PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **17** (2005), 109–136

PLISKA
STUDIA MATHEMATICA
BULGARICA

STATISTICAL INFERENCE FOR PROCESSES DEPENDING ON ENVIRONMENTS AND APPLICATION IN REGENERATIVE PROCESSES

Christine Jacob, Nadia Lalam, Nicolas Yanev

We consider a process $\{Z_n\}_{n\in\mathbb{N}}$, recursively defined by $Z_n = f(F_{n-1}, E_n) + \eta_n$, where $F_{n-1} = \{Z_k\}_{k\leq n-1}$, $E_n = \{C_k\}_{k\leq n}$, $\{C_n\}_n$ is an observed exogenous process and $\{\eta_n\}_n$ is a martingale difference sequence for the filtration generated by (F_{n-1}, E_n) such that $Var(\eta_n|F_{n-1}, E_n)g(F_{n-1}, E_n) < \infty$, a.s. for some known function $\{g(F_{n-1}, E_n)\}_n$. This class of models covers a very broad range of models such as regression models, ANOVA models, autoregressive processes, branching processes, regenerative processes, ... We assume that $f(F_{n-1}, E_n)$ depends on an unknown parameter μ_0 and that $f(.) \overset{notation}{=} f_{\mu_0}(.)$ may be decomposed according to $f_{\mu_0}(.) = f^{(1)}_{\theta_0}(.) + f^{(2)}_{\mu_0}(.)$, where $\theta_0 \in \mathbb{R}^d$, $d < \infty$, is asymptotically identifiable in $f^{(1)}_{\theta_0}(.)$ as $n \to \infty$ at some rate $v(.)$ whereas $f^{(2)}_{\mu_0}(.)v(.)$ is asymptotically negligible. We build the Conditional Least Squares Estimator of θ_0 based on the observation of a single trajectory of $\{Z_k, C_k\}_k$, and give conditions ensuring its strong consistency. The particular case of general linear models according to $\mu_0 = (\theta_0, \nu_0)$ and among them, regenerative processes, are studied more particularly. In this frame, we may also prove the consistency of the estimator of ν_0 although it belongs to an asymptotic negligible part of the model, and the asymptotic law of the estimator may also be calculated.

1. Introduction

We consider the following one-dimensional nonlinear autoregressive process $\{Z_n\}_{n\in\mathbb{N}}$ that may depend on a multidimensional exogenous process $\{C_n\}_{n\in\mathbb{N}}$:

Z_0 is given and for $n \geq 1$,

$$(1) \qquad Z_n = f(F_{n-1}, E_n) + \eta_n;\ E_n = \{C_k\}_{k \leq n},\ F_{n-1} = \{Z_k\}_{k \leq n-1}.$$

We assume that $f(F_{n-1}, E_n)$ is a measurable function of $(F_{n-1}, E_n)$and $\{\eta_n\}_n$ is a martingale difference sequence for the filtration generated by F_{n-1}, E_n, that is, denoting in the same way the variables (F_{n-1}, E_n) and the $\sigma-$algebra they generate, $E(\eta_n | F_{n-1}, E_n) = 0$. We also assume that there exists $\sigma^2 < \infty$ and $g(F_{n-1}, E_n)$, a measurable and known function of (F_{n-1}, E_n), such that

$$\overline{\lim_n} E(\eta_n^2 | F_{n-1}, E_n) g(F_{n-1}, E_n) \overset{a.s.}{<} \sigma^2.$$

We assume that $f(F_{n-1}, E_n)$ depends on an unknown parameter μ_0 which may be of infinite dimension, that $\{g(F_{n-1}, E_n)\}_n$ does not depend on μ_0, and that $f(.) \overset{notation}{=} f_{\mu_0}(.)$ may be decomposed according to $f_{\mu_0}(.) = f^{(1)}_{\theta_0}(.) + f^{(2)}_{\mu_0}(.)$, where $f^{(1)}_{\theta_0}(F_{k-1}, E_k)$ depends on $\theta_0 \in \Theta \subset \mathbb{R}^d$, $d < \infty$, and $f^{(1)}_{\theta}(F_{k-1}, E_k)$ is a continuous function of θ at θ_0; θ_0 is the parameter to be estimated, while $f^{(2)}_{\mu_0}(.)$ is the nuisance part of the model.

This class of models covers a very large set of processes such as linear or nonlinear stochastic or deterministic regression models, ANOVA models, linear or nonlinear ARMA processes, regenerative processes and branching processes. It is a generalization of the NARX models (nonlinear autoregressive models with exogenous inputs) given in [15]. The model presented here may be explosive in its first two moments, which is the case, for example, of supercritical branching processes.

When $f^{(1)}_{\theta}(F_{k-1}, E_k)$ is infinitely continuously differentiable at any $\theta \in \Theta$, we build the CLSE (Conditional Least Squares Estimator) of θ_0 from $n - h$ observations of a single trajectory of $\{Z_k, C_k\}_{k \leq n:\delta(F_{k-1}, E_k) \neq 0}$, where $\delta(F_{k-1}, E_k) = 1$ if the observation (Z_k, C_k) is taken into account in the estimator, and is zero otherwise. We study its asymptotic properties, mainly the consistency, as $n \to \infty$, with either h or $n - h$ maintained constant. We give general conditions for the strong (or weak) consistency of the estimator, which are easily checked either theoretically or by numerical simulations. In the general case where $f^{(1)}_{\theta}(F_{k-1}, E_k)$ is a continuous function of θ at θ_0, not necessarily differentiable, we build a DCLSE (Discrete CLSE) by minimizing the conditional sum of squares on a discrete subset of Θ. In both cases, the conditions for consistency are extensions of those given in [16] in the setting of size-dependent branching processes and they are the same for the two estimators. The first condition concerns the asymptotic identifiability of θ_0 in $f^{(1)}_{\theta_0}(.)$ at some rate $v(.)$, the second one concerns the asymptotic negligibility of $(f^{(2)}_{\mu_0}(.) - \widehat{f^{(2)}_{\mu_0}}(.))v(.)$, where $\widehat{f^{(2)}_{\mu_0}}(.)$ is any estimation of $f^{(2)}_{\mu_0}(.)$. And the third condition concerns the amount of information $D_n = \sum_{k=h+1}^{n} \delta(F_{k-1}, E_k)[v(F_{k-1}, E_k)]^{-2} g(F_{k-1}, E_k)$ D_n which has to tend to infinity, as $n \to \infty$. This last condition appears to be not only a sufficient condition but also a necessary one when the first two conditions are checked. The identifiability condition ensures that the model $f^{(1)}_{\theta_0}(.)$ is uniquely defined from θ_0, that is $f^{(1)}_{\theta_0}(.)$ is not asymptotically equivalent to $f^{(1)}_{\theta'_0}(.)$ with $\theta'_0 \neq \theta_0$. Simultaneous consistency which occurs under the simultaneous identifiability of the parameters, allows to study the asymptotic distribution of the estimator. But we will show that the simultaneous identifiability is not a necessary condition for consistency.

We study more deeply the class of linear models $Z_n = \mu_0^T W_n + \eta_n$, where the vector W_n is a measurable function of (F_{n-1}, E_n). This class of models covers autoregressive processes ($W_n = F_{n-1}$), ARMA processes ($W_n = F_{n-1}, \{\eta_k\}_{k \leq n-1}$), regression models ($W_n = \{C_k\}_{k \leq n}$, where C_k is a vector of explicative deterministic or stochastic variables) and ANOVA models ($W_n = \{C_k\}_{k \leq n}$ where C_k is a vector of 0 and 1).

