Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Nonparametric Regression with Warped Wavelets and Strong Mixing Processes

Luz M. Gómez · Rogério F. Porto · Pedro A. Morettin

Received: date / Revised: date

1 Supplementary Material: Proof

1.1 Proof of Proposition 1

Proof of item 1

This is already proved in Kerkyacharian and Picard (2004, p.1059) but we reproduce here for completeness. The proof is the same either the error or the predictor follows a strong-mixing stochastic process and the other term follows an IID sequence.

$$
\begin{aligned}
E(\hat{\beta}_{j,k}) &= \frac{1}{n}\sum_{i=1}^{n} E(\psi_{j,k}(G(X_i))(f(X_i)+\epsilon_i)) \\
&= E(\psi_{j,k}(G(X_i))f(X_i)) \\
&= \int_a^b \psi_{j,k}(G(x))f(x)g(x)\,dx \\
&= \int_a^b \psi_{j,k}(y)f(G^{-1}(y))\,dy \\
&= \beta_{j,k}.
\end{aligned}
$$

Proof of item 2

Luz M. Gómez (corresponding author)
Institute of Mathematics and Statistics, University of São Paulo, SP, Brazil.
Rua do Matão, 1010 - CEP 05508-090 - São Paulo, SP, Brazil
E-mail: lgomez928@gmail.com

Rogério F. Porto
Bank of Brazil, Brazil.
SAUN Quadra 5, Lote B, Ed. Green Towers - CEP 70742-010, Brasília, DF, Brazil.

Pedro A. Morettin
Institute of Mathematics and Statistics, University of São Paulo, SP, Brazil.
Rua do Matão, 1010 - CEP 05508-090 - São Paulo, SP, Brazil.

This is a special case of item 3.

Proof of item 3

When either the error or the predictor follows a strong-mixing stochastic process and the other term follows an IID sequence, we have that

$$\begin{aligned}
&\mathrm{Cov}\left(\hat{\beta}_{j,k}, \hat{\beta}_{j',k'}\right) \\
&= \mathrm{Cov}\left(n^{-1}\sum_{r=1}^{n}\psi_{j,k}(G(X_r))Y_r, n^{-1}\sum_{s=1}^{n}\psi_{j',k'}(G(X_s))Y_s\right) \\
&= n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\mathrm{Cov}\left(\psi_{j,k}(G(X_r))Y_r, \psi_{j',k'}(G(X_s))Y_s\right) \\
&= n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\Big[\mathrm{Cov}\left(\psi_{j,k}(G(X_r))f(X_r), \psi_{j',k'}(G(X_s))f(X_s)\right) \\
&\quad + \mathrm{Cov}\left(\psi_{j,k}(G(X_r))f(X_r), \psi_{j',k'}(G(X_s))\epsilon_s\right) \\
&\quad + \mathrm{Cov}\left(\psi_{j,k}(G(X_r))\epsilon_r, \psi_{j',k'}(G(X_s))f(X_s)\right) \\
&\quad + \mathrm{Cov}\left(\psi_{j,k}(G(X_r))\epsilon_r, \psi_{j',k'}(G(X_s))\epsilon_s\right)\Big] \\
&= n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\left[A + B + C + D\right].
\end{aligned}$$

The terms B, C and D have the same value when either the error is strong mixing and the predictor is IID or vice-versa. To see this, we write

$$\begin{aligned}
B &= E\left(\psi_{j,k}(G(X_r))f(X_r)\psi_{j',k'}(G(X_s))\epsilon_s\right) \\
&\quad - E\left(\psi_{j,k}(G(X_r))f(X_r)\right)E\left(\psi_{j',k'}(G(X_s))\epsilon_s\right) \\
&= E\left(\psi_{j,k}(G(X_r))f(X_r)\psi_{j',k'}(G(X_s))\right)E(\epsilon_s) \\
&\quad - E\left(\psi_{j,k}(G(X_r))f(X_r)\right)E\left(\psi_{j',k'}(G(X_s))\right)E(\epsilon_s).
\end{aligned}$$

Thus, $B = 0$ and, by symmetry, $C = 0$. For the term D, we write

$$\begin{aligned}
D &= E\left(\psi_{j,k}(G(X_r))\epsilon_r\psi_{j',k'}(G(X_s))\epsilon_s\right) \\
&\quad - E\left(\psi_{j,k}(G(X_r))\epsilon_r\right)E\left(\psi_{j',k'}(G(X_s))\epsilon_s\right) \\
&= E\left(\psi_{j,k}(G(X_r)))\psi_{j',k'}(G(X_s))\right)E(\epsilon_r\epsilon_s).
\end{aligned}$$

If $r = s$

$$\begin{aligned}
D &= \sigma^2 E(\psi_{j,k}(G(X))\psi_{j',k'}(G(X))) \\
&= \sigma^2\int_a^b \psi_{j,k}(G(x))\psi_{j',k'}(G(x))g(x)\,dx.
\end{aligned}$$

Thus, $|D| \leq \sigma^2 2^{(j+j')/2}\|\psi\|_\infty^2$.

If $r \neq s$, $D = 0$ because, when the error is IID, $E(\epsilon_r\epsilon_s) = E(\epsilon_r)E(\epsilon_s) = 0$; and when the predictor is IID, $E(\psi_{j,k}(G(X_r))) = \int_a^b \psi_{j,k}(G(x))g(x)\,dx = \int_0^1 \psi_{j,k}(y)\,dy = 0$.

It remains to evaluate the value of the term A. We begin with the case of dependent errors and IID predictors by writing

$$\begin{aligned} A = {} & E\left(\psi_{j,k}(G(X_r))f(X_r)\psi_{j',k'}(G(X_s))f(X_s)\right) \\ & - E\left(\psi_{j,k}(G(X_r))f(X_r)\right)E\left(\psi_{j',k'}(G(X_s))f(X_s)\right). \end{aligned}$$

It is easy to see that $A = 0$ if $r \neq s$. However, if $r = s$,

$$\begin{aligned} A &= E\left(\psi_{j,k}(G(X_1))\psi_{j',k'}(G(X_1))f^2(X_1)\right) \\ &\quad - E\left(\psi_{j,k}(G(X_1))f(X_1)\right)E\left(\psi_{j',k'}(G(X_1))f(X_1)\right) \\ &= \int_0^1 f^2(G^{-1}(y))\psi_{j,k}(y)\psi_{j',k'}(y)\,dy - \beta_{j,k}\beta_{j',k'}. \end{aligned}$$

Thus, $|A| \leq \|f\|_\infty^2 2^{(j+j')/2}\|\psi\|_\infty^2$.

Since the only non-zero terms occur when $r = s$, and they are constant, then

$$n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}[A+B+C+D] = n^{-1}[A+D] = O(n^{-1}).$$

Now, for the case of dependent predictors and IID errors, we note that the sequence $\{W_{r,j,k} = \psi_{j,k}(G(X_r))f(X_r), r = 1,\ldots,n\}$ is a portion of an α-mixing stochastic process.

Let the dependence coefficient between $W_{t,j,k}$ and $W_{t+h,j',k'}$ be given by $\alpha_{W,h}$. Then, as given for instance by Theorem 3 of Chapter 1.2 in Doukhan (1994), we can bound the term A as

$$|\mathrm{Cov}(W_{r,j,k}, W_{s,j',k'})| \leq 8\alpha_{W,|r-s|}^{1/r'}\left(E(|W_{r,j,k}|^p)\right)^{1/p}\left(E(|W_{s,j',k'}|^q)\right)^{1/q},$$

for any $p, q, r' \geq 1$ and $1/p + 1/q + 1/r' = 1$. Calculating the expected values we have that

$$\begin{aligned} E(|W_{r,j,k}|^p) &= E(|\psi_{j,k}(G(X_r))f(X_r)|^p) \\ &= \int_0^1 |\psi_{j,k}(y)f(G^{-1}(y))|^p\,dy \\ &\leq ||f||_\infty^p \int_0^1 |\psi_{j,k}(y)|^p\,dy \\ &\leq ||f||_\infty^p 2^{j(p-2)/2}||\psi||_\infty^{p-2}\int_0^1 |\psi_{j,k}(y)|^2\,dy \\ &= ||f||_\infty^p 2^{j(p-2)/2}||\psi||_\infty^{p-2}. \end{aligned}$$

Similarly calculating $E(|W_{s,j',k'}|^q)$, we have the following bound:

$$\begin{aligned} |A| &= |\mathrm{Cov}\left(\psi_{j,k}(G(X_r))f(X_r), \psi_{j',k'}(G(X_s))f(X_s)\right)| \\ &= |\mathrm{Cov}(W_{r,j,k}, W_{s,j',k'})| \\ &\leq 8\alpha_{W,|r-s|}^{1/r'}||f||_\infty^2 2^{j(p-2)/(2p)+j'(q-2)/(2q)}||\psi||_\infty^{(p-2)/p+(q-2)/q} \\ &= \alpha_{W,|r-s|}^{1/r'}C(f,j,j',p,q,\psi), \end{aligned}$$

where $C(f,j,j',p,q,\psi) = 8||f||_\infty^2 2^{j(p-2)/(2p)+j'(q-2)/(2q)}||\psi||_\infty^{(p-2)/p+(q-2)/q}$.

Since the α-mixing coefficient $\alpha_{W,|r-s|}$ is defined on a sub-σ-algebra induced by the definition $W_{r,j,k} = \psi_{j,k}(G(X_r))f(X_r)$, by Assumption 1, we have that

$$\sum_{h=1}^{\infty}(h+1)^{c-2}\alpha_{W,h}^{1/r'} = C',$$

for some $c \in 2\mathbb{N} = 0, 2, 4, 6, \ldots$ and $r' > 0$, where $C' < \infty$ does not depend on n.

