PUB. IRMA, LILLE 2010
Vol. 70, N° VI

Estimating and forecasting partially linear models with periodic covariances*

Xavier Brossat[1] Georges Oppenheim[2] and Marie-Claude Viano[3]

[1] Département OSIRIS – Service de Recherche et Développement – EDF
1, av. du Général de Gaulle F-92140 Clamart
[2] Laboratoire d'Analyse et de Mathématiques Appliquées Université Paris-Est
5, bd. Descartes, Champs-sur-Marne F-77454 Marne-la-Vallée Cedex 2
[3] Laboratoire Paul Painlevé UMR CNRS 8524 – Bât M2.
Université Lille 1, Villeneuve d'Ascq, F-59655 Cedex

Article déposé le 8 février 2010

Résumé

This paper presents a backfitting-type method to estimate and forecast a periodically correlated partially linear model with exogeneous variables and heteroskedastic input noise. A rate of convergence of the estimator is given. The results are valid even if the period is unknown.

Résumé

On utilise une procédure itérative de type backfitting pour estimer les paramètres d'une classe de modèles partiellement linéaires périodiquement corrélés présentés pour modéliser l'évolution de la consommation d'électricité. On obtient une vitesse de convergence des estimateurs et un intervalle de prévision pour la consommation.

MSC 2000 subject classifications. 62E20, 62H12, 62G20, 62J02, 62M10, 62M20.

Key words and phrases. α-mixing, additive models, backfitting, electricity consumption, semiparametric regression smoothing.

*Preprint

1 Introduction

In this paper, we focus on partially linear models of type

$$X_n = \sum_{j=1}^{p} a_j X_{n-j} + \sum_{j=0}^{q} b_j(e_{n-j}) + \sigma(e_n, \ldots, e_{n-q'})\varepsilon_n. \qquad (1)$$

The parameters $p \geq 1$ and $q \geq 0$ are supposed to be known while the coefficients a_j as well as the functions b_j and σ are unknown. The sequence (ε_n) is an unobserved system noise. The aim is to predict X_{n+h}, for some $h \geq 1$, from $((X_n, e_n), (X_{n-1}, e_{n-1}), \ldots)$, the observed past set of values available at date n.

During the last 20 years, partially linear autoregressive models such as (1) have gained attention, as being a good compromise between linear and purely non parametric ones. Such models, first introduced in [5] to represent the relation between weather and electricity consumption are now widely used in the literature. See for example [14] where a chapter is devoted to models like (1). The functions b_j are expanded on a convenient basis and the first coefficients of this expansion, together with the a_j's, are estimated via a L.M.S method. With the same type of partially linear models [6, 10] use wavelets in the estimation scheme. In [2], the b_j's are treated as nuisance parameters. Let us also mention [11, 12], devoted to pure autoregressive models where some past values operate in a linear form and the others in a functional one.

In model (1), the linear character of the autoregressive part presents several advantages. First, it reduces the so called curse of dimensionality. Second, linear autoregression is preserved when expressing the future values $(X_{n+h})_{h=1,\ldots}$ from the past $(X_{n-h}, e_{n-h})_{h=0,\ldots}$, which makes easier, and in some sense coherent, forecasting at lags greater than 1. Last, model (1) is specially well fitted to the situation where the output X_n is the electricity consumption at date n and the input e_n the temperature at the same date, since it is well-known that the effect of temperature on electricity sales is highly non-linear on the domains of extreme temperatures. Notice that, in practical situations, the temperature at date n is either measured or forecasted by Météo-France. In both cases, the value of the exogeneous variable e_n is known.

For simplicity and convenience, we only consider in this paper the situation $q = q' = 0$ given by

$$X_n = a_1 X_{n-1} \ldots + a_p X_{n-p} + b(e_n) + \sigma(e_n)\varepsilon_n, \qquad n \in \mathbb{Z}. \qquad (2)$$

The algorithm presented below can easily be adapted to the general case $q, q' > 0$, and the results of theorem 6 still hold if $q' = 0$, the case $q' > 0$ leading to a loss of speed if σ has not an additive form.

1.1 First hypotheses

We adopt the following basic hypotheses $(\mathcal{H})$

- $\mathcal{H}_1$: Periodicity. The exogeneous sequence (e_n) is the sum of a periodic deterministic sequence (s_n) and a zero-mean strong white noise

$$e_n = s_n + \eta_n \quad \forall n \tag{3}$$

- $\mathcal{H}_2$: Whiteness of the system noise. (ε_n) is an i.i.d sequence of zero-mean variables, and $\mathrm{Var}(\varepsilon_n) = 1$.
- $\mathcal{H}_3$: Stability. The autoregressive dynamic is stable. In other words, the polynomial

$$A(z) = z^p - \left(\sum_{j=1}^{p} a_j z^{p-j}\right)$$

does not vanish on the domain $|z| \geq 1$.
- $\mathcal{H}_4$: Independence of the inputs. The two sequences (ε_n) and (η_n) are independent.

Keeping in mind the example of electricity consumption, the first hypothesis $\mathcal{H}_1$ allows some periodicity in the random structure of the input sequence (e_n).

1.2 Backfitting-type estimation

In this paper, we chose an iterative estimation procedure. Backfitting methods, first proposed by [1], are usually recommended for additive models which involve several explanatory variables, each with unknown functional form. The method is well described in [15]. See also [7, 20, 21] where the estimation algorithms use local polynomial regression. See also [19] based on projections on polynomial spaces. The qualities of backfitting procedures when autoregression is involved are less well understood. In [26], for the non linear stationary autoregressive model with exogeneous variables

$$X_n = a(X_{n-1}) + b(e_n) + \varepsilon_n,$$

the algorithm works in two steps: the first step builds a preliminary estimator of a et b by piecewise constant functions. Then, from the obtained pseudo remainders, the second step builds kernel estimators of the same functiions. The author obtains the limit law for the estimation error.

In the present paper, the choice of a backfitting method presents also the advantage of allowing the period of the input sequence (e_n) to remain unknown. In fact, simulation studies in progress seem to indicate that the method still works when the period shows slight variations.

2 Estimation of the parametric and non parametric components

The aim is to estimate the functions $b(.)$ and $\sigma(.)$ and the vector parameter

$$\theta = {}^t(a_1, \ldots, a_p).$$

Denoting

$$\phi_k = {}^t(X_{k-1}, \ldots, X_{k-p}),$$

the model can be written

$$X_n = {}^t\phi_n\theta + b(e_n) + \sigma(e_n)\varepsilon_n. \tag{4}$$

We chose a kernel K, and a smoothing parameter h_n.

Starting from an initialisation, and having chosen a stopping rule, the iterative method consists in estimating θ (resp. b) by using an estimation of the residual calculated from the previous estimation of b (resp. θ).