Usual consistency criteria in linear models with $g(.) = 1$ and no nuisance parameter, are often based on the relative rate of growth to infinity of $\lambda_{min}(A_n)$ and $\lambda_{max}(A_n)$, where $A_n = \sum_{k=h+1}^{n} W_k W_k^T$ (see for example [1], [3], [12], [13], [14], [15]). The weakest assumptions obtained in this setting are those given in [13]: the least squares estimator is strongly consistent if $[\ln \lambda_{max}(A_n)]^{\rho}[\lambda_{min}(A_n)]^{-1}$ converges a.s. to 0 for some $\rho > 1$ with $\lambda_{min}(A_n)$ converging to ∞. We extend and weaken this condition. Let

$$\begin{aligned} D_n^{(i)} &= \sum_{k=h+1}^{n} ||W_k^{(i)}||^2_{L_2} \delta(F_{k-1}, E_k) g(F_{k-1}, E_k), i = 1, 2 \\ D_n^{(1,2)} &= \sum_{k=h+1}^{n} ||W_k^{(1)}||_{L_2} ||W_k^{(2)}||_{L_2} \delta(F_{k-1}, E_k) g(F_{k-1}, E_k). \end{aligned}$$

Assume the asymptotic identifiability of θ_0 in $\{\theta_0^T W_n^{(1)}\}_n$ and that of ν_0 in $\{\nu_0^T W_n^{(2)}\}_n$. If there exists a deterministic sequence $\{\phi_n\}_n$ such that
i) $\lim_n \phi_n ||\theta_0 - \widehat{\theta}_{h,n,\nu_0}||_{L_2} \exists$ in distribution(resp. $\overline{\lim}_n \phi_n ||\theta_0 - \widehat{\theta}_{h,n,\nu_0}||_{L_2} \overset{a.s.}{<} \infty$),
ii) $\lim_n D_n^{(1)} [\phi_n^2 D_n^{(2)}]^{-1} \overset{P(resp.\, a.s.)}{=} 0$; $\lim_n D_n^{(2)} [D_n^{(1)}]^{-1} \overset{a.s.}{=} 0$,
iii) $\overline{\lim}_n \phi_n D_n^{(1,2)} [D_n^{(1)}]^{-1} \overset{P(resp.\, a.s.)}{<} \infty$,
then $\lim_n \widehat{\theta}_{h,n} \overset{a.s.}{=} \theta_0$ and $\lim_n \widehat{\nu}_{h,n} \overset{P(resp.\, a.s.)}{=} \nu_0$. For example, assume $d = 2$,

$|W_{k,1}| = O(1)$, $|W_{k,2}| = O(k^{-1/2})$, i.i.d. $\{\eta_n\}_n$. Then i), ii), iii) are satisfied in probability with $\phi_n = n^{1/2}$ and a.s. with $\phi_n = n^{1/2}[\ln\ln n]^{-1/2}$. But the condition $\lim_n [\ln \lambda_{max}(A_n)]^\rho [\lambda_{min}(A_n)]^{-1} \stackrel{a.s.}{=} 0$ is not fulfilled here even in the limit case $\rho = 1$, since $\lambda_{max}(A_n) = O(n)$ and $\lambda_{min}(A_n) = O(\ln n)$.

When the vector W_n is orthogonal, for all n, *i.e.* $W_{n,j} W_{n,j'} = 0$, for all $j \neq j'$, the simultaneous identifiability means that the individual amounts of information, $\{D_{n,i}\}_{i=1,d}$ are balanced between the different components $\{\theta_{0,i}\}_{i=1,d}$. But the only condition $\lim_n D_{n,i} \stackrel{a.s.}{=} \infty$ ensures the individual strong consistency of the estimator of $\theta_{0,i}$. In this frame, we study more particularly the strong consistency and the asymptotic normality of the estimators of the offspring mean and the immigration mean for the regenerative Bienaymé-Galton-Watson branching process with immigration only allowed in the state 0.

2. Identifiability and negligibility

Assume $\theta_0 \in \overset{\circ}{\Theta}$, Θ being a compact set of $\mathbb{R}^d$, $d < \infty$. Let $\delta > 0$ and $B_\delta^c = \{\theta = (\theta_1, \ldots, \theta_d) \in \Theta : ||\theta_k - \theta_{0k}||_{L_2} \geq \delta\}$. Let $||.||_n$ be a norm on the space of functions $\{f_{k,n}\}_{k \leq n}$. Let $v(F_{k-1}, E_k)$ a measurable function of (F_{k-1}, E_k) which may depend on θ_0, $\Delta_{\theta_0,\theta}(F_{k-1}, E_k) = (f^{(1)}_{\theta_0}(F_{k-1}, E_k) - f^{(1)}_{\theta}(F_{k-1}, E_k))v(F_{k-1}, E_k)$, for some measurable function $v(F_{k-1}, E_k)$, and let us introduce the following definitions:

Definition 1. *θ_0 is asymptotically identifiable in $\{f^{(1)}_{\theta_0}(F_{k-1}, E_k)\}_k$ for $\{||.||_n\}_n$ if there exists $\{v(F_{k-1}, E_k)\}_k$ depending only on F_{k-1}, E_k, such that, for all $\delta > 0$, $B1 : \underline{\lim}_{n\to\infty} \inf_{\theta \in B_\delta^c} ||\Delta_{\theta_0,\theta}(F_{.-1}, E_.)||_n \stackrel{a.s.}{>} 0$ is satisfied. If moreover, condition $B2 : \overline{\lim}_{n\to\infty} \sup_{\theta \in B_\delta^c} ||\Delta_{\theta_0,\theta}(F_{.-1}, E_.)||_n \stackrel{a.s.}{<} \infty$ is satisfied, then $v(.)$ is called a rate of identifiability of θ_0.*

Notice that $B1$ and $B2$ are satisfied when the stronger conditions
$B1s : \underline{\lim}_{n\to\infty} \inf_{\theta \in B_\delta^c} |\Delta_{\theta_0,\theta}(F_{n-1}, E_n)| \stackrel{a.s.}{>} 0$ and
$B2s : \overline{\lim}_{n\to\infty} \sup_{\theta \in B_\delta^c} |\Delta_{\theta_0,\theta}(F_{n-1}, E_n)| \stackrel{a.s.}{<} \infty$ are satisfied.

Definition 2. *The process $\{r(F_{k-1}, E_k)\}_k$ is asymptotically negligible if $B3 : \overline{\lim}_{n\to\infty} ||r(F_{.-1}, E_.)||_n \stackrel{a.s.}{=} 0$.*

3. Conditional Least Squares Estimator

We aim to estimate θ_0 considering the unknown process $\{f^{(2)}_{\mu_0}(F_{k-1}, E_k)\}_k$ as a nuisance process. If the unknown part of $\{f^{(2)}_{\mu_0}(F_{k-1}, E_k)\}_k$ is given by a finite dimensional parameter ν_0, then ν_0 is set to a given vector ν_n based on the observations until n. For example we may take $\nu_n = \widehat{\nu}_{h,n}$ defined by $(\widehat{\theta}_{h,n}, \widehat{\nu}_{h,n}) = \arg\min_{(\theta,\nu)\in\Theta\times N} \widetilde{S}_{h,n,\nu}(\theta)$, where $\Theta \times N$ is compact, or we may set $\nu_n = 0$. Now if the unknown part of $\{f^{(2)}_{\mu_0}(F_{k-1}, E_k)\}_k$ is of infinite dimension, then we set $f^{(2)}_{\mu_0}(F_{k-1}, E_k)$ to 0, for all k. For simplifying the notations, we will write ν instead of ν_n, $\{\widehat{f^{(2)}_{\mu_0}}(F_{k-1}, E_k)\}_k$ for any estimation of $\{f^{(2)}_{\mu_0}(F_{k-1}, E_k)\}_k$, and $f_{\theta_0,\nu}(F_{k-1}, E_k) = f^{(1)}_{\theta_0}(F_{k-1}, E_k) + \widehat{f^{(2)}_{\mu_0}}(F_{k-1}, E_k)$.