Then

$$\begin{aligned}
n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}[A+B+C+D] &= n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}A + n^{-1}D \\
&\le \left[n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\alpha_{W,|r-s|}^{1/r'}C(f,j,j',p,q,\psi)\right] + n^{-1}D \\
&\le \left[n^{-2}\sum_{r=1}^{n}C'C(f,j,j',p,q,\psi)\right] + n^{-1}D \\
&= O(n^{-1}).
\end{aligned}$$

1.2 Proof of Proposition 2

Proof of item 1

The proof is almost the same either the error or the predictor follows a strong-mixing stochastic process and the other term follows an IID sequence. The only difference occurs because it uses the proof third item of Proposition 1, which depends on the error or the predictor being IID or strong-mixing.

We begin by noting that

$$\begin{aligned}
E(|\hat{\beta}_{j,k} - \beta_{j,k}|^p) &= E\left(\left|n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))Y_i - d_{j,k}\right|^p\right) \\
&= E\left(n^{-p}\left|\sum_{i=1}^{n}[\psi_{j,k}(G(X_i))Y_i - d_{j,k}]\right|^p\right) \\
&= E\left(n^{-p}\left|\sum_{i=1}^{n}Z_{i;j,k}\right|^p\right).
\end{aligned}$$

By the first item of Proposition 1, $E(Z_{i;j,k}) = 0$, for all i, j and k considered. Following the proof of the third item of Proposition 1, we also have that, for all $r \neq s$, $\text{Cov}\,(Z_{r;j,k}, Z_{s;j,k}) = 0$, and when $r = s$, $\text{Cov}\,(Z_{r;j,k}, Z_{r;j,k}) < \infty$. Then $Z_{i;j,k}$ belongs to $L_2[a,b]$ and is centered.

Let the dependence coefficient between $Z_{t,j,k}$ and $Z_{t+h,j,k}$ be given by $\alpha_{Z,h}$. Since this coefficient is defined on a sub-σ-algebra induced by the definition of

$Z_{t,j,k}$ in terms of X_t and ϵ_t, by Assumption 1, given $p > 1$, there exists $c > p$, $c \in 2\mathbb{N} = 0, 2, 4, 6, \ldots$ and $\delta > 0$, such that $\sum_{h=1}^{\infty}(h+1)^{c-2}\alpha_{Z,h}^{\delta/(c+\delta)} \leq \infty$.

We are going to use the following Lemma.

Lemma 1 Let X be a random variable with finite mean μ and finite moment of order p. Then $E(|X-\mu|^p) \leq CE(|X|^p)$, for some constant C which depends on p.

Proof. By the Hölder inequality, $E(|X-\mu|^p) = \int_0^\infty P(|X-\mu|^p > y)dy$ is less than or equal to

$$\begin{aligned}\int_0^\infty P(2^{p-1}(|X|^p + |-\mu|^p) > y)dy &= \int_0^\infty P(|X|^p > \frac{y}{2^{p-1}} - |-\mu|^p)dy \\ &= \int_{-|-\mu|^p}^\infty P(|X|^p > z)2^{p-1}dz \\ &= \int_0^\infty P(|X|^p > z)2^{p-1}dz \\ &= 2^{p-1}E(|X|^p).\end{aligned}$$

Thus, by applying Theorem 2 of Chapter 1.4 in Doukhan (1994), for $p > 2$, and the previous lemma, we have that

$$\begin{aligned}& n^{-p}E\left(\left|\sum_{i=1}^n Z_{i;j,k}\right|^p\right) \\ &= \frac{C}{n^p}\max\left\{\sum_{i=1}^n\left[E\left(|Z_{i;j,k}|^{p+\delta}\right)\right]^{\frac{p}{p+\delta}}, \left[\sum_{i=1}^n\left(E\left(|Z_{i;j,k}|^{2+\delta}\right)\right)^{\frac{2}{2+\delta}}\right]^{p/2}\right\} \\ &\leq \frac{C}{n^p}\left\{\sum_{i=1}^n\left[E\left(|\psi_{j,k}(G(X_i))Y_i|^{p+\delta}\right)\right]^{\frac{p}{p+\delta}}\right. \\ &\quad \left. + \left[\sum_{t=i}^n\left(E\left(|\psi_{j,k}(G(X_i))Y_i|^{2+\delta}\right)\right)^{\frac{2}{2+\delta}}\right]^{p/2}\right\} \\ &= C\left\{\frac{\left[E\left(|\psi_{j,k}(G(X_t))Y_t|^{p+\delta}\right)\right]^{\frac{p}{p+\delta}}}{n^{p-1}} + \frac{\left(E\left(|\psi_{j,k}(G(X_t))Y_t|^{2+\delta}\right)\right)^{\frac{p}{2+\delta}}}{n^{p/2}}\right\},\end{aligned}$$

for some constant $C > 0$ that does not depend on n.

By the result at line 19 of page 1086 in Kerkyacharian and Picard (2004), this last expression is less than or equal to

$$C\left\{\frac{(1+||f||_\infty^{p+\delta})^{\frac{p}{p+\delta}}2^{j(\frac{p+\delta}{2}-1)\frac{p}{p+\delta}}}{n^{p-1}} + \frac{(1+||f||_\infty^{2+\delta})^{\frac{2}{p+\delta}}2^{j(\frac{p+\delta}{2}-1)\frac{p}{p+\delta}}}{n^{p/2}}\right\} = C(A+B).$$

For $p > 2$,

$$A \leq C(n/\log n)^{p/2} \Leftrightarrow 2^{j((p+\delta)/2-1)p/(p+\delta)} \leq Cn^{3p/2-1}/(\log n)^{p/2}$$
$$\Leftrightarrow 2^{j(p+\delta-2)p/(2(p+\delta))} \leq Cn^{(3p-2)/2}/(\log n)^{p/2}$$
$$\Leftrightarrow 2^j \leq C\left(\frac{n^{(3p-2)/p}}{\log n}\right)^{\frac{p+\delta}{p+\delta-2}}.$$

Similarly,

$$B \leq C(n/\log n)^{p/2} \Leftrightarrow 2^{j((p+\delta)/2-1)p/(p+\delta)} \leq Cn^{p}/(\log n)^{p/2}$$
$$\Leftrightarrow 2^{j(p+\delta-2)p/(2(p+\delta))} \leq Cn^{p}/(\log n)^{p/2}$$
$$\Leftrightarrow 2^j \leq C\left(\frac{n^{2}}{\log n}\right)^{\frac{p+\delta}{p+\delta-2}} \leq C\left(\frac{n^{(3p-2)/p}}{\log n}\right)^{\frac{p+\delta}{p+\delta-2}}.$$

Now, for $1 < p \leq 2$, using the same previous arguments, we have that

$$\begin{aligned} & n^{-p}E\left(\left|\sum_{i=1}^{n} Z_{i;j,k}\right|^p\right) \\ &= \frac{C}{n^p}\sum_{i=1}^{n}\left[E\left(|Z_{i;j,k}|^{p+\delta}\right)\right]^{\frac{p}{p+\delta}} \\ &\leq \frac{C}{n^p}\sum_{i=1}^{n}\left[E\left(|\psi_{j,k}(G(X_i))Y_i|^{p+\delta}\right)\right]^{\frac{p}{p+\delta}} \\ &\leq \frac{C}{n^{p-1}}\left[(1+||f||_{\infty}^{p+\delta})2^{j(\frac{p+\delta}{2}-1)}\right]^{\frac{p}{p+\delta}}. \end{aligned}$$

Similarly to the case when $p > 2$, this result is bounded from above by $C(n/\log n)^{p/2}$ if

$$2^j \leq C\left(\frac{n^{(3p-2)/p}}{\log n}\right)^{\frac{p+\delta}{p+\delta-2}}.$$

Note that

$$\left(\frac{n}{\log n}\right)^{1/2} \leq \left(\frac{n^{(3p-2)/p}}{\log n}\right)^{\frac{p+\delta}{p+\delta-2}}$$

for all $\delta > 0$, when $p \geq 2$ and, when $1 < p < 2$, this is also true for all $\delta \geq 1$. This is not true only when $1 < p < 2$ and $0 < \delta < 1$. Thus,

$$E(|\hat{d}_{j,k} - d_{j,k}|^p) \leq C(n/\log n)^{p/2},$$

for all $2^{J_1} \leq C\min\{n_1, n_2\}$, where

$$n_1 = \left(\frac{n}{\log n}\right)^{1/2}, \quad n_2 = \left(\frac{n^{(3p-2)/p}}{\log n}\right)^{\frac{p+\delta}{p+\delta-2}}.$$

Proof of item 2

The main steps of this proof are the same in Kerkyacharian and Picard (2004), used to prove their result (65) at page 1086. However, since the errors are dependent, we need to make the following adaptations at page 1087 in Kerkyacharian and Picard (2004).

