Rafał Kulik

6
Wavelet regression, random design and LRD

Estimator
The estimator we are going to consider has the form as in (2) with:

\[
\hat{\beta}_{jk} := \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i,
\]

(5)
The estimator we are going to consider has the form as in (2) with:

$$\hat{\beta}_{jk} := \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i)) Y_i,$$

(5)

$^3$The idea originates from Kerkyacharian and Picard (2004) - warped wavelets
The estimator we are going to consider has the form as in (2) with:

\[
\hat{\beta}_{jk} := \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i, \tag{5}
\]

where

\[
2^{j_1} \sim \frac{n}{\log n},
\]

\(^{3}\text{The idea originates from Kerkyacharian and Picard (2004) - warped wavelets}\)
The estimator we are going to consider has the form as in (2) with:

\[ \hat{\beta}_{jk} := \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i, \]  

(5)

where

\[ 2^{j_1} \sim \frac{n}{\log n}, \]

\[ \lambda = \lambda_{n,j} = \max \left\{ 2^{-j/2} \mathbf{E} \left[ \psi_{jk}(G(X_1)) \sigma(X_1) \right] \frac{(\log n)^{1/2}}{n^{\alpha/2}}, \frac{\log n}{\sqrt{n}} \right\}. \]

\[ ^{3}\text{The idea originates from Kerkyacharian and Picard (2004) - warped wavelets} \]
### Estimator

The estimator we are going to consider has the form as in (2) with:

\[
\hat{\beta}_{jk} := \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))Y_i,
\]

where

\[
2^{j_1} \sim \frac{n}{\log n},
\]

\[
\lambda = \lambda_{n,j} = \max \left\{ 2^{-j/2} \mathbb{E} [\psi_{jk}(G(X_1)) \sigma(X_1)] \frac{(\log n)^{1/2}}{n^{\alpha/2}}, \frac{\log n}{\sqrt{n}} \right\}.
\]

**Note:** In most regular cases, we have \( \mathbb{E} [\cdot] = 0. \)

---

3 The idea originates from Kerkyacharian and Picard (2004) - warped wavelets
Rates of convergence
Rates of convergence

**Theorem 2.** Assume that $p > 1$, $2^{j_1} \sim n/\log(n)$, $f \circ G^{-1} \in B_{\pi,r}^s$ with $s \geq \frac{1}{\pi} \lor \frac{1}{2}$, then there exists a constant $C > 0$ such that for all $n \geq 0$

$$
E\|f - \hat{f}_n\|_{L^p(g)}^p \leq Cn^{-\gamma(s)}(\log n)^{\gamma(s)},
$$

$$
\gamma(s) = \begin{cases} 
\frac{ps}{2s+1}, & \text{if } s > \frac{p-\pi}{2\pi} \text{ and } \alpha > \alpha_D := \frac{2s}{2s+1}, \\
\frac{s-(1/\pi-1/p)}{1+2(s-1/\pi)}, & \text{if } s < \frac{p-\pi}{2\pi} \text{ and } \alpha > \alpha_S := \frac{2(s-(1/\pi-1/p))}{2(s-1/\pi)+1}.
\end{cases}
$$

Otherwise, $E\|f - \hat{f}_n\|_p^p \leq Cn^{-p\alpha/2}$. 

Rafał Kulik

8
Comments
Comments

• The result is stated in weighted norm $\|f - h\|_{L^p(g)}^p = \int |f - h|^p dG$. 
Comments

• The result is stated in weighted norm $\|f - h\|_{L^p(g)}^p = \int |f - h|^p dG$.

• The estimator matches the minimax rates.
Comments

• The result is stated in weighted norm $\|f - h\|_{L^p(g)}^p = \int |f - h|^p dG$.

• The estimator matches the minimax rates.

• Trichotomy.
Comments

• The result is stated in weighted norm $\|f-h\|^p_{L^p(G)} = \int |f-h|^p dG$.

• The estimator matches the minimax rates.

• Trichotomy.

• One can replace $G$ with $G_n$ (empirical cdf) and get the same rates, however, we must stop at

$$2^{j_1} \sim \sqrt{\frac{n}{\log n}}.$$
Comments

• The result is stated in weighted norm $\|f - h\|_{L^p(g)}^p = \int |f - h|^p dG$.

• The estimator matches the minimax rates.

• Trichotomy.