In the case of $f^{(1)}_\theta(.)$ infinitely differentiable at any θ, we define the CLSE estimator $\widehat{\theta}_{h,n,\nu}$ of θ_0 in the following way

$$\text{(2)} \qquad \widehat{\theta}_{h,n,\nu} = \arg\min_{\theta\in\Theta} \widetilde{S}_{h,n,\nu}(\theta)$$

$$\text{(3)} \qquad \widetilde{S}_{h,n,\nu}(\theta) = \sum_{k=h+1}^{n} (Z_k - f_{\theta,\nu}(F_{k-1}, E_k))^2 \delta(F_{k-1}, E_k) g(F_{k-1}, E_k),$$

where $\delta(F_{k-1}, E_k)$ is a Bernoulli variable, measurable function of (F_{k-1}, E_k), equal to 1 when $Z_k - f_{\theta,\nu}(F_{k-1}, E_k)$ is taken into account in the estimator. For example, if the environmental condition C_k necessarily leads to a bad observation of Z_k, we do not take into account $Z_k - f_{\theta,\nu}(F_{k-1}, E_k)$.

In the general case ($f^{(1)}_\theta(.)$ continuous at θ_0 but not necessarily differentiable), we define the DCLSE (Discrete CLSE) by:

$$\widehat{\theta}_{m,h,n,\nu} = \arg\min_{\theta\in\Theta_m} \widetilde{S}_{h,n,\nu}(\theta),$$

where Θ_m is a finite countable subset of Θ.

4. Strong consistency of the Conditional Least Squares Estimators

Assume first that $f^{(1)}_\theta(F_{k-1}, E_k)$ is infinitely continuously differentiable at any $\theta \in \Theta$. Let $v(F_{k-1}, E_k)$ as in the previous section and

$$\Delta_{\theta_0,\nu_0;\theta,\nu}(F_{k-1}, E_k) = (f_{\theta_0,\nu_0}(F_{k-1}, E_k) - f_{\theta,\nu}(F_{k-1}, E_k))v(F_{k-1}, E_k).$$

Then $\Delta_{\theta_0,\nu;\theta,\nu}(F_{k-1}, E_k) = \Delta_{\theta_0;\theta}(F_{k-1}, E_k)$.

Since $\widetilde{S}_{h,n,\nu}(\theta)$ defined by (2) may be written as

$$\widetilde{S}_{h,n,\nu}(\theta) = \sum_{k=h+1}^{n} (\eta_k v(F_{k-1}, E_k) + \Delta_{\theta_0,\nu_0;\theta,\nu}(F_{k-1}, E_k))^2 a(F_{k-1}, E_k)$$

$$a(F_{k-1}, E_k) = \delta(F_{k-1}, E_k)[v(F_{k-1}, E_k)]^{-2} g(F_{k-1}, E_k),$$

the estimator $\widehat{\theta}_{h,n,\nu}$ also satisfies $\widehat{\theta}_{h,n,\nu} = \mathrm{argmin}_{\theta\in\Theta} S_{h,n,\nu}(\theta)$, where $S_{h,n,\nu}(\theta) = \widetilde{S}_{h,n,\nu}(\theta)D_n^{-1}$ with $D_n = \sum_{k=h+1}^n a(F_{k-1}, E_k)$. This leads to the natural norm $||f_{k,n}||_n^2 = [\sum_{k=h+1}^n f_{k,n}^2 a(F_{k-1}, E_k)]D_n^{-1}$ on the space of functions $\{\{f_{k,n}\}_{k\le n}\}$. ¿From now on, we use this norm. In the following proposition, we prove that if θ_0 is asymptotically identifiable in $\{f_{\theta_0}^{(1)}(F_{k-1}, E_k)\}_k$ at the rate $\{v(F_{k-1}, E_k)\}_k$ and if $\{(f_{\mu_0}^{(2)}(F_{k-1}, E_k) - \widehat{f_{\mu_0}^{(2)}}(F_{k-1}, E_k))v(F_{k-1}, E_k)\}_{k\le n}$ is asymptotically negligible, then the strong consistency of $\{\widehat{\theta}_{h,n,\nu}\}_n$ is ensured under some weak additional conditions.

Proposition 3. *Assume $f_\theta(.)$ infinitely continuously differentiable at any θ, and the following conditions $B1$ to $B5$:*
1) $B1$: $\underline{\lim}_{n\to\infty} \inf_{\theta\in B_\delta^c} ||\Delta_{\theta_0,\theta}(F_{.-1}, E_.))||_n \overset{a.s.}{>} 0$.
$B2s$: $\overline{\lim}_{n\to\infty} \sup_{\theta\in B_\delta^c} |\Delta_{\theta_0,\theta}(F_{n-1}, E_n)| \overset{a.s.}{<} \infty$.
2) $B3$: $\overline{\lim}_{n\to\infty} ||(f_{\mu_0}^{(2)}(F_{.-1}, E_.) - \widehat{f_{\mu_0}^{(2)}}(F_{.-1}, E_.))v(F_{.-1}, E_.)||_n \overset{a.s.}{=} 0$;
3) $B4$: $\{D_n\}_n$ is a.s. increasing to ∞;
4) $B5$: for all $\delta > 0$ and (F_{k-1}, E_k), $\sup_{\theta\in B_\delta^c} f_\theta^{(1)}(F_{k-1}, E_k)$ is attained at some $\theta_{F_{k-1},E_k}^{sup}$ (respectively $\inf_{\theta\in B_\delta^c} f_\theta^{(1)}(F_{k-1}, E_k)$ is attained at some $\theta_{F_{k-1},E_k}^{inf}$).
Then, $\{\widehat{\theta}_{h,n,\nu}\}_n$ is strongly consistent, i.e. $\lim_n \widehat{\theta}_{h,n,\nu} \overset{a.s.}{=} \theta_0$.

If $B3$ is checked in probability instead of almost surely, then $\{\widehat{\theta}_{h,n,\nu}\}_n$ is weakly consistent, i.e. $\lim_n \widehat{\theta}_{h,n,\nu} \overset{P}{=} \theta_0$.

Remarks.
1. $B2s$ may be replaced by the weaker assumptions $B2$ and
$B2w : \overline{\lim}_{n\to\infty} \sum_{k=h+1}^{n} \sup_{\theta\in B_\delta^c} \Delta_{\theta_0,\theta}^2(F_{k-1}, E_k)a(F_{k-1}, E_k)D_k^{-2} \overset{a.s.}{<} \infty$.
2. When the nuisance parameter ν_0 is of finite dimension, then $B3$ implies that ν_0 is not asymptotically identifiable in $f_{\mu_0}^{(2)}(.)$ at the rate $v(.)$.
3. Assume that θ_0 is asymptotically identifiable at the rate $v_1(.)$ and ν_0 is asymptotically identifiable at the rate $v_2(.)$ with $v_2(.) > v_1(.)$. Then we may first prove the consistency of $\widehat{\theta}_{h,n}$ using the fact that $f_{\nu_0}^{(2)}(.)v_1(.)$ is asymptotically negligible, and then we may prove the consistency of $\widehat{\nu}_{h,n}$ if $(f_{\theta_0}^{(1)}(.) - f_{\widehat{\theta}_{h,n}}^{(1)}(.))v_2(.)$ is asymptotically negligible, which will be checked if $\widehat{\theta}_{h,n}$ converges sufficiently rapidly. We will detail this problem in the linear case (following section).