First note that given data $(X_1 = x_1, \ldots, X_n = x_n)$, the random variable $n^{-1}\sum_{i=1}^n \psi_{j,k}(G(X_i))\epsilon_i$ is normally distributed with mean zero and

$$\begin{aligned}
&\mathrm{Var}\left(n^{-1}\sum_{i=1}^n \psi_{j,k}(G(x_i))\epsilon_i\right)\\
&= E\left(\left(n^{-1}\sum_{i=1}^n \psi_{j,k}(G(x_i))\epsilon_i\right)^2\right)\\
&= n^{-2}\sum_{r=1}^n\sum_{s=1}^n \psi_{j,k}(G(x_r))\psi_{j,k}(G(x_s))E(\epsilon_r\epsilon_s)\\
&= n^{-2}\sum_{r=1}^n\sum_{s=1}^n \psi_{j,k}(G(x_r))\psi_{j,k}(G(x_s))\mathrm{Cov}(\epsilon_r,\epsilon_s).
\end{aligned}$$

Now, let $Z_r = \sum_{s=1}^n \psi_{j,k}(G(X_r))\psi_{j,k}(G(X_s))\mathrm{Cov}(\epsilon_r,\epsilon_s)$. We have that these variables are bounded

$$|Z_r| \le 2^j||\psi||_\infty^2 \sum_{s=1}^\infty |\mathrm{Cov}(\epsilon_r,\epsilon_s)| < M < \infty,$$

because, by Assumption 1 and Theorem 3 of Chapter 1.2 in Doukhan (1994),

$$\sum_{s=1}^\infty |\mathrm{Cov}(\epsilon_r,\epsilon_s)| \le \sum_{s=1}^\infty 8\alpha_{\epsilon,s}^{1/m}\left(E(|\epsilon_r|^p)\right)^{1/p}\left(E(|\epsilon_s|^q)\right)^{1/q}$$

for any p, q, $m \ge 1$, $1/p + 1/q + 1/m = 1$, and we can choose p, q and m such that, by Assumption 1,

$$\begin{aligned}
&\sum_{s=1}^\infty 8\alpha_{\epsilon,s}^{1/m}\left(E(|\epsilon_r|^p)\right)^{1/p}\left(E(|\epsilon_s|^q)\right)^{1/q}\\
&\le \sum_{s=1}^\infty (s+1)^{\delta(m-1)-2}\left(\alpha_{\epsilon,s}\right)^{1/m}\left(E(|\epsilon_r|^p)\right)^{1/p}\left(E(|\epsilon_s|^q)\right)^{1/q}\\
&= \left(E(|\epsilon_r|^p)\right)^{1/p}\left(E(|\epsilon_s|^q)\right)^{1/q}\sum_{s=1}^\infty (s+1)^{c-2}\left(\alpha_{\epsilon,s}\right)^{\delta/(c+\delta)} < \infty.
\end{aligned}$$

Also we have,

$$E(Z_r) = \sum_{s=1}^n E\left(\psi_{j,k}(G(X_r))\psi_{j,k}(G(X_s))\right)\mathrm{Cov}(\epsilon_r,\epsilon_s) = E\left(\psi_{j,k}^2(G(X_r))\right)\sigma^2 = 1.$$

In the last step we used $\sigma = 1$, without loss of generality and to keep up with Kerkyacharian and Picard (2004). Then we continue noting that $\mathrm{Var}(Z_r) = E(Z_r^2) - 1$ where, again using Assumption 1,

$$\begin{aligned}
E(Z_r^2) &= E\Bigg(\psi_{j,k}^2(G(X_r))\sum_{l=1}^{n}\sum_{m=1}^{n}\psi_{j,k}(G(X_l))\psi_{j,k}(G(X_m)) \\
&\qquad \times \mathrm{Cov}(\epsilon_r,\epsilon_l)\mathrm{Cov}(\epsilon_r,\epsilon_m)\Bigg) \\
&= E\left(\psi_{j,k}^2(G(X_r))\sum_{l=1}^{n}\psi_{j,k}^2(G(X_l))\mathrm{Cov}(\epsilon_r,\epsilon_l)^2\right) \\
&= E\left(\psi_{j,k}^4(G(X_r)) + \psi_{j,k}^2(G(X_r))\sum_{\substack{l=1\\ l\neq r}}^{n}\psi_{j,k}^2(G(X_l))\mathrm{Cov}(\epsilon_r,\epsilon_l)^2\right) \\
&= \int_a^b \psi_{j,k}^4(G(x))g(x)\,dx + \sum_{\substack{l=1\\ l\neq r}}^{n}\mathrm{Cov}(\epsilon_r,\epsilon_l)^2 \\
&\leq 2^{2j}||\psi||_\infty^4 + \sum_{\substack{l=1\\ l\neq r}}^{\infty}\mathrm{Cov}(\epsilon_r,\epsilon_l)^2.
\end{aligned}$$

Thus, Z_r are random variables independent but not identically distributed.

Let

$$\sigma_* = \sup_r \sum_{\substack{l=1\\ l\neq r}}^{\infty}\mathrm{Cov}(\epsilon_r,\epsilon_l)^2$$

and

$$b_n^2 = \sum_{r=1}^{n} E((Z_r-1)^2) = \sum_{r=1}^{n} E(Z_r^2) - n.$$

Now, using the Hoeffding's inequality as given at item (i) of Corollary C.1 in Härdle et al (1998), we have

$$
\begin{aligned}
P\left(n^{-1}|\sum_{i=1}^{n}(Z_i-1)|>\alpha\right) &= P\left(|\sum_{i=1}^{n}(Z_i-1)|>n\alpha\right)\\
&\leq 2\exp\left(-\frac{n^2\alpha^2}{2\left(\sum_{i=1}^{n}E(Z_i^2)-n+n\alpha M/3\right)}\right)\\
&\leq 2\exp\left(-\frac{n^2\alpha^2}{2\left(\sum_{i=1}^{n}E(Z_i^2)+n\alpha M\right)}\right)\\
&\leq 2\exp\left(-\frac{n^2\alpha^2}{2\left(n2^{2j}||\psi||_\infty^4+n\sigma_*+n\alpha M\right)}\right)\\
&\leq 2\exp\left(-\frac{2n\alpha^2}{C_r2^{2j}}\right)\\
&\leq 2\exp\left(-\frac{2n\alpha^2}{C_r}\log n\right)\\
&= 2n^{-2\alpha^2/C_r},
\end{aligned}
$$

if $2^j\leq\sqrt{n/\log n}=n_3$, where $C_r=4(||\psi||_\infty^4+\sigma_*+||\psi||_\infty^4\sum_{s=1}^{\infty}\text{Cov}(\epsilon_r,\epsilon_s)$.

This is equivalent to result (66) in Kerkyacharian and Picard (2004) such that we just need to follow the rest of the proof of their result (65).

Proof of item 3

The main steps of this proof are the same in Kerkyacharian and Picard (2004), used to prove their result (65) at page 1086. But now, since the predictors are dependent, we need to make the following adaptations at page 1087 in Kerkyacharian and Picard (2004).

$$
\begin{aligned}
\hat{\beta}_{j,k}-\beta_{j,k} &= n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\left(f(X_i)+\epsilon_i\right)-\beta_{j,k}\\
&= n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))f(X_i)-E\left(\psi_{j,k}(G(X_i))f(X_i)\right)\\
&\quad+n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i
\end{aligned}
$$

Initially, observe that given $X_1 = x_1, \ldots, X_n = x_n$, the random variable $n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i$ is normally distributed with mean zero and variance

$$\begin{aligned}
\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i\right) &= E\left(\left[n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i\right]^2\right)\\
&= n^{-2}\sum_{i=1}^{n}E\left(\psi_{j,k}^2(G(X_i))\epsilon_i^2\right)\\
&= n^{-2}\sum_{i=1}^{n}\psi_{j,k}^2(G(X_i)E\left(\epsilon_i^2\right)\\
&= n^{-2}\sigma^2\sum_{i=1}^{n}\psi_{j,k}^2(G(X_i)).
\end{aligned}$$

We know that $|\psi_{j,k}^2(G(X_i))| \le 2^j \left\|\psi\right\|_\infty^2$, hence

$$\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i\right) \le n^{-1}2^j \left\|\psi\right\|_\infty^2 \sigma^2.$$

Following Kerkyacharian and Picard (2004) at page 1087, hereafter we consider $\sigma = 1$ and have

$$\begin{aligned}
&P\left(\left|n^{-1}\sum_{i=1}^{n}\psi_{j,k}(G(X_i))\epsilon_i\right| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right)\\
&\le P\left(\left|\sum_{i=1}^{n}\left[\psi_{j,k}^2(G(X_i)) - 1\right]\right| > \alpha\right) + \exp\left(-\frac{\kappa^2\log n}{8(1+\alpha)}\right).
\end{aligned}$$

Note that the sequence $\left\{\psi_{j,k}^2(G(X_i)), i = 1, \ldots, n\right\}$ is α-mixing, with random variables bounded by $2^j||\psi||_\infty^2$ and $E(\psi_{j,k}^2(G(X_i))) = 1$. Then, if Conditions $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ of Corollary 1 in Doukhan and Louhichi (1999) are satisfied, we can use it to find an adequate upper bound for the previous probability.

By Lemma 6 in Doukhan and Louhichi (1999), we have that $C_{r,q} \le 4M^q\alpha_r$, hence Condition $(\mathcal{H}_1)$ is satisfied with $\epsilon_r = \alpha_r$.

Now we check that Condition $(\mathcal{H}_2)$ is satisfied.

$$\begin{aligned} M_{q,n} &= n \sum_{r=0}^{n-1} (r+1)^{q-2} C_{r,q} \\ &\leq n \sum_{r=0}^{n-1} (r+1)^{q-2} 4 M^q \alpha_r \\ &\leq 4nM^q \sum_{r=0}^{n-1} (r+1)^{q-2} \alpha_r \\ &\leq 4nM^q \sum_{r=0}^{\infty} (r+1)^{q-2} \alpha_r^{\delta/(\delta+q)}, \end{aligned}$$

and by Assumption 1, $\sum_{r=0}^{\infty} (r+1)^{q-2} \alpha_r^{\delta/(\delta+q)} < L < \infty$. Hence

$$M_{q,n} \leq 4nM^q L \leq q! n M^q L = A_n \frac{q!}{\beta^q},$$

for $q > 2$, where $A_n = nL$ and $\beta = 1/M = 1/(2^j \|\psi\|_\infty^2)$. Taking $\alpha = n \log n \sqrt{A_n}$ and using Corollary 1 in Doukhan and Louhichi (1999), we have that

$$\begin{aligned} P\left(\left| \sum_{i=1}^{n} \psi_{j,k}^2 (G(X_i)) - 1 \right| > \alpha \right) &= P\left(\left| \sum_{i=1}^{n} \psi_{j,k}^2 (G(X_t)) - 1 \right| > n^{3/2} \log n \sqrt{L} \right) \\ &\leq A \exp\left(-B \sqrt{\frac{n \log n}{2^j \|\psi\|_\infty^2}} \right) \\ &\leq A \exp\left(-B \sqrt{\frac{n (\log n)^2}{n \|\psi\|_\infty^2}} \right) \\ &= A \exp\left(-B \frac{\log n}{\|\psi\|_\infty} \right) \\ &= A n^{-B/\|\psi\|_\infty}. \end{aligned}$$

for $2^j \leq \sqrt{n/(\log n)} \leq n/(\log n)$.