• One can replace $G$ with $G_n$ (empirical cdf) and get the same rates, however, we must stop at

$$2^j \sim \sqrt{\frac{n}{\log n}}.$$ 

• If we estimate $f^* = f - E[f(X_1)]$ we do not have LRD effect.
Proof
Proof

\[ \hat{\beta}_{jk} - \beta_{jk} = \frac{1}{n} \sum_{i=1}^{n} \left( \psi_{jk}(G(X_i))Y_i - \mathbb{E}[\psi_{jk}(G(X_i))Y_i] \right) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \left( \psi_{jk}(G(X_i))f(X_i) - \mathbb{E}[\psi_{jk}(G(X_1))f(X_1)] \right) \]

\[ + \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i =: A_0 + A_1. \]
Proof

$$
\hat{\beta}_{jk} - \beta_{jk} = \frac{1}{n} \sum_{i=1}^{n} (\psi_{jk}(G(X_i))Y_i - E[\psi_{jk}(G(X_i))Y_i])
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} (\psi_{jk}(G(X_i))f(X_i) - E[\psi_{jk}(G(X_1))f(X_1)])
$$

$$
+ \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i =: A_0 + A_1.
$$

(6)

Note that $A_0$ is the sum of iid random variables, whereas the dependence structure is included in $A_1$ only.
Proof

\[ \hat{\beta}_{jk} - \beta_{jk} = \frac{1}{n} \sum_{i=1}^{n} (\psi_{jk}(G(X_i))Y_i - \mathbb{E}[\psi_{jk}(G(X_i))Y_i]) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} (\psi_{jk}(G(X_i))f(X_i) - \mathbb{E}[\psi_{jk}(G(X_1))f(X_1)]) \]

\[ + \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i =: A_0 + A_1. \] (6)

Note that \( A_0 \) is the sum of iid random variables, whereas the dependence structure is included in \( A_1 \) only.

The part \( A_1 \) is decomposed further. Let \( \mathcal{F}_i = \sigma(X_i, \eta_i, X_{i-1}, \eta_{i-1}, \ldots) \). Let \( \epsilon_{i,i-1} = \epsilon_i - \eta_i \). Then,

\[ \mathbb{E}[\psi_{jk}(G(X_i))\sigma(X_i)\epsilon_i|\mathcal{F}_{i-1}] = \epsilon_{i,i-1}\mathbb{E}[\psi_{jk}(G(X_1))\sigma(X_1)]. \]
Proof - ctd.
Proof - ctd.

Write

\[
A_1 = \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i)) \sigma(X_i) \epsilon_i
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (\psi_{jk}(G(X_i)) \epsilon_i - \mathbb{E}[\psi_{jk}(G(X_i)) \sigma(X_i) \epsilon_i | \mathcal{F}_{i-1}])
\]

\[
+ \frac{1}{n} \mathbb{E}[\psi_{jk}(G(X_1)) \sigma(X_1)] \sum_{i=1}^{n} \epsilon_{i,i-1} =: A_2 + A_3.
\]
Proof - ctd.

Write

\[ A_1 = \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(G(X_i)) \sigma(X_i) \epsilon_i \]

\[ = \frac{1}{n} \sum_{i=1}^{n} (\psi_{jk}(G(X_i)) \epsilon_i - E[\psi_{jk}(G(X_i)) \sigma(X_i) \epsilon_i | \mathcal{F}_{i-1}]) \]

\[ + \frac{1}{n} E[\psi_{jk}(G(X_1)) \sigma(X_1)] \sum_{i=1}^{n} \epsilon_{i,i-1} =: A_2 + A_3. \]

The crucial feature of this decomposition is that we reduce the problem of studying the partial sums \( \sum_{i=1}^{n} \psi_{jk}(G(X_i)) \sigma(X_i) \epsilon_i \) to the partial sums \( \sum_{i=1}^{n} \epsilon_{i,i-1} \) and the martingale \( A_2 \).
This leads to

$$E \left[ |\hat{\beta}_{jk} - \beta_{jk}|^p \right] \leq \begin{cases} C n^{-p/2} & \text{if } 2^j > n^{1-\alpha}, \\ C 2^{-j p/2} n^{-p \alpha / 2} & \text{if } 2^j < n^{1-\alpha} \end{cases}$$

and appropriate large deviations inequalities.
This leads to

$$\mathbb{E} \left[ |\hat{\beta}_{jk} - \beta_{jk}|^p \right] \leq \begin{cases} C n^{-p/2} & \text{if } 2^j > n^{1-\alpha}, \\ C 2^{-j p/2} n^{-p\alpha/2} & \text{if } 2^j < n^{1-\alpha} \end{cases}$$

and appropriate large deviations inequalities.

In principle, low resolution levels are affected by long memory, high resolution levels are not.
Dichotomy