Proof. The proof relies on the martingale difference structure of η_n ([9]) and on a sufficient condition for consistency of minimum contrast estimators ([20]). Let $B_\delta^c = \{\theta \in \Theta : \sum_{j=1}^d |\theta_j - \theta_{0,j}| > \delta\}$. If for all $\delta > 0$,

$\underline{\lim}_{n\to\infty}(\inf_{\theta\in B_\delta^c} S_{h,n,\nu}(\theta) - S_{h,n,\nu}(\theta_0)) > 0$ a.s. (resp. in probability), then $\{\widehat{\theta}_{h,n,\nu}\}_n$ is strongly (resp. weakly) consistent (proof in the a.s. case: assume that $\{\widehat{\theta}_{h,n,\nu}\}_n$ is not a.s. consistent; then there exists a non negligible set of trajectories ω such that, for each ω, there exists δ and an infinite subsequence $\{\widehat{\theta}_{h,n_i,\nu}\}_{n_i}$ with $\widehat{\theta}_{h,n_i,\nu} \in B_\delta^c$, for all n_i, implying that $S_{h,n_i,\nu}(\widehat{\theta}_{h,n_i,\nu}) > S_{h,n_i,\nu}(\theta_0)$, which is in contradiction with the definition of $\widehat{\theta}_{h,n_i,\nu}$; in the probability case, δ and $\{n_i\}_i$ do not depend on ω).

According to $B5$, there exists θ_n such that

$$\inf_{\theta\in B_\delta^c} S_{h,n,\nu}(\theta) - S_{h,n,\nu}(\theta_0) = S_{1n}(\theta_n) + 2S_{2n}(\theta_n) + 2S_{3n}(\theta_n),$$

where $S_{1n}(\theta_n) = \sum_{k=h+1}^n [\Delta_{\theta_0,\theta_n}(F_{k-1}, E_k)]^2 a(F_{k-1}, E_k) D_n^{-1}$,
$S_{2n}(\theta_n) = \sum_{k=h+1}^n \Delta_{\theta_0,\nu_0;\theta_0,\nu}(F_{k-1}, E_k)\Delta_{\theta_0,\theta_n}(F_{k-1}, E_k) a(F_{k-1}, E_k) D_n^{-1}$,
$S_{3n}(\theta_n) = \sum_{k=h+1}^n \eta_k v(F_{k-1}, E_k)\Delta_{\theta_0,\theta_n}(F_{k-1}, E_k) a(F_{k-1}, E_k) D_n^{-1}$.
We successively study each $S_{in}(\theta_n)$, $i \in \{1, 2, 3\}$.
1. Since $S_{1n}(\theta_n) = ||\Delta_{\theta_0,\theta_n}(F_{\cdot-1}, E_\cdot)||_n^2$, then

$$\underline{\lim}_n S_{1n}(\theta_n) \geq \underline{\lim}_n \inf_{\theta\in B_\delta^c} ||\Delta_{\theta_0,\theta}(F_{\cdot-1}, E_\cdot)||_n^2.$$

Using $B1$, the right-hand side is strictly positive yielding $\underline{\lim}_n S_{1n}(\theta_n) > 0$ a.s..
2. First notice that $S_{2n}(\theta_n) = 0$, if $\nu = \nu_0$. Otherwise, according to Hölder's inequality, $|S_{2n}(\theta_n)| \leq ||\Delta_{\theta_0,\nu_0;\theta_0,\nu}(F_{\cdot-1}, E_\cdot)||_n ||\Delta_{\theta_0,\theta_n}(F_{\cdot-1}, E_\cdot)||_n$. implying

$$|\underline{\lim}_n S_{2n}(\theta_n)| \leq \overline{\lim}_n ||\Delta_{\theta_0,\nu_0;\theta_0,\nu}(F_{\cdot-1}, E_\cdot)||_n . \overline{\lim}_n \sup_{\theta\in B_\delta^c} ||\Delta_{\theta_0,\theta}(F_{\cdot-1}, E_\cdot)||_n.$$

The right-hand side is equal to 0, due to $B2$ and $B3$, implying $\underline{\lim}_n S_{2n}(\theta_n) \overset{a.s.}{=} 0$.
3. Consider $S_{3n}(\theta_n)$. Assume first that h is constant. Let Θ_{ε_*} a neighborhood of Θ such that all the conditions valid on Θ are also checked on Θ_{ε_*} ($B1$, ..., $B5$, $f_\theta^{(1)}(.)$ infinitely differentiable). Let

$$\begin{aligned} \tilde{\Delta}_{\theta;\theta_0}(F_{k-1}, E_k) &= f_\theta(F_{k-1}, E_k) - f_{\theta_0}(F_{k-1}, E_k) \\ L_n(\theta) &= \sum_{k=h+1}^n \eta_k \tilde{\Delta}_{\theta;\theta_0}(F_{k-1}, E_k) g(F_{k-1}, E_k) \end{aligned}$$

$\{L_n(\theta)\}_n$ is a martingale and $|S_{3n}(\theta_n)| = |L_n(\theta_n)| D_n^{-1}$. Using lemma 5, $\lim_n S_{3n}(\theta_n) \overset{a.s.}{=} 0$.

Assume now $n-h$ constant and denote $L_{1,n}(\theta_n)$ for $L_n(\theta_n)$ when $h = 0$. Then $L_n(\theta_n) = L_{1,n}(\theta_n) - L_{1,h}(\theta_n)$. Then, as above, since $L_{1,n}(\theta_n) \leq \sup_\theta L_{1,n}(\theta)$ and

$L_{1,h}(\theta_n) \le \sup_\theta L_{1,h}(\theta)$, and using $D_n \ge D_h$, we get $\lim_n L_{1,n}(\theta_n)D_n^{-1} \overset{a.s.}{=} 0$ and $\lim_n L_{1,h}(\theta_n)D_n^{-1} \overset{a.s.}{=} 0$, implying $\lim_n S_{3n}(\theta_n) \overset{a.s.}{=} 0$. □

Lemma 4. *Let $a_k \ge 0$, for all k, with $a_1 > 0$, and $S_n = \sum_{k=1}^n a_k$ with $\lim_n S_n \le \infty$. Then $\sum_{k=1}^\infty a_k S_k^{-2} \le 2a_1^{-1} - \lim_n S_n^{-1}$.*

Proof. We have

$$S_1^{-1} - S_n^{-1} = \sum_{k=2}^{n}(S_{k-1}^{-1} - S_k^{-1}) = \sum_{k=2}^{n} a_k[S_{k-1}S_k]^{-1} \ge \sum_{k=2}^{n} a_k S_k^{-2}.$$

Then the result follows from $S_1^{-1} = a_1^{-1}$ and $\sum_{k=2}^n a_k S_k^{-2} = \sum_{k=1}^n a_k S_k^{-2} - a_1 S_1^{-2}$. This result, in the weaker form $\sum_{k=1}^\infty a_k S_k^{-2} \le 2a_1^{-1}$, is given in [9] (p.158) and is based on another proof.

□

Lemma 5. *Let Θ_{ε_*} a neighborhood of Θ such that $f_\theta^{(1)}(.)$ is infinitely continuously differentiable at any $\theta \in \Theta_{\varepsilon_*}$. Assume also*
$B2s$: $\overline{\lim}_k \sup_{\theta\in\Theta_{\varepsilon_*}}[\tilde{\Delta}_{\theta,\theta_0}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 \overset{a.s.}{<} \infty$.
Then $\lim_n L_n(\theta_n)D_n^{-1} \overset{a.s}{=} 0$.