Now consider the other inequality,

$$P\left(\left| n^{-1} \sum_{i=1}^{n} \psi_{j,k}(G(X_i)) f(X_i) - E\left(\psi_{j,k}(G(X_i)) f(X_i) \right) \right| > \frac{\kappa}{2} \sqrt{\frac{\log n}{n}} \right).$$

Since $Q_{i;j,k} = \psi_{j,k}(G(X_i)) f(X_i) - E\left(\psi_{j,k}(G(X_i)) f(X_i)\right)$ is α-mixing, with $E\left(Q_{i;j,k}\right) = 0$, then we can use the Corollary 1 in Doukhan and Louhichi (1999). We can use this corollary because, similarly to the previous paragraphs, the variable $Q_{i;j,k}$ satisfies the Conditions $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ $\beta = 1/M$ and $x = n \log n$. Therefore we have

$$\begin{aligned} P\left(n^{-1}\left|\sum_{i=1}^{n} Q_{i;j,k}\right| > \frac{\kappa}{2}\sqrt{\frac{\log n}{n}}\right) &\leq A \exp\left(-B\sqrt{\frac{n\log n}{2^j||\psi||_\infty^2}}\right) \\ &\leq A \exp\left(-B\sqrt{\frac{n(\log n)^2}{n||\psi||_\infty^2}}\right) \\ &= A \exp\left(-B\frac{\log n}{||\psi||_\infty}\right) \\ &= An^{-B/\|\psi\|_\infty} \end{aligned}$$

for $2^j \leq \sqrt{n/(\log n)} \leq n/(\log n)$.

1.3 Proof of Proposition 3

Proof of item 1

We have that

$$|E(\hat{\beta}^{@}_{j,k}) - \beta_{j,k}| \leq E(|\hat{\beta}^{@}_{j,k} - \beta_{j,k}|) \leq \left(E(|\hat{\beta}^{@}_{j,k} - \beta_{j,k}|^2)\right)^{1/2} \leq \frac{C}{\sqrt{n}},$$

where C is a general positive constant, by Jensen's inequality and Proposition 6 in Kerkyacharian and Picard (2004, p.1072), which is valid, as can be seen from its proof, when either the error or the predictor follows a strong-mixing stochastic process and the other term follows an IID sequence.

Proof of item 2

This is a special case of item 3.

Proof of item 3

Follow the proof of Proposition 1, item 3, using $\hat{G}$ instead of G and conclude that $B = C = 0$.

For the term D, when $r = s$, since $|\psi_{j,k}(\hat{G}(x))| \leq 2^{j/2}\|\psi\|_\infty$, we have that $|D| \leq \sigma^2 2^{(j+j')/2}\|\psi\|^2_\infty$.

Now, when $r \neq s$ and the error is IID, $D = 0$ because $E(\epsilon_r\epsilon_s) = E(\epsilon_r)E(\epsilon_s) = 0$. When the predictor is IID, take $f \equiv 1$ in model (7) such that

$$\beta_{j,k} = \int_a^b \psi_{j,k}(G(x))f(x)g(x)\,dx = E\left(\psi_{j,k}(G(X_r))\right),$$

for $r = 1, \ldots, n$, and

$$E(\hat{\beta}^{@}_{j,k}) = E\left(\frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(\hat{G}(X_i))(1+\epsilon_i)\right) = E\left(\psi_{j,k}(\hat{G}(X))\right).$$

Thus, $D \to 0$ as $n \to \infty$ because,

$$E\left(\psi_{j,k}(\hat{G}(X))\right) \leq E\left(|\hat{\beta}^{@}_{j,k}|\right) \leq \left(E\left(|\hat{\beta}^{@}_{j,k}|^2\right)\right)^{1/2} \leq \frac{C}{\sqrt{n}},$$

where C is a general positive constant, by Jensen's inequality and Proposition 6 in Kerkyacharian and Picard (2004, p.1072).

For the term A, when the errors are dependent and predictors are IID, if $r = s$, we have that $|A| \leq 2\|f\|_\infty^2 2^{(j+j')/2}\|\psi\|_\infty^2$ and $A = 0$, if $r \neq s$. Thus,

$$n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}[A+B+C+D] = n^{-2}\sum_{r=1}^{n}[A+D]+n^{-2}\sum_{r=1}^{n}\sum_{\substack{s=1\\ s\neq r}}^{n} D \leq n^{-1}[A+D]+\frac{C}{n} = O(n^{-1}).$$

Finally, when the errors are IID and predictors are dependent, we still have $|A| \leq \alpha_{W,|r-s|}^{1/r'} C(f, j, j', p, q, \psi)$ and $n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}[A+B+C+D] \leq O(n^{-1})$.

1.4 Proof of Theorem 1

When the error is strong mixing and the predictor is IID

As given in Section 3, the random variables X_i, $i = 1, 2, \ldots, n$, have all the same known density g, which is compactly supported on the interval $[a, b]$. The respective distribution function $G(x) = \int_a^x g(u)\,du$ is continuous and strictly monotone from $[a, b]$ to $[0, 1]$. Its inverse $G^{-1}(x)$ is also continuous and strictly monotone. These conditions on the distribution function and its inverse imply that $G(G^{-1}(x)) = x$ and $G^{-1}(G(x)) = x$, almost sure for all $x \in [a, b]$. Thus, by Theorem 1 in Kerkyacharian and Picard (2004), the collection formed by $\{\psi_{j,k}(G(\cdot)), j \geq -1, k = 0, \ldots, 2^j-1\}$ satisfies the shrinkage (or unconditional) and the p-Temlyakov properties, as given by Properties 1 and 2 in Kerkyacharian and Picard (2004).

By Theorem 5 in Kerkyacharian and Picard (2004) and Assumption 2, we have that $\nu\{(j,k)\} = \|\psi_{j,k}(G(\cdot))\|_p^p = 2^{jp/2}\omega(I_{j,k})$, where w is the used Muckenhoupt weight. Since ω belongs to L_1 and $2^J \log n/n$ is bounded, then

$$\sum_{j=-1}^{J} 2^{jp/2}\sum_{k=0}^{2^j-1}\omega(I_{j,k})\left(\frac{\log n}{n}\right)^{p/2} < \infty,$$

such that

$$\sup_n \nu\{(j,k) : |k| \leq N2^j\}\left(\frac{\log n}{n}\right)^{p/2} < \infty,$$

for $-1 \leq j \leq J$, where N is the support of ψ.

The results from these previous paragraphs together with the results of our Propostion 2, permit us to apply Theorem 4 in Kerkyacharian and Picard (2004) to obtain Theorem 3, also in Kerkyacharian and Picard (2004), which is the desired result.

When the predictor is strong mixing and the error is IID

The proof follows exactly the same steps of the previous case, since Theorem 4 in Kerkyacharian and Picard (2004) is still valid when the predictor is strong mixing and the error is IID. To see this, suppose that t_n is a sequence of real numbers tending to zero and $\mathcal{J}_n$ is a set of pairs (j,k) such that $\sup_n \nu\{\mathcal{J}_n\} t_n^p < \infty$.

Take $t_n = \sqrt{(\log n)/n}$ and write

$$E\left\|\hat{f}-f\right\|_p^p \leq 2^{p-1}\left[E\left(\left\|\sum_{(j,k)\in\mathcal{J}_n}\left(\hat{\beta}_{j,k} I\left(|\hat{\beta}_{j,k}| > \frac{\kappa t_n}{2}\right) - \beta_{j,k}\right)\psi_{j,k}(G(x))\right\|_p^p + \left\|\sum_{(j,k)\notin\mathcal{J}_n}\beta_{j,k}\psi_{j,k}(G(x))\right\|_p^p\right)\right].$$

Note that

$$\begin{aligned}
&E\left(\left\|\sum_{(j,k)\in\mathcal{J}_n}\left(\hat{\beta}_{j,k} I\left(|\hat{\beta}_{j,k}| > \frac{\kappa t_n}{2}\right) - \beta_{j,k}\right)\psi_{j,k}(G(x))\right\|_p^p\right)\\
&= E\left(\left\|\left(\sum_{(j,k)\in\mathcal{J}_n}(\hat{\beta}_{j,k}-\beta_{j,k}) I\left(|\hat{\beta}_{j,k}| > \frac{\kappa t_n}{2}\right)\right.\right.\right.\\
&\qquad \left.\left.\left. - \beta_{j,k} I\left(|\hat{\beta}_{j,k}| < \frac{\kappa t_n}{2}\right)\right)\psi_{j,k}(G(x))\right\|_p^p\right)\\
&\leq E\left(\left\|\sum_{(j,k)\in\mathcal{J}_n}(\hat{\beta}_{j,k}-\beta_{j,k}) I\left(|\hat{\beta}_{j,k}| > \frac{\kappa t_n}{2}\right)\psi_{j,k}(G(x))\right\|_p^p\right)\\
&\qquad + E\left(\left\|\sum_{(j,k)\in\mathcal{J}_n} I\left(|\hat{\beta}_{j,k}| < \frac{\kappa t_n}{2}\right)\beta_{j,k}\psi_{j,k}(G(x))\right\|_p^p\right)\\
&= A + B
\end{aligned}$$