- Initialisation. Fix the first value $\hat{\theta}^{(1)}$
- Step 1. Estimate the function b by a kernel estimator based on the partial residuals of the backfitting

$$\hat{b}_n^{(1)}(e) = \frac{\sum_{l=p+1}^{n} \left(X_l - {}^t\phi_l\hat{\theta}^{(1)}\right) K_n\left(e - e_l\right)}{\sum_{l=p+1}^{n-1} K_n\left(e - e_l\right)}$$

where

$$K_n(e) := K\left(\frac{e}{h_n}\right).$$

- Step 2. Update the estimation of θ by a least mean square estimator based on the new partial residuals of the backfitting

$$\begin{aligned}\hat{\theta}_n^{(2)} &= \text{Argmin}_\theta \sum_{l=p+1}^{n} (X_l -{}^t \phi_l\theta - \hat{b}_n^{(1)}(e_l))^2 \\ &= \Sigma_n^{-1} \sum_{l=p+1}^{n-1} \phi_l(X_l - \hat{b}_n^{(1)}(e_l))\end{aligned}$$

with

$$\Sigma_n = \sum_{l=p+1}^{n} \phi_l^t\phi_l. \tag{5}$$

Finally, transition from step $k-1$ to step k writes

$$\hat{b}_n^{(k-1)}(e) = \frac{\sum_{l=p+1}^{n} \left(X_l -{}^t \phi_l\hat{\theta}^{(k-1)}\right) K_n\left(e - e_l\right)}{\sum_{l=p+1}^{n-1} K_n\left(e - e_l\right)} \tag{6}$$

$$\hat{\theta}_n^{(k)} = \Sigma_n^{-1} \sum_{l=p+1}^{n} \phi_l(X_l - \hat{b}_n^{(k-1)}(e_l)). \tag{7}$$

- Chosing a stopping time k for the iterations, the variance $\sigma^2(e)$ is then estimated by a kernel method using the partial residuals based on the estimates $\hat{\theta}_n^{(k)}$ and $\hat{b}_n^{(k-1)}$

$$\hat{\sigma}_{n,k}^2(e) = \frac{\sum_{l=p+1}^{n-1} \left(X_l -{}^t \phi_l\hat{\theta}_n^{(k)} - \hat{b}_n^{(k-1)}(e_l)\right)^2 K_n\left(e - e_l\right)}{\sum_{l=p+1}^{n-1} K_n\left(e - e_l\right)} \tag{8}$$

3 Main results

3.1 More about Hypotheses

The results below shall be proved under hypotheses $\mathcal{H}$, and the following extra assumptions:

- $\mathcal{H}_5$: On the input sequences. The variables ε_1 and η_1 are bounded. Both have a density.
- $\mathcal{H}_6$: On the densities. The density f of η_1 is continuous and non vanishing on the support $[-m_\eta, m_\eta]$ of η_1. The density g of ε_1 is C_1 and never vanishes on the support $[-m_\varepsilon, m_\varepsilon]$ of ε_1.
- $\mathcal{H}_7$: On the functions. Let $\mathcal{E} = \cup_{j=1}^{T}[s_j - m_\eta, s_j + m_\eta]$ denote the union of the T compact supports of the variables e_j.

  1. The function b is γ-Hölderian on $\mathcal{E}$, for some $0 < \gamma \leq 1$, which means that
  $$\sup_{e_1, e_2 \in \mathcal{E}} \frac{|b(e_1) - b(e_2)|}{|e_1 - e_2|^\gamma} < \infty \tag{9}$$
  2. The variance $\sigma^2(e)$ of the input noise is γ_1-Hölderian on $\mathcal{E}$, for some $0 < \gamma_1 \leq 1$, and
  $$\inf_{e \in \mathcal{E}} \sigma(e) > 0. \tag{10}$$

- $\mathcal{H}_8$: On the kernel. The kernel K is lipschitzian, and satisfies
  $$\int K(u)du = 1$$
  The kernel is not necessarily positive. See for example Lemma 9, where positivity is exluded when $\ell \geq 2$. As will be proved, the denominator in (6) and (8), when conveniently normalized, converges to a stritly positive value.

3.2 Preliminaries about the process (X_n) and its covariances

In what follows, we work on the solution of (2) given by the $MA\infty$ expansion

$$X_n = \sum_{j \geq 0} g_j \left(b(e_{n-j}) + \sigma(e_{n-j})\varepsilon_{n-j}\right) \qquad n \in \mathbb{Z} \tag{11}$$

where the geometrically vanishing sequence (g_j) is defined by

$$\frac{1}{1 - a_1 z - \ldots - a_p z^p} = \sum_{j \geq 0} g_j z^j$$

3.2.1 Periodicity

Let $n|T$ denote the remainder obtained when n is divided by T. Since the sequence (s_n) is T-periodic, only the $2T$ stationary sequences $b_{n|T}(\eta_n)$ and $\sigma_{n|T}(\eta_n)$ defined by

$$\begin{aligned} b_{n|T}(\eta_n) &:= b(e_n) = b(s_n + \eta_n) \\ \sigma_{n|T}(\eta_n) &:= \sigma(e_n) = \sigma(s_n + \eta_n) \end{aligned}$$

appear in the expansion (11), which can be rewritten, with $\xi_n = (\eta_n, \varepsilon_n)$ and $H_{n|T}(\xi_n) = b_{n|T}(\eta_n) + \sigma_{n|T}(\eta_n)\varepsilon_n$, as

$$X_n = \sum_{j\geq 0} g_j H_{n-j|T}(\xi_{n-j}). \tag{12}$$

by unfolding T successive random values $X_{kT}, \ldots, X_{(k+1)T-1}$, we notice that

$$\begin{aligned} X_{kT} &= \sum_{j=0}^{\infty} g_{jT} H_0(\xi_{(k-j)T}) + \sum_{j=0}^{\infty} g_{jT+1} H_{T-1}(\xi_{(k-j)T-1}) + \ldots + \\ &+ \sum_{j=0}^{\infty} g_{(j+1)T-1} H_1(\xi_{(k-j-1)T+1}) \\ X_{kT+1} &= \sum_{j=0}^{\infty} g_{jT} H_1(\xi_{(k-j)T+1}) + \sum_{j=0}^{\infty} g_{jT+1} H_0(\xi_{(k-j)T}) + \ldots + \\ &+ \sum_{j=0}^{\infty} g_{(j+1)T-1} H_2(\xi_{(k-j-1)T+2}) \\ &\vdots \\ X_{(k+1)T-1} &= \sum_{j=0}^{\infty} g_{jT} H_{T-1}(\xi_{(k+1-j)T-1}) + \sum_{j=0}^{\infty} g_{jT+1} H_{T-2}(\xi_{(k+1-j)T-2}) \\ &+ \ldots + \sum_{j=0}^{\infty} g_{(j+1)T-1} H_0(\xi_{(k-j)T}) \end{aligned}$$

The T-dimensional vector sequence $Z_k = {}^t\left(X_{kT}, \ldots, X_{(k+1)T-1}\right)$ is a strictly stationary process, each coordinate being the sum of T linear scalar processes based on T independent white noises. In other words, the process (X_n) is periodically correlated (see for example [18] for a review on periodically correlated time series).

Hereafter, we use the stationarity of Z_k mainly to prove convergence results via the law of large numbers.

3.2.2 Covariance matrix

The lemma below proves that, almost surely, the matrix $\Sigma_n = \sum_{l=p+1}^{n} \phi_l^t \phi_l$ appearing in (5) and used in estimating the parameter θ is invertible at least for large enough n. Denoting

$$\phi_k^{(l)} = {}^t(X_{kT+l-1}, X_{kT+l-2}, \ldots, X_{kT+l-p}), \tag{13}$$

Lemma 1. *Under the hypotheses $\mathcal{H}_{1,2,3,4}$, the matrix Σ_n being defined in (5), as $n \to \infty$,*

(i)

$$\frac{\Sigma_n}{n} \xrightarrow{a.s.} M = \frac{1}{T}\sum_{l=0}^{T-1} \mathbb{E}\left(\phi_0^{(l)\,t}\phi_0^{(l)}\right) = \frac{1}{T}\sum_{l=0}^{T-1}\left[\mu^{(l)\,t}\mu^{(l)} + \Gamma^{(l)}\right]$$

where $\mu^{(l)} = \mathbb{E}(\phi_0^{(l)})$ and $\Gamma^{(l)}$ is the covariance matrix of $\phi_0^{(l)}$.