Proof. First notice that $L_n(\theta_n)$ is generally not a martingale. Assume first that $\Theta = [\theta_{min}, \theta_{max}] \subset \mathbb{R}$. Let Θ_* a random regular grid of size $\varepsilon_* \le \theta_{max} - \theta_{min}$, independent of $\{Z_n\}_n$ and which covers Θ, that is $\Theta_* = \{\theta_{*i}\}_{i=1,\dots,I}$, with $\theta_{*i+1} - \theta_{*i} = \varepsilon_*$, $\theta_{*I} \overset{a.s.}{\ge} \theta_{max}$, $\theta_{*1} \overset{a.s.}{\le} \theta_{min}$, and θ_{*1} follows a uniform law on $(\theta_{min} - \varepsilon_*, \theta_{min})$. This implies that for any $\theta \in \Theta$, $\theta_*(\theta) - \theta$ is uniformly distributed on $(-\varepsilon_*/2, +\varepsilon_*/2)$, where $\theta_*(\theta)$ is the point of Θ_* the nearest from θ. If $\Theta \subset \mathbb{R}^d$, we assume this on each coordinate j, $j = 1, \dots, d$. We have

$$L_n(\theta_n) = L_n(\theta_*(\theta_n)) + [L_n(\theta_n) - L_n(\theta_*(\theta_n))].$$

We prove first $\lim_n |L_n(\theta_{*i})D_n^{-1}| \overset{a.s.}{=} 0$, for any $\theta_{*i} \in \Theta_*$; $\{L_n(\theta_{*i})\}_n$ is a martingale because $\{\eta_k\}_k$ is a martingale difference sequence and, for each k, given F_{k-1}, E_k, $\Delta_{\theta_{*i};\theta_0}(F_{k-1}, E_k)$ is independent of η_k, since θ_{*i} is independent of $\{Z_n\}_n$. Moreover D_n is (F_{n-1}, E_n)- measurable and increases with n, and according to $B2s$ and lemma 1

$$\sum_{k=h+1}^{\infty} E([\eta_k \tilde{\Delta}_{\theta_{*i};\theta_0}(F_{k-1}, E_k)g(F_{k-1}, E_k)]^2 | F_{k-1}, E_k) D_k^{-2} \le$$

$$\sigma \sup_{k,\theta}[\tilde{\Delta}_{\theta;\theta_0}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 \sum_{k=h+1}^{\infty} a(F_{k-1}, E_k)D_k^{-2} \overset{a.s.}{<}$$

$$\infty$$

Therefore the SLLNM may be applied, implying the result. Then, since Θ_* is finite, $\lim_n |L_n(\theta_*(\theta_n))|D_n^{-1} \leq \lim_n \max_{\theta_{*i}\in\Theta_*} |L_n(\theta_{*i})|D_n^{-1} \overset{a.s.}{=} 0$.

Next we are going to prove that $\lim_n |L_n(\theta_n) - L_n(\theta_*(\theta_n))|D_n^{-1} \overset{a.s.}{=} 0$, by using the fact that this quantity depends on the difference $\theta_n - \theta_*(\theta_n)$ and not on the particular value taken by θ_n. Let

$$U^*_{m,n}(\theta) = \sum_{k=m}^{n} \eta_k \tilde{\Delta}_{\theta;\theta_*(\theta)}(F_{k-1}, E_k) g(F_{k-1}, E_k) D_k^{-1} \overset{notation}{=} \sum_{k=m}^{n} Y^*_k(\theta)$$

$$U_{m,n}(\theta) = \sum_{k=m}^{n} \eta_k \tilde{\Delta}_{\theta;\theta_0}(F_{k-1}, E_k) g(F_{k-1}, E_k) D_k^{-1}.$$

Since $U^*_{m,n}(\theta) = U_{m,n}(\theta) - U_{m,n}(\theta_*(\theta))$, where $\{U_{m,n}(\theta)\}_n$ and $\{U_{m,n}(\theta_*(\theta))\}_n$ are martingales, then $\{U^*_{m,n}(\theta)\}_n$ is a martingale, and according to Jensen's inequality, this implies that $\{\sup_\theta |U^*_{m,n}(\theta)|\}_n$ is a submartingale. Therefore using th.2.1 from Hall and Heyde (p.14), we get

$$\lambda P(\max_{n:m\leq n\leq m'} \sup_\theta |U^*_{m,n}(\theta)| > \lambda) \leq E(\sup_\theta |U^*_{m,m'}(\theta)|) \leq E(\sup_\theta |\sum_{m}^{m'} Y^*_k(\theta)|)$$

Denote $\theta_{m'} = \arg\sup_\theta |U^*_{m,m'}(\theta)|$. Using Hölder's inequality, for any $\lambda > 0$,

$$\lambda P(\max_{n:m\leq n\leq m'} \sup_\theta |U^*_{m,n}(\theta)| > \lambda) \leq E[\sum_{m}^{m'} Y^*_k(\theta_{m'}) D_k^{-1}]^2)^{1/2}.$$

Let $k \in \{m, \ldots, m'\}$. Using the definition of η_k and Taylor's expansion of $\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k)$ at $\theta_*(\theta_{m'})$ which depends only on each coordinate of $\theta_{m'} - \theta_*(\theta_{m'})$ given F_{k-1}, E_k, we get

$$P(\{\eta_k \in E\} \cap \{\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k) \in D\}|F_{k-1}, E_k) =$$
$$P(\{Z_k \in E + f_{\mu_0}(F_{k-1}, E-k)\} \cap \{\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k) \in D\}|F_{k-1}, E_k) =$$
$$\int_e \int_t P(\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k) \in D|Z_k =$$
$$e + f_{\mu_0}(F_{k-1}, E_k), \theta_{m'} - \theta_*(\theta_{m'}) = t, F_{k-1}, E_k).$$
$$dP(\theta_{m'} - \theta_*(\theta_{m'}) = t|Z_k = e + f_{\mu_0}(F_{k-1}, E_k), F_{k-1}, E_k).$$
$$dP(Z_k = e + f_{\mu_0}(F_{k-1}, E_k)|F_{k-1}, E_k)$$

Since, given F_{k-1}, E_k, $\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k)$ depends only on coordinates of the difference $\theta_{m'} - \theta_*(\theta_{m'})$, which follow a uniform law on $(-\varepsilon_*/2, +\varepsilon_*/2)$, then

$$P(\{\eta_k \in E\} \cap \{\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k) \in D\}|F_{k-1}, E_k) =$$
$$\int_e \int_t P(\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k) \in D|\theta_{m'} - \theta_*(\theta_{m'}) = t, F_{k-1}, E_k).$$
$$dP(\theta_{m'} - \theta_*(\theta_{m'}) = t|F_{k-1}, E_k)dP(Z_k = e + f_{\mu_0}(F_{k-1}, E_k)|F_{k-1}, E_k) =$$
$$P(\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k) \in D|F_{k-1}, E_k)P(\eta_k \in E|F_{k-1}, E_k)$$

that is, η_k and $\Delta_{\theta_{m'},\theta_*(\theta_{m'})}(F_{k-1}, E_k)$ are independent, given F_{k-1}, E_k, leading to $E[\sum_m^{m'} Y_k^*(\theta_{m'})D_k^{-1}]^2 = E[\sum_m^{m'} E(Y_k^{*2}(\theta_{m'})D_k^{-2}|F_{k-1}, E_k)]$. Consequently

$$\lambda P(\max_{n:m\le n\le m'} \sup_\theta |U^*_{m,n}(\theta)| > \lambda) \le$$
$$E[\sum_m^{m'} E(Y_k^{*2}(\theta_{m'})D_k^{-2}|F_{k-1}, E_k)])^{1/2} \le$$
$$\sigma(E[\sum_{k=m}^{\infty} \sup_\theta [\Delta_{\theta,\theta_*(\theta)}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 a(F_{k-1}, E_k)D_k^{-2}])^{1/2} \le$$
$$\sigma(E([\sup_{k>m} \sup_\theta [\Delta_{\theta,\theta_*(\theta)}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 \sum_{k=m}^{\infty} a(F_{k-1}, E_k)D_k^{-2}]))^{1/2}$$