By the p-Temlyakov property, as given by Property 2 in Kerkyacharian and Picard (2004), $A \leq C(A_1 + A_2)$, where $C > 0$ is a general constant that does not depend on n, and A_1 and A_2 are respectively given by

$$\int_a^b\left(\sum_{(j,k)\in\mathcal{J}_n} I(|\beta_{j,k}| \leq \frac{\kappa t_n}{2})\left|\hat{\beta}_{j,k}-\beta_{j,k}\right|^2 I(\left|\hat{\beta}_{j,k}-\beta_{j,k}\right| > \frac{\kappa t_n}{2})\,|\psi_{jk}(G(x))|^2\right)^{p/2} dG(x)$$

and

$$\int_a^b\left(\sum_{(j,k)\in\mathcal{J}_n} I(|\beta_{j,k}| > \frac{\kappa t_n}{2})\left|\hat{\beta}_{j,k}-\beta_{j,k}\right|^2 I(\left|\hat{\beta}_{j,k}\right| > \frac{\kappa t_n}{2})\,|\psi_{jk}(G(x))|^2\right)^{p/2} dG(x).$$

Again by the p-Temlyakov property, we have that

$$A_1 \leq C \int_a^b \sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|\leq\kappa t_n/2} \left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^p I(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| > \frac{\kappa t_n}{2}) \left\|\psi_{jk}\right\|_\infty^p dG(x).$$

Now applying the Cauchy-Schwartz inequality we have

$$\begin{aligned}
&\int_a^b \left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^p I(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| > \frac{\kappa t_n}{2}) \left\|\psi_{jk}\right\|_\infty^p dG(x) \\
&= E\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^p I(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| > \frac{\kappa t_n}{2}) \left\|\psi_{jk}\right\|_\infty^p\right) \\
&\leq \left(P\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| > \frac{\kappa t_n}{2}\right)\right)^{1/2} \left(E\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right|^{2p}\right)\right)^{1/2} \left\|\psi_{jk}\right\|_\infty^p,
\end{aligned}$$

such that, by the results of Proposition 2 and the definition of $\nu\{(j,k)\}$, we have that

$$A_1 \leq O(n^{-B/\|\psi\|_\infty}) O\left(\left(\frac{\log n}{n}\right)^p\right) 2^{jp/2} \nu\{(j,k) \in \mathcal{J}_n : |\beta_{j,k}| \leq \kappa t_n/2\}.$$

For A_2, similarly to what we did for A_1, we obtain:

$$A_2 \leq O(n^{-B/\|\psi\|_\infty}) O\left(\left(\frac{\log n}{n}\right)^p\right) 2^{jp/2} \nu\{(j,k) \in \mathcal{J}_n : |\beta_{j,k}| > \kappa t_n/2\}.$$

Now we analyze the term B which, again by the p-Temlyakov property, $B \leq C(B_1 + B_2)$, where $C > 0$ is a general constant that does not depend on n, and B_1 and B_2 are respectively given by:

$$\int_a^b \left(\sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2} |\beta_{j,k}|^2 I\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| < \frac{\kappa t_n}{2}\right) |\psi_{jk}(G(x)|^2\right)^{p/2} dG(x)$$

and

$$\int_a^b \left(\sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|\leq\kappa t_n/2} |\beta_{j,k}|^2 I\left(\left|\hat{\beta}_{j,k}\right| < \frac{\kappa t_n}{2}\right) |\psi_{jk}(G(x)|^2\right)^{p/2} dG(x).$$

For $p \leq 2$,

$$\begin{aligned}
B_1 &\leq \int_a^b \left(\sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2} |\beta_{j,k}\psi_{j,k}(G(x))|^2 I\left(\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| \geq \kappa t_n\right)\right)^{p/2} dG(x) \\
&\leq C t_n^{2p} \int_a^b \left(\sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2} |\beta_{j,k}\psi_{j,k}(G(x))|^2\right)^{p/2} dG(x) \\
&\leq C t_n^{2p} \|f\|_p^p.
\end{aligned}$$

In the first line it is used Jensen inequality, in line two we used the Cauchy-Schwartz inequality result from A_1 and then it is applied the shrinkage property, as given by Property 1 in Kerkyacharian and Picard (2004).

For $p \geq 2$, using Minkowski inequality,

$$
\begin{aligned}
B_1 &\leq \int_a^b \left\{ \sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2} \left[E\,|\beta_{j,k}|^p\, I\left(\left|\hat{\beta}_{j,k}-\beta_{j,k}\right| \geq \kappa t_n/2\right) |\psi_{j,k}(G(x))|^p \right]^{2/p} \right\}^{p/2} d\,G(x) \\
&\leq C t_n^{2p} \int_a^b \left\{ \sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2} |\beta_{j,k}\psi_{j,k}(G(x))|^2 \right\}^{p/2} d\,G(x) \\
&\leq C t_n^{2p} \int_a^b \sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2} |\beta_{j,k}\psi_{j,k}(G(x))|^p \, d\,G(x) \\
&\leq C t_n^{2p} \left\| f \right\|_p^p .
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
B_2 &\leq \int_a^b \left\{ \sum_{(j,k)\in\mathcal{J}_n:|\beta_{jk}|\leq\kappa t_n/2} |\beta_{j,k}|^2 \, |\psi_{j,k}(G(x))|^2 \right\}^{p/2} d\,G(x) \\
&\leq K^p \left\| \sum_{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|\leq\kappa t_n/2} \beta_{j,k}\psi_{j,k}(G(x)) \right\|_p^p \\
&\leq K^p \left[\sum_{k=0}^{\infty} \left\| \sum_{(j,k)\in\mathcal{J}_n:2^{-k}\kappa t_n/2\leq|\beta_{j,k}|\leq 2^{-k+1}\kappa t_n} \beta_{j,k}\psi_{j,k}(G(x)) \right\|_p \right]^p \\
&\leq K^p \left[\sum_{k=0}^{\infty} 2^{-k+1}\kappa t_n \left\| \sum_{(j,k)\in\mathcal{J}_n,2^{-k}\kappa t_n/2\leq|\beta_{j,k}|\leq 2^{-k+1}\kappa t_n} \psi_{j,k}(G(x)) \right\|_p \right]^p \\
&\leq K^p \left[\sum_{k=0}^{\infty} 2^{-k+1}\kappa t_n \left\{ \sum_{(j,k)\in\mathcal{J}_n,2^{-k}\kappa t_n/2\leq|\beta_{j,k}|\leq 2^{-k+1}\kappa t_n} \left\| \psi_{j,k}(G(x)) \right\|_p \right\}^{1/p} \right]^p \\
&\leq K^p \left[\sum_{k=0}^{\infty} 2^{-k+1}\kappa t_n \left\{ 2^{-k}\kappa t_n \right\}^{-q/p} \left\| f \right\|_{l_{q,\infty}}^{q/p} \right]^p \\
&\leq C t_n^{p-q} \left\| f \right\|_{l_{q,\infty}}^q .
\end{aligned}
$$

In line 1 it is used the shrinkage property, in line 3, the triangular inequality, in line 4 again the shrinkage property, and in line 5 the p-Temlyakov property.

Summarizing,

$$
\begin{aligned}
E\left\|\hat{f}-f\right\|_p^p &\leq O(n^{-B/\|\psi\|_\infty})O\left(\left(\frac{\log n}{n}\right)^p\right)2^{jp/2}\nu\{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|\leq\kappa t_n/2\} \\
&\quad + O(n^{-B/\|\psi\|_\infty})O\left(\left(\frac{\log n}{n}\right)^p\right)2^{jp/2}\nu\{(j,k)\in\mathcal{J}_n:|\beta_{j,k}|>\kappa t_n/2\} \\
&\quad + Ct_n^{2p}\|f\|_p^p + Ct_n^{p-q}\|f\|_{l_{q,\infty}}^q + \left\|\sum_{(j,k)\notin\mathcal{J}_n}\beta_{j,k}\psi_{j,k}(G(x))\right\|_p^p \\
&\leq t_n^{p-q}A\left(\|f\|_{l_{q,\infty}}^q + t_n^{p+q}\left[\|f\|_p^p + \nu\left\{(j,k)\in\mathcal{J}_n\right\}\right]\right) \\
&\quad + \left\|\sum_{(j,k)\notin\mathcal{J}_n}\beta_{j,k}\psi_{j,k}(G(x))\right\|_p^p,
\end{aligned}
$$

where A is a constant that depends on p, q, k and C. Expressing $p-q$ in terms of s, we have for the last term:

$$
\begin{aligned}
E\left\|\hat{f}-f\right\|_p^p &\leq t_n^{2sp/(1+2s)}A\left(\|f\|_{l_{q,\infty}}^q + t_n^{p+q}\left[\|f\|_p^p + \nu\left\{(j,k)\in\mathcal{J}_n\right\}\right]\right) \\
&\quad + \left\|\sum_{(j,k)\notin\mathcal{J}_n}\beta_{j,k}\psi_{j,k}(G(x))\right\|_p^p.
\end{aligned}
$$