(ii) The limit matrix M is regular.

The proof is in the Appendix.

3.3 Convergence results

3.3.1 Analysis of estimation errors

We first focus on the error in the parameter estimator and on the function estimator,

$$\tilde{\theta}_n^{(k)} = \theta - \hat{\theta}_n^{(k)}.$$

and

$$\tilde{b}_n^{(k-1)}(e) = b(e) - \hat{b}_n^{(k-1)}(e)$$

From (4),

$$\tilde{\theta}_n^{(k)} = -\Sigma_n^{-1}\sum_{l=p+1}^{n} \phi_l\left(\tilde{b}_n^{(k-1)}(e_l) + \sigma(e_l)\varepsilon_l\right) \tag{14}$$

$$\tilde{b}_n^{(k-1)}(e) = \frac{\sum_{l=p+1}^{n}\left(-{}^t\phi_l\tilde{\theta}_n^{(k-1)} + b(e) - b(e_l) - \sigma(e_l)\varepsilon_l\right)K_n(e-e_l)}{\sum_{l=p+1}^{n}K_n(e-e_l)} \tag{15}$$

leading to

$$\tilde{\theta}_n^{(k)} = -\Sigma_n^{-1}\sum_{l=p+1}^{n}\phi_l \times$$
$$\times\left[\sigma(e_l)\varepsilon_l + \frac{\sum_{j=p+1}^{n}\left(-{}^t\phi_j\tilde{\theta}_n^{(k-1)} + b(e_l) - b(e_j) - \sigma(e_j)\varepsilon_j\right)K_n(e_l-e_j)}{\sum_{j=p+1}^{n}K_n(e_l-e_j)}\right]$$

easily rewritten as

$$\tilde{\theta}_n^{(k)} = A_n\tilde{\theta}_n^{(k-1)} + R_n^{(1)} + R_n^{(2)} \tag{16}$$

where

$$\begin{aligned} A_n &= \Sigma_n^{-1}\left(\sum_{l=p+1}^{n}\phi_l\frac{\sum_{j=p+1}^{n}{}^t\phi_j K_n\left(e_l-e_j\right)}{\sum_{j=p+1}^{n}K_n\left(e_l-e_j\right)}\right)\\ R_n^{(1)} &= \Sigma_n^{-1}\sum_{l=p+1}^{n}\phi_l\frac{\sum_{j=p+1}^{n}\left(b(e_j)-b(e_l)+\sigma(e_j)\varepsilon_j\right)K_n\left(e_l-e_j\right)}{\sum_{j=p+1}^{n}K_n\left(e_l-e_j\right)}\\ R_n^{(2)} &= \Sigma_n^{-1}\sum_{l=p+1}^{n}\phi_l\sigma(e_l)\varepsilon_l \end{aligned}$$

We are going to prove that, as $n\to\infty$ the matrix operator A_n converges to a contractant one, and that the remainder terms $R_n^{(1)}$ and $R_n^{(2)}$ tend to zero.

3.3.2 Convergence of $R_n^{(1)}$

Lemma 2. *Under the assumptions $\mathcal{H}_{1,\ldots,8}$, and if the smoothing parameter h_n is such that as $n\to\infty$,*

$$h_n\sim n^{\beta_1}(\ln n)^{\beta_2},\ \textit{with some } \beta_1<0,$$

we have

$$R_n^{(1)}=O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right)+O(h_n^\gamma)$$

Proof. Given the convergence of Σ_n/n to a regular matrix, it is enough to prove the wanted result for

$$\frac{1}{n}\sum_{l=p+1}^{n}\phi_l\frac{\sum_{j=p+1}^{n}\left(b(e_j)-b(e_l)+\sigma(e_j)\varepsilon_j\right)K_n\left(e_l-e_j\right)}{\sum_{j=p+1}^{n}K_n\left(e_l-e_j\right)}. \tag{17}$$

We prove in the Appendix the uniform convergence

$$\sup_e\left|\frac{\sum_{j=p+1}^{n}\left(b(e_j)-b(e)+\sigma(e_j)\varepsilon_j\right)K_n\left(e-e_j\right)}{\sum_{j=p+1}^{n}K_n\left(e-e_j\right)}\right|=O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right)+O(h_n^\gamma). \tag{18}$$

The result then follows from the fact that, thanks to the law of large numbers applied to each subsequence $(\phi_k^{(l)})_k$, $(k=0,\ldots,T-1)$, the arithmetic mean $n^{-1}\sum_{p+1}^{n}\phi_k$ almost surely converges. □

3.3.3 Convergence of $R_n^{(2)}$

The vector sequence $\phi_k\sigma(e_k)\varepsilon_k$ is a martingale difference sequence since $\mathbb{E}(\varepsilon_k)=0$ and since $\phi_k e_k$ and ε_k are independent. Moreover,

$$\mathbb{E}(\|\phi_k\sigma(e_k)\varepsilon_k\|_2^2)=\sigma^2\mathbb{E}\left(b(e_k)^2\right)\mathbb{E}(\|\phi_k\|_2^2).$$

where $\mathbb{E}(\|\phi_k\|_2^2)$ and $\mathbb{E}\left(b(e_k)^2\right)$ are periodic. Hence, for every $\beta > 1/2$

$$\sum \frac{\mathbb{E}(\|\phi_k \sigma(e_k)\varepsilon_k\|_2^2)}{k^{2\beta}} < \infty,$$

implying, from theorem 3.3.1 of [25],

$$n^{-\beta} \sum_{p+1}^{n} \phi_k \sigma(e_k)\varepsilon_k \xrightarrow{a.s.} 0.$$

Finally, using the convergence of Σ_n/n we obtain

Lemma 3. *Under the assumptions $\mathcal{H}_{1,2,3,4}$,*

$$R_n^{(2)} = o_{as}\left(n^{\gamma}\right) \quad \forall \gamma > -1/2$$

3.3.4 Convergence of the coefficient A_n

We prove the convergence of A_n, the matrix coefficient of $\tilde{\theta}_n^{(k-1)}$ in (16).

Lemma 4. *Under the assumptions $\mathcal{H}_{1,\dots,8}$,*
(i) As $n \to \infty$,

$$\begin{aligned} A_n \quad &= \quad \Sigma_n^{-1} \left(\sum_{l=p+1}^{n} \phi_l \frac{\sum_{j=p+1}^{n} {}^t\phi_j K_n(e_l - e_j)}{\sum_{j=p+1}^{n} K_n(e_l - e_j)} \right) \\ &\xrightarrow{a.s.} \quad M^{-1} \sum_{l,j=0}^{T-1} \mu^{(l)\,t}\mu^{(j)} \int \frac{f(u-s_j)f(u-s_l)}{\sum_{i=0}^{T-1} f(u-s_i)} du =: A \end{aligned}$$

where M is defined in Lemma 1.
(ii) Moreover $\|A_n - A\| = 0_{as}\left(\sqrt{\frac{\ln n}{n h_n}}\right) + O(h_n^{\gamma})$.

Lemma 4, together with Lemma 5 below, shows that the passage (16) from step $k-1$ to step k is a fixed point iteration, at least for n large enough.

Lemma 5. *There exists $k_0 \geq 1$ such that*

$$\sup_{v} \frac{\|A^{k_0} v\|_2}{\|v\|_2} < 1. \tag{19}$$

Moreover, $k_0 = 1$ when $p = 1$.

This lemma is proved in the Appendix.