According to $B2s$, to lemma 1, and to $\Delta_{\theta,\theta_*(\theta)}(.) = \Delta_{\theta,\theta_0(\theta)}(.) + \Delta_{\theta_0,\theta_*(\theta)}(.)$, $\overline{\lim}_k \sup_\theta [\Delta_{\theta,\theta_*(\theta)}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 \sum_{k=h+1}^{\infty} a(F_{k-1}, E_k)D_k^{-2}$ is a.s. finite implying that $\sup_{\theta,k>m}[\Delta_{\theta_s,\theta_*(\theta)}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 \sum_{k\ge m} a(F_{k-1}, E_k)D_k^{-2}$ converges a.s. to 0, as $m \to \infty$. Consequently according to Beppo-Levi lemma, $E[\sup_{\theta,k>m}[\Delta_{\theta,\theta_*(\theta)}(F_{k-1}, E_k)v(F_{k-1}, E_k)]^2 \sum_{k\ge m} a(F_{k-1}, E_k)D_k^{-2}]^2$ tends to 0, as $m \to \infty$. Moreover

$$P(\sup_{n:m\le n} \sup_\theta |U^*_{m,n}(\theta)| > \lambda) = \lim_{m'} P(\max_{m\le n\le m'} \sup_\theta |U^*_{m,n}(\theta)| > \lambda).$$

Therefore $P(\sup_{n:m\le n} \sup_\theta |U^*_{m,n}(\theta)| > \lambda)$ tends to 0 as $m \to \infty$. This implies that $\sup_{n:m\le n} \sup_\theta |U^*_{m,n}(\theta)|$ converges to 0 in probability, and therefore there exists a subsequence $\sup_{n:m_i\le n} \sup_\theta |U^*_{m_i,n}(\theta)|$ which converges a.s. to 0, as $m_i \to \infty$. But, for $m > m_i$, $U^*_{m,n}(\theta) = U^*_{m_i,n}(\theta) - U^*_{m_i,m-1}(\theta)$ which implies

$$\begin{aligned}\sup_\theta |U^*_{m,n}(\theta)| &\le \sup_\theta |U^*_{m_i,n}(\theta)| + \sup_\theta |U^*_{m_i,m-1}(\theta)| \\ &\le \sup_\theta |U^*_{m_i,n}(\theta)| + \sup_{m':m'\ge m_i} \sup_\theta |U^*_{m_i,m'}(\theta)|\end{aligned}$$

This implies $\sup_{n:n\geq m}\sup_\theta |U^*_{m,n}(\theta)| \leq 2\sup_{n:n\geq m_i}\sup_\theta |U^*_{m_i,n}(\theta)|$ and therefore the left-hand member converges a.s. to 0, as $m\to\infty$, since the right-hand member converges to 0.

Then, it remains to show that $\lim_n |L_n(\theta_n) - L_n(\theta_*(\theta_n))|D_n^{-1} \overset{a.s.}{=} 0$. Denote $S^*_{k,n} = \sum_{l=1}^{k-1} Y^*_l(\theta_n) \overset{definition}{=} U^*_{1,k-1}(\theta_n)$. We have

$$\begin{aligned}\frac{L_n(\theta_n) - L_n(\theta_*(\theta_n))}{D_n} &= \sum_{k=h+1}^{n} \frac{Y^*_k(\theta_n)D_k}{D_n} = \sum_{k=h+1}^{n} \frac{(S^*_{k+1,n} - S^*_{k,n})D_k}{D_n} \\ &= S^*_{n+1,n} - \sum_{k=h+1}^{n} \frac{S^*_{k,n}(D_k - D_{k-1})}{D_n} = \frac{\sum_{k=h+1}^{n}(S^*_{n+1,n} - S^*_{k,n})a_k}{\sum_{k=h+1}^{n} a_k}\end{aligned}$$

Since $S^*_{n+1,n} - S^*_{k,n} = U^*_{k,n}(\theta_n)$, then

$$\frac{L_n(\theta_n) - L_n(\theta_*(\theta_n))}{D_n} = \frac{\sum_{k=h+1}^{n} U^*_{k,n}(\theta_n)a_k}{\sum_{k=h+1}^{n} a_k}$$

implying

$$\lim_n \frac{|L_n(\theta_n) - L_n(\theta_*(\theta_n))|}{D_n} \leq \lim_N \lim_n \frac{\sum_1^N a_k}{\sum_1^n a_k} \sup_{k<N} |U^*_{k,n}(\theta_n)| + \lim_N \lim_n \sup_{N\leq k\leq n} |U^*_{k,n}(\theta_n)|$$

Now using $U^*_{k,n}(\theta_n) = U^*_{k,N-1}(\theta_n) + U^*_{N,n}(\theta_n)$, for the first term, and $U^*_{k,n}(\theta_n) = U^*_{N,n}(\theta_n) - U^*_{N,k-1}(\theta_n)$, for the second term, we have

$$\begin{aligned}\lim_n \frac{|L_n(\theta_n) - L_n(\theta_*(\theta_n))|}{D_n} \leq\ & \lim_N \lim_n \frac{\sum_1^N a_k}{\sum_1^n a_k}[\sup_{k<N}\sup_\theta |U^*_{k,N-1}(\theta)| + \sup_\theta U^*_{N,n}(\theta)] + \\ & \lim_N \lim_n [\sup_\theta |U^*_{N,n}(\theta)| + \sup_{N\leq k\leq n}\sup_\theta |U^*_{N,k-1}(\theta)|]\end{aligned}$$

Since $\lim_n [\sum_1^N a_k][\sum_1^n a_k]^{-1} \sup_{k<N}\sup_\theta |U^*_{k,N-1}(\theta)| \overset{a.s.}{=} 0$, then

$$\begin{aligned}\lim_n \frac{|L_n(\theta_n) - L_n(\theta_*(\theta_n))|}{D_n} \leq\ & \lim_N \lim_n \frac{\sum_1^N a_k}{\sum_1^n a_k}[\sup_\theta U^*_{N,n}(\theta)] + \\ & \lim_N \lim_n \sup_\theta |U^*_{N,n}(\theta)| + \lim_N \sup_{N\leq k}\sup_\theta |U^*_{N,k-1}(\theta)|] \\ \leq\ & 3\lim_N \sup_{n>N}\sup_\theta U^*_{N,n}(\theta)]\end{aligned}$$

which is null a.s.. □

Consider next the case where $f_\theta^{(1)}(.)$ is not necessarily differentiable. Let G_m a regular grid of $\mathbb{R}^d$ of size ε_m, that is $G_m = \Pi_{j=1}^d G_{m,j}$, where $G_{m,j} = \{x_{m,j}\}_j$, $x_{m,j+1} - x_{m,j} = \varepsilon_m$. Assume $\lim_m \varepsilon_m = 0$. Let $\Theta_m = \Theta \cap G_m$ and define the DCLSE of θ_0 by $\widehat{\theta}_{m,h,n,\nu} = \arg\min_{\theta\in\Theta_m} \widetilde{S}_{h,n,\nu}(\theta)$.