2 Supplementary Material: Simulations results

	SNR=1		SNR=7	
n	Symmlet9	Coiflet3	Symmlet9	Coiflet3
Sine-Uniform				
128	0.0669	0.0624	0.0306	0.0303
256	0.0487	0.0446	0.0221	0.0200
512	0.0404	0.0380	0.0161	0.0147
1024	0.0293	0.0271	0.0116	0.0107
2048	0.0264	0.0255	0.0098	0.0094
Sine-Sine				
128	0.0667	0.0617	0.0313	0.0314
256	0.0482	0.0438	0.0228	0.0206
512	0.0411	0.0394	0.0161	0.0148
1024	0.0295	0.0277	0.0115	0.0108
2048	0.0258	0.0248	0.0098	0.0094
Heavisine-Uniform				
128	0.0709	0.0703	0.0409	0.0451
256	0.0549	0.0511	0.0355	0.0328
512	0.0424	0.0401	0.0223	0.0216
1024	0.0322	0.0301	0.0190	0.0180
2048	0.0274	0.0266	0.0139	0.0137
Heavisine-Sine				
128	0.0712	0.0709	0.0426	0.0474
256	0.0560	0.0534	0.0384	0.0384
512	0.0424	0.0407	0.0222	0.0216
1024	0.0316	0.0304	0.0183	0.0181
2048	0.0265	0.0257	0.0138	0.0135
Doppler-Uniform				
128	0.1305	0.1260	0.1186	0.1158
256	0.1192	0.1242	0.1124	0.1186
512	0.0897	0.0923	0.0834	0.0860
1024	0.0852	0.0909	0.0815	0.0886
2048	0.0614	0.0645	0.0578	0.0601
Doppler-Sine				
128	0.1301	0.1277	0.1156	0.1154
256	0.1211	0.1224	0.1143	0.1165
512	0.1006	0.0974	0.0936	0.0911
1024	0.0970	0.0920	0.0942	0.0898
2048	0.0631	0.0650	0.0591	0.0615

Table 1 Average of 200 root mean square error (RMSE) for the simulation study, with $Y_i = f(X_i) + \epsilon_i$, $\epsilon_i = 0.2\epsilon_{i-1} + u_i$, i.e., ϵ_i is α-mixing, for each pair function-density, sample size n, signal-to-noise ratio (SNR) and wavelets Symmlet9 and Coiflet3.

n	SNR=1		SNR=7	
	Symmlet9	Coiflet3	Symmlet9	Coiflet3
Sine-Uniform				
128	0.0507	0.0477	0.0258	0.0275
256	0.0365	0.0335	0.0195	0.0185
512	0.0296	0.0278	0.0133	0.0126
1024	0.0217	0.0206	0.0099	0.0094
2048	0.0193	0.0186	0.0077	0.0075
Sine-Sine				
128	0.0478	0.0450	0.0251	0.0270
256	0.0358	0.0333	0.0191	0.0187
512	0.0291	0.0276	0.0132	0.0124
1024	0.0216	0.0205	0.0096	0.0094
2048	0.0188	0.0182	0.0075	0.0074
Heavisine-Uniform				
128	0.0694	0.0652	0.0391	0.0369
256	0.0515	0.0488	0.0298	0.0306
512	0.0418	0.0400	0.0222	0.0219
1024	0.0316	0.0311	0.0185	0.0193
2048	0.0275	0.0267	0.0143	0.0143
Heavisine-Sine				
128	0.0685	0.0636	0.0394	0.0372
256	0.0514	0.0490	0.0306	0.0310
512	0.0414	0.0400	0.0224	0.0221
1024	0.0318	0.0309	0.0185	0.0188
2048	0.0270	0.0263	0.0143	0.0143
Doppler-Uniform				
128	0.1284	0.1257	0.1116	0.1145
256	0.1168	0.1230	0.1090	0.1179
512	0.0860	0.0839	0.0774	0.0761
1024	0.0778	0.0823	0.0725	0.0783
2048	0.0551	0.0520	0.0494	0.0462
Doppler-Sine				
128	0.1353	0.1336	0.1196	0.1199
256	0.1252	0.1299	0.1173	0.1239
512	0.0868	0.0840	0.0788	0.0761
1024	0.0837	0.0762	0.0802	0.0723
2048	0.0526	0.0526	0.0465	0.0471

Table 2 Average of 200 root mean square error (RMSE) for the simulation study, with $Y_i = f(X_i) + \epsilon_i$, $X_i = 0.2X_{i-1} + u_i$, i.e., X_i is α-mixing, for each pair function-density, sample size n, signal-to-noise ratio (SNR) and wavelets Symmlet9 and Coiflet3.

3 Supplementary Material: Example

Example of a process satisfying Assumption 1

We provide only one example of processes satisfying Assumption 1. The autoregressive model used for the simulations is considered and we try to make clear the precise effect of the parameters.

Consider the first order autoregressive process given by $\epsilon_i = \theta\epsilon_{i-1} + u_i$, where u_i are IID normally distributed random variables with zero mean and variance σ^2, for $i \in \mathbb{Z}$ and $\theta \in (0, 1)$.

Since the normal density of u_i is bounded, by Remark 2 in Andrews (1983) we have that the Condition K1 of Theorem 2 also in Andrews (1983) holds with $q = 1$. Condition K2 of this theorem is well known for normal densities and, thus, Condition S_ρ of its Theorem 1 is satisfied. Since $E(|u_i|) < \infty$, for all $i \in \mathbb{Z}$, we get $\nu = 1$ at Theorem 1 in Andrews (1983) and

$$\alpha_{\epsilon,h} \leq \begin{cases} 2(C+1)E(|\epsilon_i|)|\theta|^h & \text{for } h \geq h_0; \\ 1 & \text{for } 1 \leq h < h_0; \end{cases}$$

where C is a general positive constant.

Using the expectation of a half-normal random variable, it is easy to see that

$$E(|\epsilon_i|) \leq \sum_{j=0}^{\infty} |\theta|^j E(|u_{i-j}|) = \frac{1}{1-\theta}\sigma\sqrt{2/\pi},$$

such that Assumption 1 will be satisfied if

$$\sum_{h=1}^{\infty} (h+1)^{c-2} |\theta|^{h\delta/(c+\delta)} < \infty,$$

for given $p > 1$ and some chosen vales c and δ, where $c > p$, $c \in 2\mathbb{N} = \{0, 2, 4, 6, \ldots\}$, and $\delta > 0$.

For $1 < p < 2$, it is enough to take $c = 2$ and $\delta = 1$ because

$$\sum_{h=1}^{\infty} (h+1)^{c-2} |\theta|^{h\delta/(c+\delta)} \leq \sum_{h=0}^{\infty} |\theta|^{h/3} = \left(1 - |\theta|^{1/3}\right)^{-1}.$$

For $p \geq 2$, we may use the following identity from, for instance, Mood et al (1974, p.533):

$$(1-x)^t = \sum_{h=0}^{\infty} \binom{t}{h} (-x)^h,$$

for $-1 < x < 1$ and $t \in \mathbb{Z}$.

As an illustration, for $2 \leq p < 4$, take $c = 4$ and $\delta = 1$, such that

$$\begin{aligned} \sum_{h=1}^{\infty} (h+1)^{c-2} |\theta|^{h\delta/(c+\delta)} &\leq \sum_{h=0}^{\infty} (h+1)^2 |\theta|^{h/5} \\ &\leq 2 \frac{\sum_{h=0}^{\infty} (h+2)(h+1) \left(|\theta|^{1/5}\right)^h}{2} \\ &= 2 \sum_{h=0}^{\infty} \binom{-3}{h} (-|\theta|^{1/5})^h \\ &= 2 \left(1 - |\theta|^{1/5}\right)^{-3}. \end{aligned}$$

4 Supplementary Material: Application

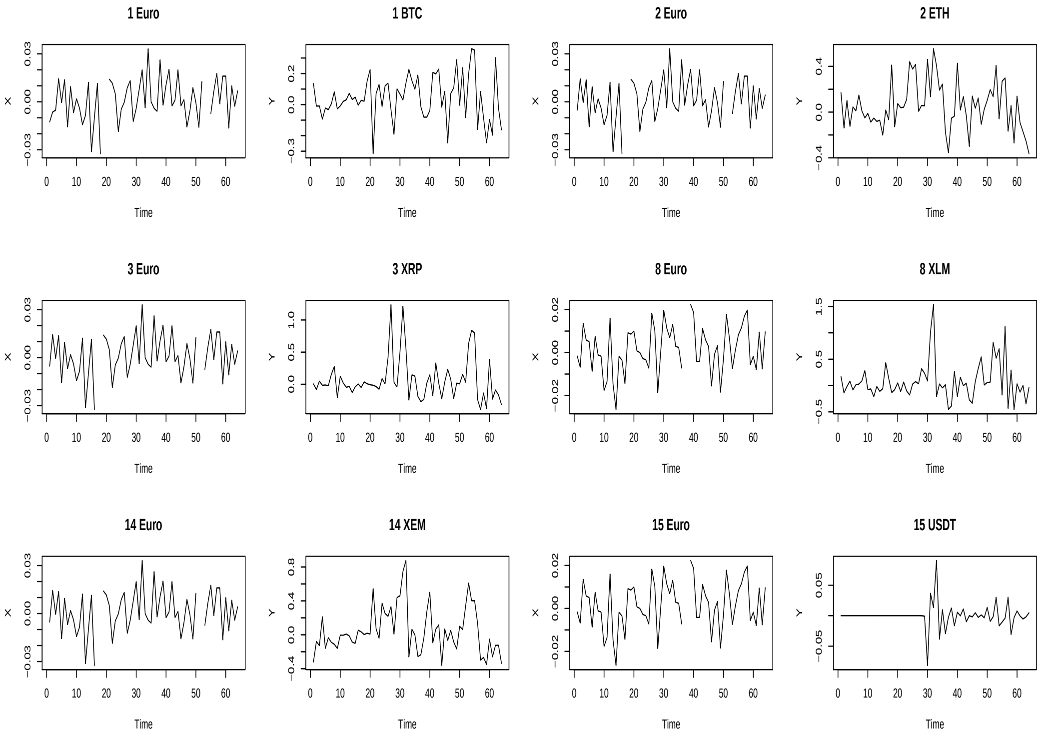

Fig. 1 Time series of 7-days log returns of the Euro and six studied cryptocurrencies. They are plotted in pairs, over the matching period of the Euro and each studied cryptocurrency.