3.3.5 Summing up

Gathering the above partial results easily leads to the convergence of $\hat{\theta}_n^{(k)}$ and of $\hat{b}_n^{(k)}(e)$ where k is the number of iterations and n the sample size, leading to

Theorem 6. *With the assumptions of section 3.1, if the smoothing parameter is such that, as $n \to \infty$ $h_n \sim n^{\beta_1}(\ln n)^{\beta_2}$ and if the kernel K is Lipschitzian, with $\int K(e)de = 1$, there exists $\beta \in]0,1[$ such that*

$$\left.\begin{array}{l} \|\hat{\theta}_n^{(k)} - \theta\|_2 \\ \sup_{e\in\mathcal{E}} |\hat{b}_n^{(k)}(e) - b(e)| \end{array}\right\} = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O_{as}(h_n^{\gamma}) + O_{as}(\beta^k)$$

and

$$\sup_{e\in\mathcal{E}} |\hat{\sigma}^2_{n,k}(e) - \sigma^2(e)| = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O_{as}(h_n^{\min\{\gamma,\gamma'\}}) + O_{as}(\beta^k) \qquad (20)$$

where the $0(.)$'s are uniform with respect to k and n.

We see that the rate of convergence of $\hat{\sigma}^2_{n,k}(e)$ cannot exceed the rate of the other parameters and can even be slower when $b(e)$ is smoother than $\sigma(e)$. Equality (20) is proved in the Appendix.

As a result, an optimal rate is obtained by chosing convenient values for β_1 and β_2.

Corollary 7. *Under the same hypotheses, if $h_n \sim (\ln n/n)^{\frac{1}{2\gamma+1}}$, there exists $\beta \in]0,1[$ such that*

$$\left.\begin{array}{l} \|\hat{\theta}_n^{(k)} - \theta\|_2 \\ \sup_{e\in\mathcal{E}} |\hat{b}_n^{(k)}(e) - b(e)| \end{array}\right\} = O_{as}\left(\frac{\ln n}{n}\right)^{\frac{\gamma}{2\gamma+1}} + O_{as}(\beta^k)$$

and

$$\sup_{e\in\mathcal{E}} |\hat{\sigma}^2_{n,k}(e) - \sigma^2(e)| = O_{as}\left(\frac{\ln n}{n}\right)^{\frac{\min\{\gamma,\gamma'\}}{2\gamma+1}} + O_{as}(\beta^k)$$

Remark 1. In Corollary 7, it is clear that, provided β is not too close to 1, the convergence to zero of the term β^k is fast. In other words, stabilisation of the iterations is easily obtained while convergence to zero of $\left(\frac{\ln n}{n}\right)^{\frac{\gamma}{2\gamma+1}}$ requires large sample size. More precisely, taking $k = k(n) \geq C \ln n$ gives

Corollary 8. *Under the same hypotheses as in Corollary 7 and with $h_n \sim (\ln n/n)^{\frac{1}{2\gamma+1}}$, if the recursive scheme stops after $k(n) \geq C \ln n$ iterations*

$$\left.\begin{array}{l} \|\hat{\theta}_n^{(k(n))} - \theta\|_2 \\ \sup_{e\in\mathcal{E}} |\hat{b}_n^{(k(n))}(e) - b(e)| \end{array}\right\} = O_{as}\left(\frac{\ln n}{n}\right)^{\frac{\gamma}{2\gamma+1}}$$

and

$$\sup_{e\in\mathcal{E}} |\hat{\sigma}^2_{n,k(n)}(e) - \sigma^2(e)| = O_{as}\left(\frac{\ln n}{n}\right)^{\frac{\min\{\gamma,\gamma'\}}{2\gamma+1}}$$

Remark 2. With the above remark in mind, it is interesting to notice that, when the autoregression is close to the instability domain, the value of β can approach 1. In such situations, a large number of iterations are needed before stabilisation of the iterative scheme. For example consider the particular model

$$X_n = aX_{n-1} + b(e_n) + \varepsilon_n$$

where the sequence (e_n) is i.i.d. From Lemma 4,

$$A = \frac{\mathbb{E}(X_n)^2}{\mathbb{E}(X_n)^2 + \sigma(0)} = \frac{1}{1 + \frac{\sigma_0}{\mathbb{E}(X_n)^2}},$$

where $\sigma(0) = c^2/(1-a^2)$ and $\mathbb{E}(X_n) = c'/(1-a)$. Hence,

$$A = \frac{1}{1 + c\frac{1-a}{1+a}} \to 1 \quad \text{if} \quad a \to 1.$$

The iterative scheme can be very slow if a is close to 1. On the opposite, when a is close to -1, the iterations stabilize very quickly.

3.4 Remarks about rates of convergence

As well-known in functional estimation, a smoother b induces, with some extra conditions on the kernel K, a better rate of convergence of the estimators.

Lemma 9. *If the function b is C_ℓ for some integer $\ell > 1$ and if the kernel satisfies*

$$\int e^k K(e)de = 0 \quad \forall k \in 1, \ldots, \ell \tag{21}$$

$$\int K(e)de = 1 \tag{22}$$

(i) if the smoothing parameter is such that, as $n \to \infty$ $h_n \sim n^{\beta_1}(\ln n)^{\beta_2}$ there exists $\beta \in]0,1[$ such that

$$\|\hat{\theta}_n^{(k)} - \theta\|_2 = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O(h_n^\ell) + O_{as}(\beta^k)$$

$$\sup_{e \in \mathcal{E}} |\hat{b}_n^{(k)}(e) - b(e)| = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O(h_n^\ell) + O_{as}(\beta^k)$$

where the $0(.)$'s are uniform with respect to k and n.
(ii) if $h_n \sim (\ln n/n)^{\frac{1}{2\ell+1}}$ the rate of the two first terms is optimal and becomes

$$O_{as}\left(\frac{\ln n}{n}\right)^{\frac{\ell}{2\ell+1}},$$

The proof of (i) is based on the fact that, using (21),

$$\int (b(vh_n + e) - b(e))K(v)f(vh_n + e)dv = O(h_n^\ell).$$

The details are omitted.

4 Prediction intervals

The natural predictor for X_{n+1},

$$\mathbb{E}(X_{n+1}|e_{n+1}, e_n, \dots, e_1, X_n, \dots, X_1) = {}^t\phi_{n+1}\theta + b(e_{n+1})$$

can be evaluated via the estimates of θ and b based on the observations up to time n. In other words, we propose the predictor

$$\hat{X}_{n+1} = {}^t\phi_{n+1}\hat{\theta}_n + \hat{b}_n(e_{n+1}).$$

It should be clear that, under conditions of Corollary 7,

$$\frac{\hat{X}_{n+1} - X_{n+1}}{\hat{\sigma}_n(e_{n+1})} \xrightarrow{\mathcal{L}} \varepsilon_1,$$

and, consequently, building a prediction interval requires an estimation of the noise's quantile function $Q(a)$. The inverse of $Q(a)$ can be consistently estimated by

$$\hat{Q}_n^{-1}(a) = \frac{1}{n}\sum_{j=1}^{n-1} \mathbb{I}_{\left|\frac{\hat{X}_{j+1,n} - X_{j+1}}{\hat{\sigma}_n(e_{j+1})}\right| > a},$$

based on the set of retroactive predictions which use the estimates of θ and b available at time n:

$$\hat{X}_{j+1,n} = {}^t\phi_{j+1}\hat{\theta}_n + \hat{b}_n(e_{j+1}), \quad j \le n-1.$$

Summarizing, we obtain for X_{n+1} the prediction interval at asymptotic level α

$$\left[\hat{X}_{n+1} - \hat{\sigma}_n(e_{n+1})\hat{Q}_n(\alpha)\ ,\ \hat{X}_{n+1} + \hat{\sigma}_n(e_{n+1})\hat{Q}_n(\alpha)\right]$$

5 Appendix

Most proofs below are classical in the field of kernel functional estimation. This is why some details are omitted. The reader can refer to [3] and [9] for complete devlopments.