Proposition 6. *Assume* $f_\theta^{(1)}(F_{k-1}, E_k)v(F_{k-1}, E_k)$ *continuous at* θ_0, *uniformly in* k, *i.e.* $\lim_{\theta_m(\theta_0)\to\theta_0} \sup_k |(f_{\theta_m(\theta_0)}^{(1)}(F_{k-1}, E_k) - f_{\theta_0}^{(1)}(F_{k-1}, E_k))v(F_{k-1}, E_k)| \overset{a.s.}{=} 0$. *Assume B*1, *B*2*s*, *B*3, *B*4 *and B*5. *Then* $\lim_m \lim_n \widehat{\theta}_{m,h,n,\nu} \overset{a.s}{=} \theta_0$ *(resp.* $\lim_m \lim_n \widehat{\theta}_{m,h,n,\nu} \overset{P}{=} \theta_0$, *if B*3 *is checked in probability).*

P r o o f. The proof is similar to the previous one and relies on Wu's lemma ([20]) applied to Θ_m: Let $\theta_m(\theta_0)$ the point of Θ_m the nearest from θ_0 and let $B_{m\delta}^c = \{\theta \in \Theta_m : \sum_{j=1}^d |\theta_j - [\theta_m(\theta_0)]_j| > \delta\}$. Then, if for all $\delta > 0$, $\underline{\lim}_m \underline{\lim}_n (\inf_{\theta\in B_{m\delta}^c} S_{h,n,\nu}(\theta) - S_{h,n,\nu}(\theta_m(\theta_0))) > 0$ a.s. (resp. in probability), then $\lim_n \widehat{\theta}_{m,h,n,\nu} \overset{a.s(resp.P).}{=} \theta_m(\theta_0)$ (proof in the a.s. case: assume that it is not true. Then there exists a non negligible set of trajectories ω such that, for each ω, there exists δ and an infinite subsequence $\{\widehat{\theta}_{m_j,h,n_i,\nu}\}_{m_j,n_i}$ with $\widehat{\theta}_{m_j,h,n_i,\nu} \in B_{m\delta}^c$, for all m_j, n_i, implying that $S_{h,n_i,\nu}(\widehat{\theta}_{m_j,h,n_i,\nu}) > S_{h,n_i,\nu}(\theta_{m_j}(\theta_0))$, for large m_j, n_i, which is in contradiction with the definition of $\widehat{\theta}_{m_j,h,n_i,\nu}$; in the probability case, δ and $\{n_i\}_i$ do not depend on ω).

According to $B5$, there exists $\theta_{m,n}$ such that

$$\inf_{\theta\in B_{m\delta}^c} S_{h,n,\nu}(\theta) - S_{h,n,\nu}(\theta_m(\theta_0)) = S_{1n}(\theta_{m,n}) + 2S_{2n}(\theta_{m,n}) + 2S_{3n}(\theta_{m,n}),$$

where $S_{1n}(\theta_{m,n}) = \sum_{k=h+1}^n [\Delta_{\theta_m(\theta_0),\theta_{m,n}}(F_{k-1}, E_k)]^2 a(F_{k-1}, E_k) D_n^{-1}$,

$S_{2n}(\theta_{m,n}) = \sum_{k=h+1}^n \Delta_{\theta_0,\nu_0;\theta_m(\theta_0),\nu}(F_{k-1}, E_k)\Delta_{\theta_m(\theta_0),\theta_{m,n}}(F_{k-1}, E_k)a(F_{k-1}, E_k)D_n^{-1}$,

$S_{3n}(\theta_{m,n}) = \sum_{k=h+1}^n \eta_k v(F_{k-1}, E_k)\Delta_{\theta_m(\theta_0),\theta_{m,n}}(F_{k-1}, E_k)a(F_{k-1}, E_k)D_n^{-1}$.

As previously, we have $\underline{\lim}_n S_{1n}(\theta_{m,n}) \overset{a.s.}{>} 0$, $\lim_n S_{3n}(\theta_{m,n}) \overset{a.s.}{=} 0$, and

$$|\underline{\lim}_m \underline{\lim}_n S_{2n}(\theta_{m,n})| \leq$$

$$[\overline{\lim}_m \overline{\lim}_n ||\Delta_{\theta_0,\theta_m(\theta_0)}||_n + \overline{\lim}_m \overline{\lim}_n ||(f_{\mu_0}^{(2)}(F_{.-1}, E_.) - \widehat{f_{\mu_0}^{(2)}}(F_{.-1}, E_.))v(F_{.-1}, E_.)||_n].$$
$$2\,\overline{\lim}_n \sup_{\theta\in B_\delta^c} ||\Delta_{\theta_0,\theta}(F_{.-1}, E_.)||_n.$$

which converges a.s. to 0 according to $B2$, $B3$ and $\lim_m \theta_m(\theta_0) = \theta_0$ and the continuity of $f_\theta^{(1)}(.)$ at θ_0. □

5. Strong consistency in the linear model

Assume $f(F_{n-1}, E_n) = \mu_0^T W_n$, where W_n is a measurable function of (F_{n-1}, E_n). Let $v^{-1}(W_k) = ||W_k||_{L_p} \stackrel{def.}{=} [\sum_{j=1}^d |W_{k,j}|^p]^{1/p}$. We also denote $||W_k||$ for $||W_k||_{L_p}$. Since by Hölder's inequality: $||W_k||_1 \leq ||W_k||_{L_p} d^{1/q}$, for any q with $p^{-1}+q^{-1} = 1$, then $B1$ with $p = 1$ is the weakest condition among conditions $B1$ with $p \geq 1$. Whereas $B3$ with $p = 1$ is the strongest one. For simplification of the notations, assume here $\delta(F_{k-1}, E_k) = 1$, for all k. Assume that we can decompose W_n according to $W_n = (W_n^{(1)}, W_n^{(2)})$, $W_n^{(2)}$ being the maximum subset of W_n such that $\lim_n D_n^{(2)} D_n^{-1} = 0$, where $D_n^{(i)} = \sum_{k=1}^n ||W_k^{(i)}||^2 g(F_{k-1}, E_k)$, $i = 1, 2$, $D_n = \sum_{k=1}^n ||W_k||^2 g(F_{k-1}, E_k)$. Writing $\overline{W}_k^{(i)} = W_k^{(i)} ||W_k||^{-1}$, $i = 1, 2$, this means that $\{||\overline{W}_k^{(2)}||\}_k$ is asymptotically negligible.

Notice that if $W_k^{(1)}$ and $W_k^{(2)}$ are orthogonal, for all k, *i.e.* $W_{k,i}^{(1)} W_{k,j}^{(2)} = 0$, for all i, j, then, for all k, there exists $i \in \{1, 2\}$ such that $W_k = W_k^{(i)}$ implying $||\overline{W}_k^{(i)}|| = 0$ or 1, *i.e.* $||\overline{W}_k^{(i)}|| = \delta_{||W_k^{(i)}||}$, where $\delta_Z = 1$ if $Z \neq 0$ and is 0 otherwise. Then

$$|| \, ||\overline{W}_.^{(2)}|| \, ||_n = ||\delta_{||W_.^{(2)}||}||_n = \frac{\sum_{k=h+1}^n \delta_{W_k^{(2)}} a(F_{k-1}, E_k)}{\sum_{k=h+1}^n (\delta_{W_k^{(1)}} + \delta_{W_k^{(2)}}) a(F_{k-1}, E_k)}.$$

Therefore, in the orthogonal case, the negligibility of $\{||\overline{W}_k^{(2)}||\}_k$ means that the mean number (or percentage) of observations of $\overline{W}_k^{(2)}$, weighted by $a(.)$, tends a.s. to 0. In the general case, according to the following lemma, we can equivalently use D_n or $D_n^{(1)}$ in proposition 3.