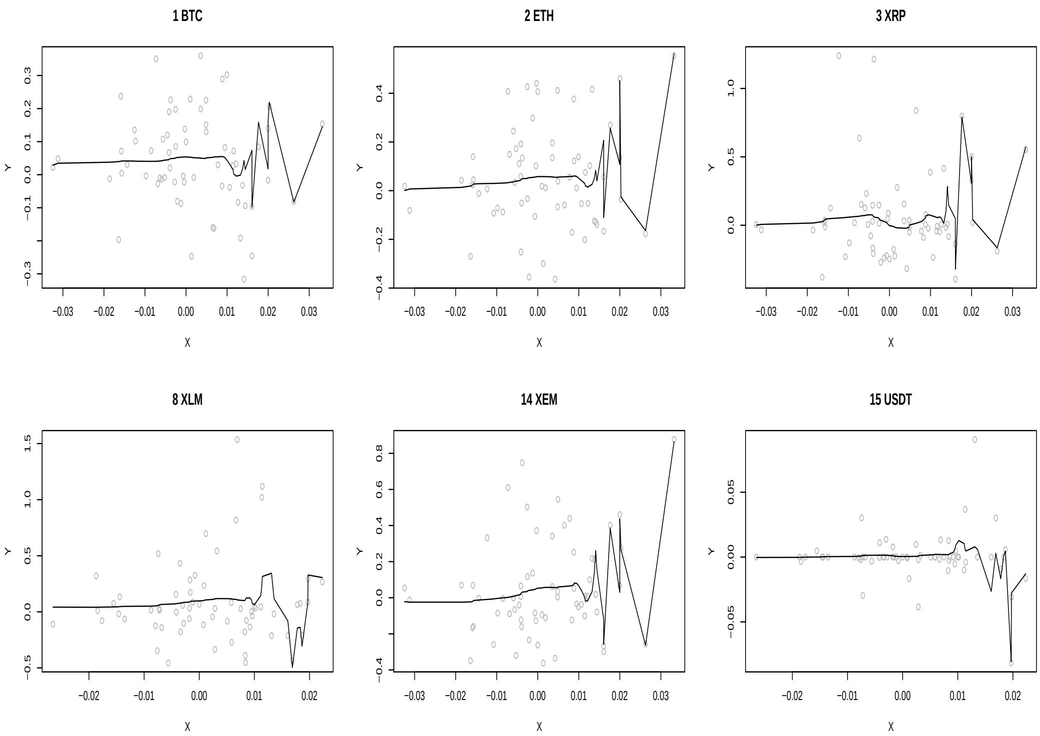

Fig. 2 Estimates of the functions f (full curves), using model (15) for the six cryptocurrencies log return (Y) as a function of the Euro log return (X). Pairs of observations X and Y are plotted as gray circles.

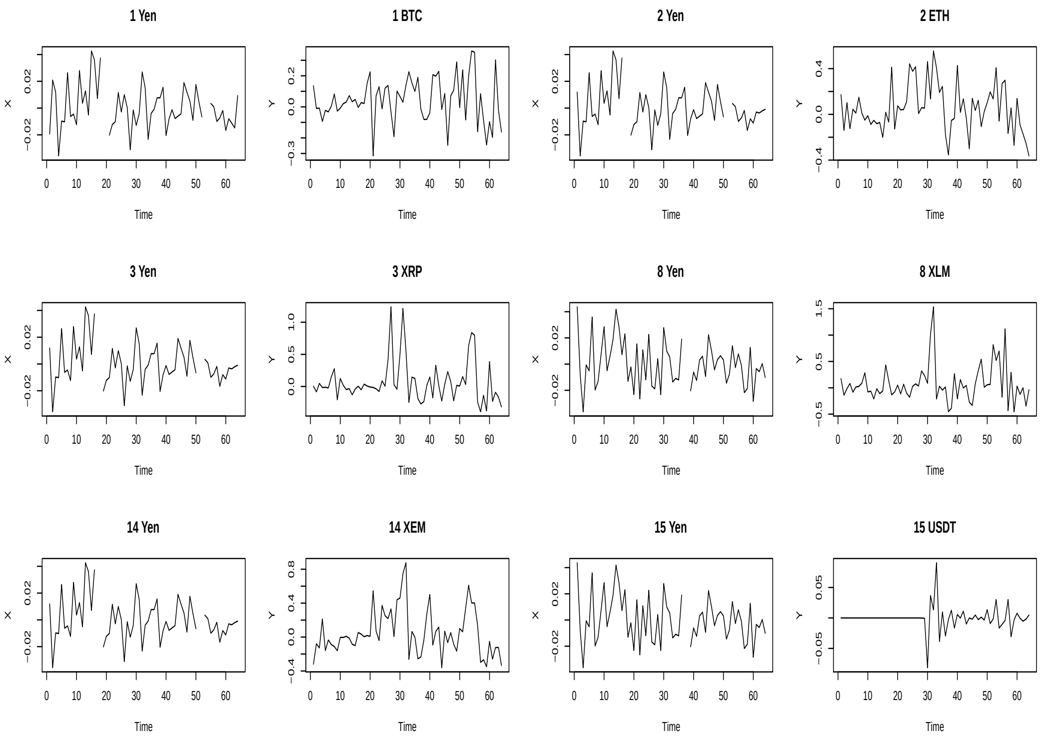

Fig. 3 Time series of 7-days log returns of the Japanese Yen and six studied cryptocurrencies. They are plotted in pairs, over the matching period of the Japanese Yen and each studied cryptocurrency.

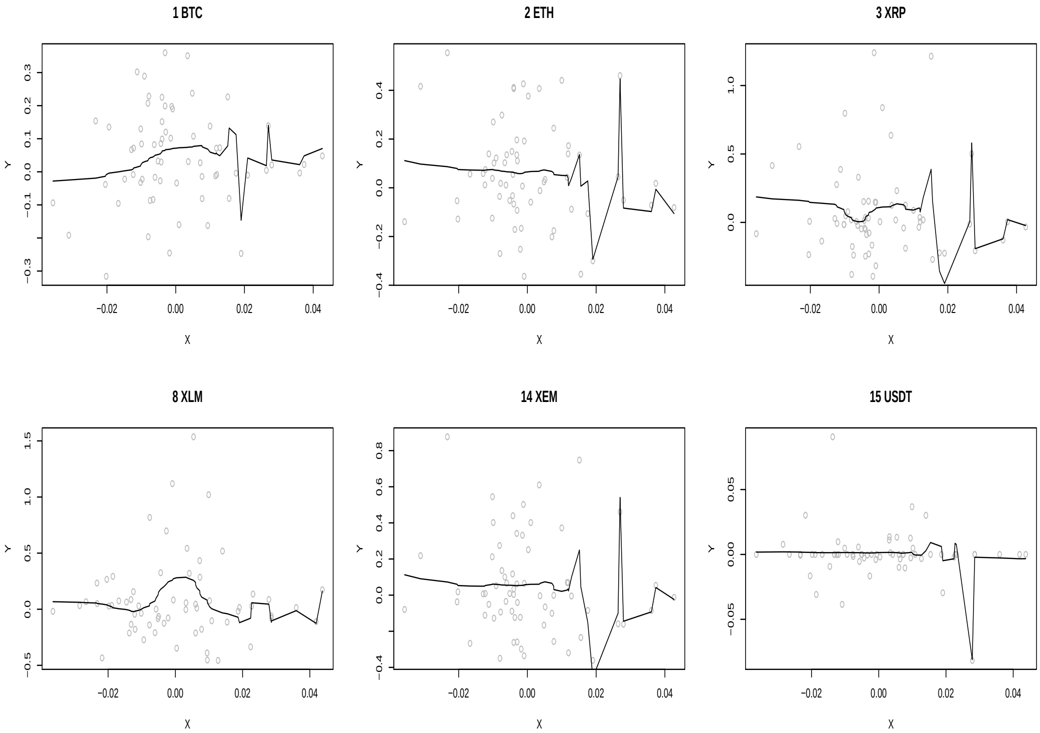

Fig. 4 Estimates of the functions f (full curves), using model (15) for the six cryptocurrencies log return (Y) as a function of the Japanese Yen log return (X). Pairs of observations X and Y are plotted as gray circles.

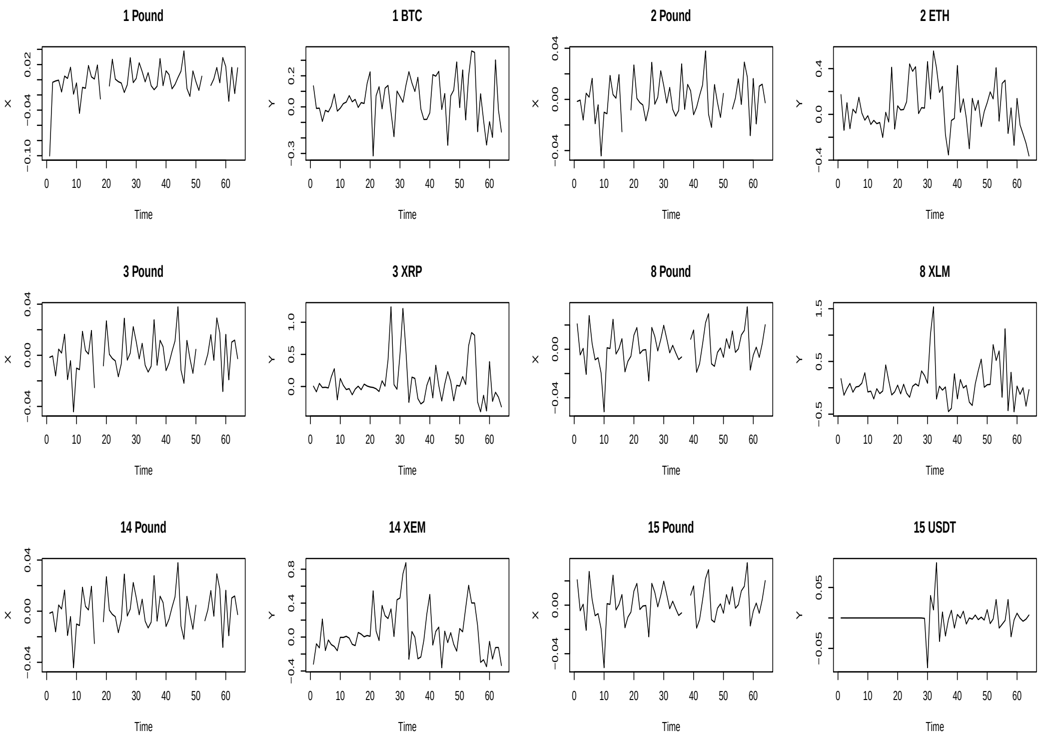

Fig. 5 Time series of 7-days log returns of the British Pound and six studied cryptocurrencies. They are plotted in pairs, over the matching period of the British Pound and each studied cryptocurrency.