5.1 Proof of Lemma 1

The proof consists in separating the sequence (ϕ_k) into the T stationary and ergodic subsequences $\{(\phi_k^{(l)})|l = 0, \ldots, T-1\}$ and using the law of large numbers. Details are omitted.

To check regularity of the limit M, consider the vector sequences (ψ_k) and $(\psi_k^{(l)})$ built from

$$Y_n = \sum_{j\geq 0} g_j \sigma(e_{n-j})\varepsilon_{n-j}$$

exactly as (ϕ_k) and $(\phi_k^{(l)})$ are built from (X_n). Similarly, consider the sequences (ψ'_k) and $(\psi_k'^{(l)})$ built from $Y'_n = \sum_{j\geq 0} g_j b(e_{n-j})$. Denoting $\Gamma'^{(l)}$ the covariance matrix of $(\psi_k'^{(l)})$, and noticing that the sequences (ψ_k) and (ψ'_k) are orthogonal,

$$\Gamma^{(l)} = \Gamma'^{(l)} + \mathbb{E}\left(\psi_0^{(l)\,t}\psi_0^{(l)}\right)$$

Hence, if M is singular, the same holds for $\sum_{l=0}^{T-1} \mathbb{E}\left(\psi_0^{(l)\,t}\psi_0^{(l)}\right)$. This in turn implies that there exist $(c_1, \ldots, c_p)$ such that, for every k,

$$c_1 S_k + \cdots + c_p S_{k-p+1} =_{as} 0 \tag{23}$$

where $S_k = Y_k + \ldots + Y_{k-T+1}$ is the sum of the Y's over a period of the noise. Now, it is clear that S_k is a stationary ARMA process having the representation

$$S_k = a_1 S_{k-1} + \ldots + a_p S_{k-p+1} + \sum_{j=0}^{T-1} \sigma(e_{k-j})\varepsilon_{k-j},$$

where the variance of the noise does not vanish since, from (10), $\sigma(e_{k-j}) > 0$. This contradicts (23).

5.2 Proof of Lemma 2

In order to prove (18) we only consider

$$\frac{\sum_{j=p+1}^{n} \left(b(e_j) - b(e)\right) K_n\left(e - e_j\right)}{\sum_{j=p+1}^{n} K_n\left(e - e_j\right)} = \frac{\frac{1}{nh_n}\sum_{j=p+1}^{n} \left(b(e_j) - b(e)\right) K_n\left(e - e_j\right)}{\frac{1}{nh_n}\sum_{j=p+1}^{n} K_n\left(e - e_j\right)}. \tag{24}$$

Now, $\mathbb{E}(\sigma(e_j)\varepsilon_j K_n(e - e_j)) = 0$ for every j, making the treatment of the other part in (17) much simpler.

• <u>The numerator</u> of (24) is conveniently splitted in two parts

$$N_1(e) \quad = \quad \frac{1}{nh_n}\sum_{j=p+1}^{n} (b(e_j) - b(e))K_n\left(e - e_j\right) - \mathbb{E}\left[((b(e_j) - b(e))K_n\left(e - e_j\right)\right]$$

and

$$N_2(e) = \frac{1}{nh_n} \sum_{j=p+1}^{n} \mathbb{E}\left[((b(e_j) - b(e))K_n\left(e - e_j\right)\right]$$

For the so called variance term $N_1(e)$, the basic tool is the exponential inequality

$$P\left(\left|\frac{\sum_{j=1}^n U_j}{n}\right| > \epsilon\right) \leq 2e^{-\frac{n\epsilon^2}{4\delta^2}} \quad \forall \epsilon \in]0, 3\delta^2/d[, \tag{25}$$

which holds for every set $(U_1, \ldots, U_n)$ of independent zero-mean variables such that $|U_j| \leq d$ and $\mathbb{E}(U_j^2) \leq \delta^2$ $(j = 1, \ldots, n)$. This inequality is easily deduced from Bernstein's one as noticed in [16], page 17.

Looking at the independent sequence

$$U_j = \frac{1}{h_n}\left((b(e_j) - b(e))K_n(e_j - e) - \mathbb{E}((b(e_j) - b(e))K_n(e_j - e))\right),$$

firstly, since b and K are bounded, it is clear that

$$|U_j| \leq \frac{c}{h_n},$$

and secondly

$$\begin{aligned} \mathbb{E}(U_j^2) &\leq \frac{1}{h_n^2} \int (b(u) - b(e))^2 K^2\left(\frac{u-e}{h_n}\right) f(u)du \\ &= \frac{1}{h_n} \int (b(vh_n + e) - b(e))^2 K^2(v) f(vh_n + e)dv \leq \frac{c}{h_n}. \end{aligned}$$

Applying inequality (25) with $d = \delta^2 = c/h_n$ we obtain

$$P\left(|N_1(e)| > \epsilon\right) \leq 2e^{-\frac{nh_n\epsilon^2}{4c}} \quad 0 < \epsilon < 1$$

and then

$$P\left(|N_1(e)| > \epsilon_0\sqrt{\frac{\ln n}{nh_n}}\right) \leq 2e^{-\frac{\epsilon_0^2 \ln n}{4c}}.$$

A suitable choice of ϵ_0 yields summability of the r.h.s. and finally, by Borel Cantelli Lemma

$$N_1(e) = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right).$$

We now turn to the biais term $N_2(e)$. From (9),

$$\begin{aligned} N_2(e) &= \frac{1}{h_n} \int (b(u) - b(e))K\left(\frac{u-e}{h_n}\right) f(u)du \\ &= \int (b(vh_n + e) - b(e))K(v)f(vh_n + e)dv = O(h_n^\gamma). \end{aligned}$$

We have thus proved that

$$\begin{aligned} N_1(e) + N_2(e) &= \frac{1}{nh_n} \sum_{j=p+1}^{n} (b(e_j) - b(e)) K_n (e - e_j) \\ &= O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O(h_n^\gamma). \end{aligned}$$

The same rate for $\sup_e(|N_1(e) + N_2(e)|)$ is obtained by covering the domain of e by well chosen intervals and using lipschitz property of the kernel. See [3] and [9] among others for the details.

• <u>The denominator</u>. A similar treatment leads to

$$\sup_{e\in\mathcal{E}} \left| \frac{\sum_{j=p+1}^{n} K_n (e - e_j)}{nh_n} - \frac{\sum_{l=0}^{T-1} f(e - s_l)}{T} \right| \xrightarrow{a.s.} 0 \tag{26}$$

This, together with the fact that $\inf_{e\in\mathcal{E}} f(e) > 0$, leads to

$$\begin{aligned} &\sup_{l} \left| \frac{\sum_{j=p+1}^{n} (b(e_j) - b(e_l)) K_n (e_l - e_j)}{\sum_{j=p+1}^{n} K_n (e_l - e_j)} \right| \\ \leq\ &\sup_{e\in\mathcal{E}} \left| \frac{\sum_{j=p+1}^{n} (b(e_j) - b(e)) K_n (e - e_j)}{\sum_{j=p+1}^{n} K_n (e - e_j)} \right| = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O(h_n^\gamma) \end{aligned}$$

and the proof of (18) is over.

5.3 Proof of Lemma 4

We consider first

$$R_n(e) := \frac{\sum_{j=p+1}^{n} {}^t\phi_j K_n (e - e_j)}{\sum_{j=p+1}^{n} K_n (e - e_j)} = \frac{\frac{\sum_{j=p+1}^{n} {}^t\phi_j K_n(e-e_j)}{nh_n}}{\frac{\sum_{j=p+1}^{n} K_n(e-e_j)}{nh_n}}.$$

The denominator has been already treated in the proof of Lemma 2 (see (26)), so we focus on the numerator and successively show that

$$\sup_{e\in\mathcal{E}} \left| \frac{\sum_{j=p+1}^{n} {}^t\phi_j K_n (e - e_j) - \mathbb{E}({}^t\phi_j K_n (e - e_j))}{nh_n} \right| = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right), \tag{27}$$

then, with $\phi_k^{(l)}$ defined in (13),

$$\sup_{e\in\mathcal{E}} \left| \frac{\sum_{j=p+1}^{n} \mathbb{E}({}^t\phi_j K_n (e - e_j))}{nh_n} - \frac{\sum_{l=0}^{T-1} {}^t\mu^{(l)} f(e - s_l)}{T} \right| = O(h_n^\gamma) \tag{28}$$

The proof of (28) uses $\int K(e)de = 1$. The details are omitted. The proof of (27) follows the lines of the proof of (18), the difference coming from the fact that the $(\phi_j K_n(e_j - e))_k$ are not independent. In fact, they are weakly dependent in so far as, conditionally to the exogeneous sequence, they are mixing.

Lemma 10.

(i) For every $e \in \mathcal{E}$ and every h, the vector sequence $(\phi_j K(\frac{e_j - e}{h}))_j$ is, conditionally to the sequence $(e_j)_j =: \overline{E}$, geometrically α-mixing.

(ii) This property holds uniformly with respect to $\overline{E}$: there exists a constant C and $\alpha \in]0,1[$ such that, $\alpha^{\overline{E}}(n)$ being the conditional mixing sequence,

$$\alpha^{\overline{E}}(n) \leq C\alpha^n \qquad \forall n.$$

Proof. Consider for example the first coordinate $K(\frac{e_j - e}{h})X_{j-1}$ of the vector sequence. Conditionally to $\overline{E}$, the sequence $K(\frac{e_j - e}{h})$ is deterministic, and it is enough to consider the sequence X_j which has the same conditional mixing coefficients as $K(\frac{e_j - e}{h})X_{j-1}$. From (11)

$$X_n = \sum_{j \geq 0} g_j \left(b(e_{n-j}) + \sigma(e_{n-j})\varepsilon_{n-j}\right)$$

is a linear time series based on the bounded noise $b(e_j) + \sigma(e_j)\varepsilon_j$, where $b(e_j)$ and $\sigma(e_j)$ are deterministic trend and variance, while ε_j is i.i.d. Let $h_j(u)$ be the conditional density of the noise. We obtain, since g is C_1 and $\inf_{e \in \mathcal{E}} \sigma(e) > 0$,

$$\begin{aligned} & \int |h_j(u+x) - h_j(u)| du \\ \leq \quad & \int \frac{1}{\sigma(e_j)} \left| g\left(\frac{u + x - b(e_j)}{\sigma(e_j)}\right) - g\left(\frac{u - b(e_j)}{\sigma(e_j)}\right) \right| \; f(v)dv \\ \leq \quad & \frac{\|g\|'_\infty}{\inf_{e \in \mathcal{E}} \sigma^2(e)} |x|. \end{aligned}$$

Now, the sequence $(X_j)_j$ is bounded and, for every j, $|g_j| \leq C\beta^j$ for a certain $\beta \in]0,1[$. Hence the theorem in [13] applies, with any $0 < \delta < 1$: the sequence $K(\frac{e_j - e}{h})X_{j-1}$ is conditionally α-mixing, with mixing coefficients satisfying

$$\alpha^{\overline{E}}(n) \leq C\left(\beta^{\frac{\delta}{1+\delta}}\right)^n =: C\alpha_1^n \qquad \forall n$$

where the constant C does not depend on $\overline{E}$. □

The reader is referred to [4] for definitions and properties of mixing sequences. Hereafter we need to replace inequality (25) by the following one, a direct consequence of theorem 6.2 in [22]:

Let (V_j) be a strong mixing sequence of centered random variables such that

$$\alpha(n) \leq c\alpha^n, \ \forall n \quad \text{and} \quad |V_j| \leq M, \ \forall j$$

Denote $s_n^2 = \sum_{1 \leq j,k \leq n} |\mathrm{Cov}(V_j, V_k)|$. For any $r > 1$ and $\lambda > 0$,

$$P\left(\left|\sum_{j=1}^n V_j\right| > 4\lambda\right) \leq 4\left(1 + \frac{\lambda^2}{r s_n^2}\right)^{-r/2} + \frac{4Mcn}{\lambda}\alpha^{\frac{\lambda}{Mr}}. \qquad (29)$$

This equality applies, conditionally to $\overline{E}$, to

$$V_j = {}^t\phi_j K_n(e-e_j) - \mathbb{E}({}^t\phi_j K_n(e-e_j)), \quad j \geq p+1.$$

For this sequence V_j, the conditional variance s_n^2 satisfies

$$s_n^2 = O(nh_n) \tag{30}$$

where the O is uniform with respect to $\overline{E}$. Indeed,

$$\begin{cases} \mathrm{Var}^{\overline{E}}(V_j) \leq ch_n & \\ |\mathrm{Cov}^{\overline{E}}(V_j, V_l)| \leq h_n^2 & \text{if} \quad |j-l| \leq \delta_n \\ |\mathrm{Cov}^{\overline{E}}(V_j, V_l)| \leq C\alpha_1^{|j-l|} & \text{if} \quad |j-l| > \delta_n \end{cases}$$

For the last bound, the reader can refer to [4]. The two first ones are directly obtained. Taking $\delta_n = 1/(h_n \ln n)$ easily leads to (30).

Now, with $M := \|K\|_\infty \mathrm{esssup}_j |X_j|$, (29) leads to

$$\begin{aligned} P^{\overline{E}}\left(\Big|\sum_{j=p+1}^n {}^t\phi_j K_n(e-e_j) - \mathbb{E}({}^t\phi_j K_n(e-e_j))\Big| > 4\lambda\right) &\leq 4\left(1+\frac{c\lambda^2}{rnh_n}\right)^{-r/2} \\ &+ \frac{4MCn}{\lambda}\alpha_1^{\frac{\lambda}{Mr}}, \end{aligned}$$

and then, if $\ln n = o(r_n)$

$$\begin{aligned} & P^{\overline{E}}\left(\left|\frac{\sum_{j=p+1}^n {}^t\phi_j K_n(e-e_j) - \mathbb{E}({}^t\phi_j K_n(e-e_j))}{nh_n}\right| > \lambda_0\sqrt{\frac{\ln n}{nh_n}}\right) \\ \leq\ & 4\left(1+\frac{c\lambda_0^2 \ln n}{16 r_n}\right)^{-r_n/2} + \frac{16MCn}{\lambda_0\sqrt{nh_n \ln n}}\alpha_1^{\frac{\lambda_0\sqrt{nh_n \ln n}}{4Mr_n}} \\ \leq\ & 4e^{-\frac{c\lambda_0^2 \ln n}{32}} + \frac{C_1}{\lambda_0}\sqrt{\frac{n}{h_n \ln n}}\alpha_1^{\frac{\lambda_0\sqrt{nh_n \ln n}}{4Mr_n}}. \end{aligned}$$