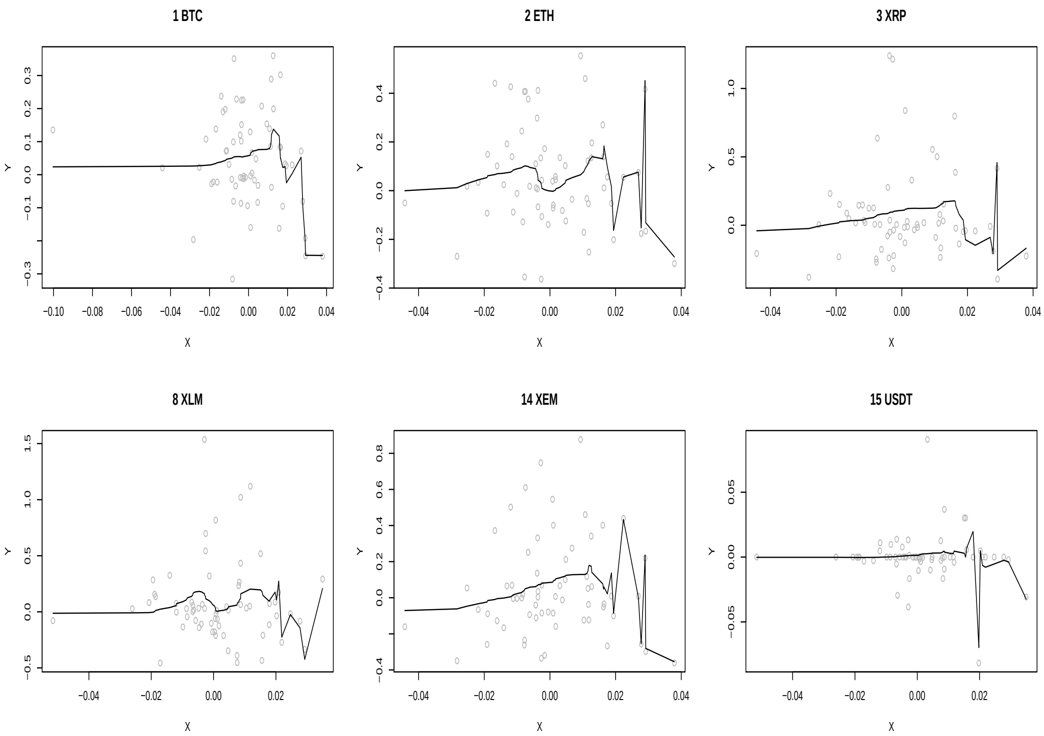

Fig. 6 Estimates of the functions f (full curves), using model (15) for the six cryptocurrencies log return (Y) as a function of the Bristish Pound log return (X). Pairs of observations X and Y are plotted as gray circles.

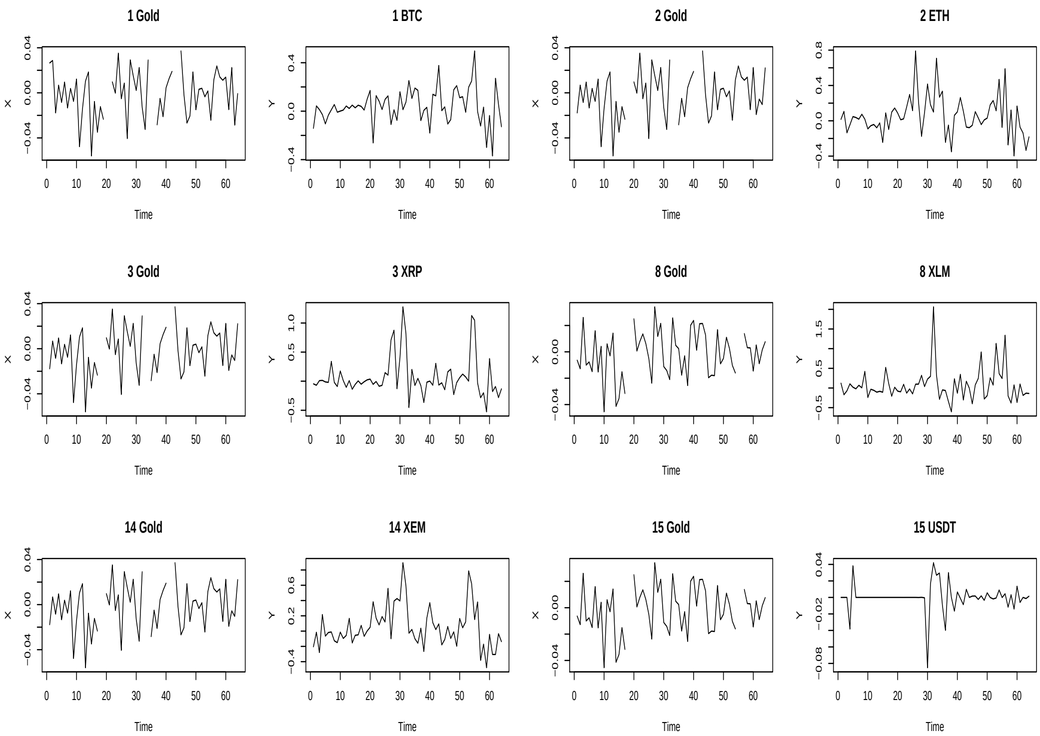

Fig. 7 Time series of 7-days log returns of gold and six studied cryptocurrencies. They are plotted in pairs, over the matching period of gold and each studied cryptocurrency.

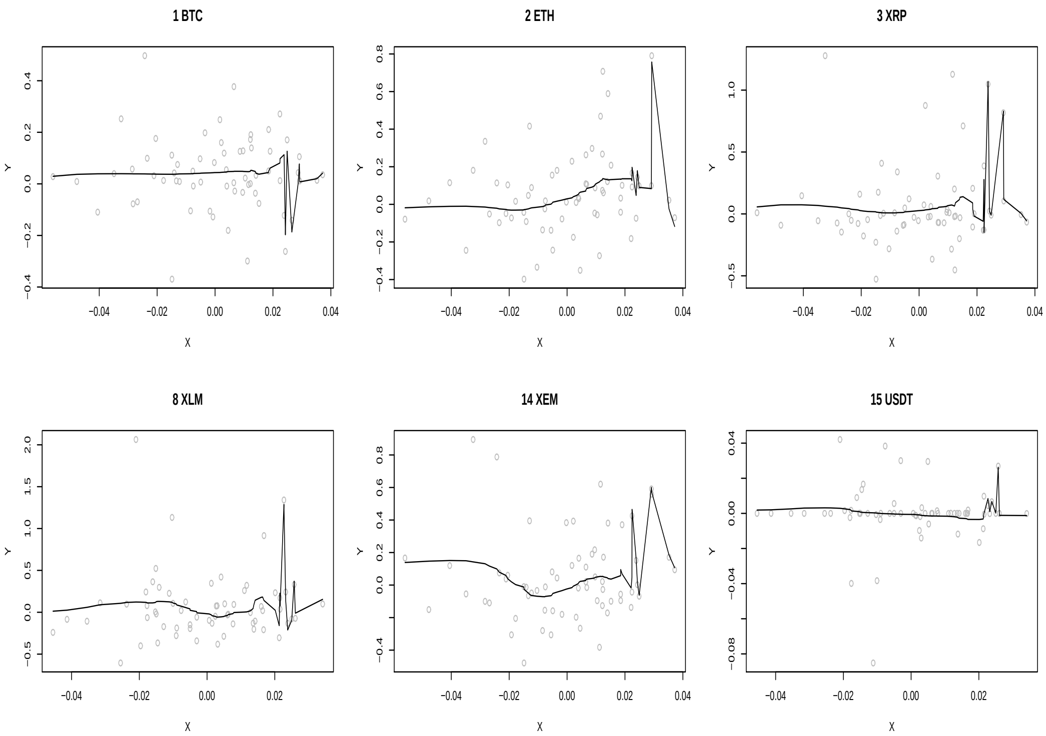

Fig. 8 Estimates of the functions f (full curves), using model (15) for the six cryptocurrencies log return (Y) as a function of gold log return (X). Pairs of observations X and Y are plotted as gray circles.

References

Andrews DWK (1983) First order autoregressive processes and strong mixing. Tech. Rep. 664, Cowles Foundation for Research in Economics at Yale University, New Haven, Connecticut

Doukhan P (1994) *Mixing: Properties and Examples*, Lecture Notes in Statistics, vol 85. Springer-Verlag, New York

Doukhan P, Louhichi S (1999) A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* 84(2):313–342

Härdle W, Kerkyacharian G, Picard D, Tsybakov A (1998) *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statistics, vol 129. Springer, New York

Kerkyacharian G, Picard D (2004) Regression in random design and warped wavelets. *Bernoulli* 10(6):1053–1105

Mood AM, Graybill FA, Boes DC (1974) *Introduction to the Theory of Statistics*, 3rd edn. McGraw-Hill, Singapore