Now, if $h_n \sim n^{\beta_1} \ln n^{\beta_2}$ with $\beta_1 > -1$, $r_n = (\ln n)^\beta$ we get, for n large enough,

$$\begin{aligned} & P^{\overline{E}}\left(\left|\frac{\sum_{j=p+1}^n {}^t\phi_j K_n(e-e_j) - \mathbb{E}({}^t\phi_j K_n(e-e_j))}{nh_n}\right| > \lambda_0\sqrt{\frac{\ln n}{nh_n}}\right) \\ \leq\ & 4n^{-c\lambda_0^2} + C_2\frac{n^{\frac{1-\beta_1}{2}}}{(\ln n)^{\frac{1+\beta_2}{2}}}\alpha_1^{\frac{\lambda_0 n^{\frac{1+\beta_1}{2}}(\ln n)^{\frac{1+\beta_2}{2}-\beta}}{4M}} \\ \leq\ & 4n^{-c\lambda_0^2} + C_2 n^{\frac{1-\beta_1}{2}+\frac{\lambda_0 \ln \alpha_1}{4M}}. \end{aligned} \tag{31}$$

Now, the constants in (31) do not depend on $\overline{E}$, implying that

$$\begin{aligned} & P\left(\left|\frac{\sum_{j=p+1}^n {}^t\phi_j K_n(e-e_j) - \mathbb{E}({}^t\phi_j K_n(e-e_j))}{nh_n}\right| > \lambda_0\sqrt{\frac{\ln n}{nh_n}}\right) \\ \leq\ & 4n^{-c\lambda_0^2} + C_2 n^{\frac{1-\beta_1}{2}+\frac{\lambda_0 \ln \alpha_1}{4M}}. \end{aligned}$$

and it is easy to select λ_0 for the r.h.s. to be the general term of a convergent series.

So we have proved that, for fixed e,

$$\frac{\sum_{j=p+1}^{n} {}^t\phi_j K_n\,(e-e_j) - \mathbb{E}({}^t\phi_j K_n\,(e-e_j))}{nh_n} = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right).$$

The same speed is obtained the sup-norm.
From (27), (28) and (26) it follows that, with

$$\tilde{R}(e) := \frac{\sum_{j=0}^{T-1} {}^t\mu_j f(e-s_j)}{\sum_{j=0}^{T-1} f(e-s_j)}$$

$$\sup_e \left| R_n(e) - \tilde{R}(e) \right| = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O_{as}(h_n^{\gamma}) \tag{32}$$

implying in turn

$$\begin{aligned} A_n &= n\Sigma_n^{-1}\frac{1}{n}\sum_{l=p+1}^{n}\phi_l R_n(e_l) \\ &= n\Sigma_n^{-1}\frac{1}{n}\sum_{l=p+1}^{n}\phi_l(R_n(e_l)-\tilde{R}(e_l)) + n\Sigma_n^{-1}\frac{1}{n}\sum_{l=p+1}^{n}\phi_l\tilde{R}(e_l) \\ &= O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right) + O_{as}(h_n^{\gamma}) + n\Sigma_n^{-1}\frac{1}{n}\sum_{l=p+1}^{n}\phi_l\tilde{R}(e_l). \end{aligned} \tag{33}$$

In (33), the last sum is separated into T ones

$$\frac{1}{n}\sum_{l=p+1}^{n}\phi_l\tilde{R}(e_l) = \sum_{l=0}^{T-1}\frac{1}{n}\sum_{kT+l\le n}^{n}\phi_k^{(l)}\tilde{R}(s_l+\eta_{kT+l}),$$

which almost surely converges to

$$\begin{aligned} \frac{1}{T}\sum_{l=0}^{T-1}\mathbb{E}\left(\phi_0^{(l)}\right)E(\tilde{R}(s_l+\eta_0)) &= \frac{1}{T}\sum_{l=0}^{T-1}\mu^{(l)}E(\tilde{R}(s_l+\eta_0)) \\ &= \frac{1}{T}\sum_{l,j=0}^{T-1}\mu^{(l)}{}^t\mu^{(j)}\int\frac{f(v-s_j)f(v-s_l)}{\sum_{i=0}^{T-1}f(v-s_i)}dv \end{aligned}$$

Moreover, the rate of this convergence, being the rate in the law large numbers for i.i.d sequences, is faster than the first two terms in (33). This, together with (33) and the almost sure convergence of $n\Sigma_n^{-1}$, leads to the desired result.

5.4 Proof of Lemma 5

For the sake of simplicity, we take $T = 2$. The general case only brings more complicated formulas. Denoting

$$S = \Gamma^{(1)} + \Gamma^{(1)},$$

and

$$\alpha_{jl} = \int \frac{f(v - s_j) f(v - s_l)}{\sum_{i=0}^{1} f(v - s_i)} dv, \tag{34}$$

$$A = \left[S + \mu_0^t \mu_0 + \mu_1^t \mu_1\right]^{-1} \left(\alpha_{00}\mu_0^t\mu_0 + \alpha_{11}\mu_1^t\mu_1 + \alpha_{01}(\mu_0^t\mu_1 + \mu_1^t\mu_0)\right). \tag{35}$$

We then apply a popular matrix inversion formula:

$$\begin{aligned}
\left[S + \mu_0^t\mu_0 + \mu_1^t\mu_1\right]^{-1} \mu_1 &= \frac{[S + \mu_0^t\mu_0]^{-1}\mu_1}{1 +^t \mu_1 [S + \mu_0^t\mu_0]^{-1}\mu_1} = \frac{S_1^{-1}\mu_1}{1 +^t \mu_1 M_1^{-1}\mu_1} \\
\left[S + \mu_0^t\mu_0 + \mu_1^t\mu_1\right]^{-1} \mu_0 &= \frac{[S + \mu_1^t\mu_1]^{-1}\mu_0}{1 +^t \mu_0 [S + \mu_1^t\mu_0]^{-1}\mu_1} = \frac{S_0^{-1}\mu_0}{1 +^t \mu_0 S_0^{-1}\mu_0}
\end{aligned}$$

where

$$S_1 = S + \mu_0^t\mu_0 \quad \text{et} \quad S_0 = S + \mu_1^t\mu_1.$$

This leads to

$$A = \frac{S_0^{-1}\mu_0}{1 + {}^t\mu_0 S_0^{-1}\mu_0} \left(\alpha_{00}{}^t\mu_0 + \alpha_{01}{}^t\mu_1\right) + \frac{S_1^{-1}\mu_1}{1 + {}^t\mu_1 S_1^{-1}\mu_1} \left(\alpha_{11}{}^t\mu_1 + \alpha_{01}{}^t\mu_0\right)$$

and finally to

$$\begin{aligned}
A &= \alpha_{00} \frac{S_0^{-1}\mu_0^t\mu_0}{1 + {}^t\mu_0 S_0^{-1}\mu_0} + \alpha_{11} \frac{S_1^{-1}\mu_1^t\mu_1}{1 + {}^t\mu_1 S_1^{-1}\mu_1} \\
&+ \alpha_{01} \left(\frac{S_0^{-1}\mu_0^t\mu_1}{1 + {}^t\mu_0 S_0^{-1}\mu_0} + \frac{S_1^{-1}\mu_1^t\mu_0}{1 + {}^t\mu_1 S_1^{-1}\mu_1} \right) \\
&= \alpha_{00} S_{00} + \alpha_{11} S_{11} + \alpha_{01}(S_{01} + S_{10})
\end{aligned} \tag{36}$$

where the last line defines the S_{jl}'s.

It is easily checked that

$$\begin{aligned}
S_{jj}^2 &= \frac{{}^t\mu_j S_j^{-1}\mu_j}{1 + {}^t\mu_j S_j^{-1}\mu_j} S_{jj} = \beta_{jj} S_{jj}, \quad j = 1, 2 \\
S_{jk}^2 &= \frac{{}^t\mu_k S_j^{-1}\mu_j}{1 + {}^t\mu_j S_j^{-1}\mu_j} S_{jk} = \beta_{jk} S_{jk} \quad j \neq k
\end{aligned}$$

Clearly, $0 \leq \beta_{jj} < 1$. Moreover,

$$|\beta_{jk}| \leq \frac{\sqrt{{}^t\mu_j S_j^{-1}\mu_j}}{1 + {}^t\mu_j S_j^{-1}\mu_j} \sqrt{{}^t\mu_k S_j^{-1}\mu_k} < 1$$

because the first factor is less than 1/2, and

$$
\begin{aligned}
{}^t\mu_k S_j^{-1}\mu_k &= {}^t\mu_k[S+\mu_k^t\mu_k]^{-1}\mu_k = {}^t\mu_k\left(S^{-1}-\frac{S^{-1}\mu_k^t\mu_k S^{-1}}{1+^t\mu_k S^{-1}\mu_k}\right)\mu_k \\
&= \frac{{}^t\mu_k M^{-1}\mu_k}{1+^t\mu_k M^{-1}\mu_k} < 1.
\end{aligned}
$$

As $\alpha_{jl}\in[0,1]$ for every j,l, it results that

$$A^2 = \alpha_{00}^{(2)}S_{00}+\alpha_{11}^{(2)}S_{11}+\alpha_{01}^{(2)}(S_{01}+S_{10})$$

where for every j,l, $|\alpha_{j,l}^{(2)}|\le\beta_{jl}\alpha_{j,l}$, whence

$$A^k = \alpha_{00}^{(k)}M_{00}+\alpha_{11}^{(k)}M_{11}+\alpha_{01}^{(k)}(M_{01}+M_{10})$$

where for every j,l, $|\alpha_{j,l}^{(k)}|\le(\beta_{jl})^{k-1}\alpha_{j,l}$. The lemma is proved.

5.5 Proof of (20) in Theorem 6

The estimation error $\tilde{\sigma}_{n,k}(e)$ is

$$
\begin{aligned}
\tilde{\sigma}_{n,k}(e) &= \hat{\sigma}_{n,k}(e)-\sigma^2(e) \\
&= \frac{\sum_{l=p+1}^{n-1}\left(\left(X_l-^t\phi_l\hat{\theta}_n^{(k)}-\hat{b}_n^{(k-1)}(e_l)\right)^2-\sigma^2(e)\right)K_n(e-e_l)}{\sum_{l=p+1}^{n-1}K_n(e-e_l)} \\
&= \frac{\sum_{l=p+1}^{n-1}\left(\sigma^2(e_l)\varepsilon_l^2-\sigma^2(e)\right)K_n(e-e_l)}{\sum_{l=p+1}^{n-1}K_n(e-e_l)}+R_{n,k}(e) \qquad (37)
\end{aligned}
$$

where, from the first part of the theorem,

$$R_{n,k}(e) = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right)+O_{as}(h_n^\gamma)+O_{as}(\beta^k).$$

Now, since the variables $\sigma^2(e_l)\varepsilon_l^2-\sigma^2(e)$ are independent and centered, the first term in (37) can be treated exactly as was (24), leading to

$$\frac{\sum_{l=p+1}^{n-1}\left(\sigma^2(e_l)\varepsilon_l^2-\sigma^2(e)\right)K_n(e-e_l)}{\sum_{l=p+1}^{n-1}K_n(e-e_l)} = O_{as}\left(\sqrt{\frac{\ln n}{nh_n}}\right)+O_{as}(h_n^{\gamma'})$$

and the proof is over.

References

[1] Buja A, Hastie T.J., Tibshirani R.J. (1989). Linear smoothers and additive models. Ann. Statist. 17. 453–555.

[2] Bhattacharya P. K., Zhao P-L. (1997) *Semiparametric inference in a partial linear model.* Ann. Stat. 25-1 244–262.

[3] Bosq D. (1998) *Nonparametric statistics for stochastic processes.* LNS. Springer.

[4] Doukhan P. (1994) *Mixing: properties and examples.* LNS. Springer.

[5] Engel R.F., Granger C.W.J., Rice J., Weiss A. (1986). Semiparametric estimates of the relation between weather and electricity sales. J. of Am. Sata. Assoc. 81. 310–320.

[6] Fadili J. M. (2005) *Penalized partially linear models using sparse representation with an application to fMRI time series.* IEEE Trans Sign. Proc. 53-9 3436–3448.

[7] Fan J., Jiang J. (2005) *Non parametric inferences for additive models.* J. of Am. Statist. Assoc. 100-471. 890–907.

[8] Fan J., Yao Q. (2003) *Non linear time series. Non parametric and parametric methods.* Springer.

[9] Ferraty F., Vieu Ph. (2001). Statistique fonctionnelle: modèles de régression pour variables aléatoires uni, multi et infiniment dimensionnées. Pub. n° LSP-2001-03.

[10] Gannaz I. (2007) Estimation par ondelettes dans les modèles partiellement linéaires. Thèse de l'Université Joseph Fourier.

[11] Gao J.(1998) *Semiparametric regression smoothing of non-linear time-series.* Scand. J. of Stat. 25. 521–539.

[12] Gao J., Yee T. (2000) *Adaptive estimation in partially linear models.* Canadian J. of Stat. 28-3. 571–588.

[13] Gorodetskii V.V. (1977) *On the strong mixing condition for linear sequences.* Theory Probab. appl. 22. 411–413.

[14] Härdle W. (1990) *Applied non-parametric regression.* Econom. Soc. Monographs.

[15] Hastie T, Tibshirani R. (1991) *Generalized additive models.* Chapman and Hall. London.

[16] Hoeffding W. (1966) *Probability inequalities for sums of bounded random variables.* J. of Am. Statist. Assoc. 58-301. 13–30.

[17] Hu Z., Wang N., Caroll, R. J. (2004) *Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data.* Biometrika. 91-2. 251-262

[18] Lund, R. B. and Basawa, I. V. (1999). *Modeling and inference for periodically correlated time series* In *Asymptotics, nonparametrics, and time series*, vol. 158 of Statist. Textbooks Monogr. 37–62. Dekker, New York.

[19] Mammen E, Linton O, and Nielsen J. (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Ann. of Stat. 27. 1443–1490.

[20] Opsomer J. D. (2000) *Asymptotic properties of backfitting estimators.* J. Mult. Anal. 73-2. 166–179.

[21] Opsomer J. D., Ruppert D. (1997). Fitting a bivariate additive model by local polynomial regression. Annals of Statistics, 25, 186-211.

[22] Rio E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Maths et Applications SMAI. Springer.

[23] Robinson, P.M. (1988). Root-N-consistent semiparametric regression. Econometrica, 56, 931-954.

[24] Severini T. A., Staniswalis J. G. (1994) Quasi-likelihood Estimation in Semiparametric models. J. of Am. Statist. Assoc. 89-426. 501–511.

[25] Stout W. F. (1974) *Almost sure convergence* Academic Press.

[26] Wang L., Yang L. (2007) Spline-backfitted kernel smoothing of non-linear autoregression model. The Ann. of stat. vol. 35. 2474–2503.

[27] Zhou Y., Liang H. (2009) *Statistical inference for semiparametric varying coefficient partially linear models with error-prone linear covariables.* Ann. Stat. 37-1 427–458.