Lemma 7. *Assume* $\lim_n D_n^{(2)} D_n^{-1} = 0$. *Then*

$$\lim_n \frac{|D_n - D_n^{(1)}|}{D_n} \stackrel{a.s.}{=} 0.$$

P r o o f. Use $[||W_k||^2]^{p/2} = ||W_k^{(1)}||^p + ||W_k^{(2)}||^p \leq (||W_k^{(1)}||^2 + ||W_k^{(2)}||^2)^{p/2}$ which implies $D_n \leq D_n^{(1)} + D_n^{(2)}$, leading to the result since $\lim_n D_n^{(2)} D_n^{-1} \stackrel{a.s.}{=} 0$.

Let θ_0, the subset of μ_0 relative to $W_n^{(1)}$ and ν_0, the subset of μ_0 relative to $W_n^{(2)}$. Then the CLSE $(\widehat{\theta}_{h,n}, \widehat{\nu}_{h,n})$ may be written in the following way

$$(4)\quad (\widehat{\theta}_{h,n}, \widehat{\nu}_{h,n}) = \\ ([\sum_{k=h+1}^{n} (Z_k - \widehat{\nu}_{h,n}^T W_k^{(2)}) W_k^{(1)T} g(F_{k-1}, E_k)][\sum_{k=h+1}^{n} W_k^{(1)} W_k^{(1)T} g(F_{k-1}, E_k)]^{-1}, \\ [\sum_{k=h+1}^{n} (Z_k - \widehat{\theta}_{h,n}^T W_k^{(1)}) W_k^{(2)T} g(F_{k-1}, E_k)][\sum_{k=h+1}^{n} W_k^{(2)} W_k^{(2)T} g(F_{k-1}, E_k)]^{-1}).$$

In the particular case where W_k is an orthogonal set of variables, for all k, that is $W_{k,i}W_{k,j} = 0$, for all k, (4) is reduced to

$$(5)\quad \widehat{\mu}_{h,n,i} = [\sum_{k=h+1}^{n} Z_k W_{k,i}^T g(F_{k-1}, E_k)][\sum_{k=h+1}^{n} W_{k,i}^2 g(F_{k-1}, E_k)]^{-1}, \\ i = 1, \dots, d.$$

This means that, in that case, we aim to prove either the individual consistency of each $\widehat{\mu}_{h,n,i}$, or the stronger property of the consistency of $\widehat{\theta}_{h,n}$, under the identifiability of θ_0, and in addition the consistency of $\widehat{\nu}_{h,n}$ under the identifiability of ν_0. The simultaneous consistency means that the rate of convergence is of the same order for all the individual estimators and is a stronger property than the individual consistency. Denote $\lambda_{min}(A)$, the smallest eigen value of A (resp. $\lambda_{max}(A)$, the largest one). Let, for $i \in \{1, 2\}$, $B2^{(i)}$, …, $B4^{(i)}$, be the conditions $B2$,…, $B4$ relative to $\{W_k^{(i)}\}_k$, and

$$\tilde{B}1^{(i)} : \underline{\lim}_n [\lambda_{min}(\sum_{k=h+1}^{n} W_k^{(i)} W_k^{(i)T} g(F_{k-1}, E_k))][D_n^{(i)}]^{-1} \overset{a.s.}{>} 0$$

□

In the following proposition, we give general conditions leading to the consistency of the CLSE of (θ_0, ν_0) although $\nu_0 W_.^{(2)} ||W_.^{(1)}||^{-1}$ is asymptotically negligible. Let $D_n^{(1,2)} = \sum_{k=h+1}^{n} ||W_k^{(1)}|| ||W_k^{(2)}|| g(F_{k-1}, E_k)$.

Proposition 8. *1. Assume* $\tilde{B}1^{(1)}$, $B4^{(1)}$ *and* $\tilde{B}3^{(1)} : \lim_n D_n^{(2)}[D_n^{(1)}]^{-1} \overset{a.s.}{=} 0$. *Then* $\lim_n \widehat{\theta}_{h,n} \overset{a.s.}{=} \theta_0$.
2. Assume in addition $p = 2$, $\tilde{B}1^{(2)}$, $B4^{(2)}$, *and* $\tilde{B}3^{(2)}$ *defined by the existence of a deterministic sequence* $\{\phi_n\}_n$ *such that:*

i) $\lim_n \phi_n ||\theta_0 - \widehat{\theta}_{h,n,\nu_0}||_{L_2}$ *exists in distribution (resp.* $\overline{\lim}_n \phi_n ||\theta_0 - \widehat{\theta}_{h,n,\nu_0}||_{L_2} \overset{a.s.}{<} \infty$*).*
ii) $\lim_n D_n^{(1)}[\phi_n^2 D_n^{(2)}]^{-1} \overset{P(resp.\, a.s.)}{=} 0$.
iii) $\overline{\lim}_n \phi_n D_n^{(1,2)}[D_n^{(1)}]^{-1} \overset{P(resp.\, a.s.)}{<} \infty$.
Then $\lim_n \widehat{\nu}_{h,n} \overset{P(resp.\, a.s.)}{=} \nu_0$.

Examples

1. Let $Z_n = \theta_0 + \nu_0 a_n^{-1} + \eta_n$, where $\lim_n a_n = \infty$, and the $\{\eta_n\}_n$ are i.i.d. $E(\eta_n) = 0$, $E(\eta_n^2) = 1$. Then θ_0 is asymptotically identifiable at the rate $v(.) = 1$, while $\nu_0 a_n^{-1}$ is asymptotically negligible. Let $h = 0$. Then $D_n^{(1)} = n$ and according to item 1 of proposition 8, $\widehat{\theta}_{0,n}$ is strongly consistent. Now ν_0 is asymptotically identifiable in $\nu_0 a_n^{-1}$ at the rate a_n, implying $D_n^{(2)} = \sum_{k=1}^n a_k^{-2}$. Moreover $\lim_n \sqrt{n}(\widehat{\theta}_{0,n,\nu_0} - \theta_0) \overset{d}{=} \mathcal{N}(0,1)$, *i.e.* $\phi_n^2 = D_n^{(1)} = n$, and $D_n^{(1,2)} = \sum_{k=1}^n a_k^{-1}$. Consequently if $\lim_n \sum_{k=1}^n a_k^{-2} = \infty$ and if $[\sum_{k=1}^n a_k^{-1}]n^{-1/2}$ is bounded, conditions of proposition 8 are all satisfied implying the weak consistency of $\widehat{\nu}_{h,n}$. In the particular case $a_k = k^\alpha$, the only solution for having both ii) and iii) is $\alpha = 1/2$. Moreover if the LIL (Law of the Iterated Logarithm) is valid, then $\phi_n = n^{1/2}[\ln\ln n]^{-1/2}$ implying the strong consistency of $\widehat{\nu}_{h,n}$. The condition given in [13] concerns $[\ln \lambda_{max}(\sum_{k=1}^n W_k W_k^T)]^\rho [\lambda_{min}(\sum_{k=1}^n W_k W_k^T)]^{-1} = [\ln n]^\rho [\ln n]^{-1}$ which does not tend to 0. Therefore the conditions described here are weaker.

2. Let $Z_n = \sum_{i=1}^{Z_{n-1}} Y_{n,i}$, where the $\{Y_{n,i}\}_i$ are i.i.d. $(\theta_0 + \nu_0 Z_{n-1}^{-\alpha}), \sigma^2(Z_{n-1}))$, given F_{n-1}, with $\theta_0 \geq 1$, $\nu_0 > 0$, $\alpha > 0$. Then $\{Z_n\}_n$ is a size-dependent branching process belonging to the class of processes studied by Klebaner [11] and which does not extinct with a nonnull probability. We have $Z_n = (\theta_0 + \nu_0 Z_{n-1}^{-\alpha})Z_{n-1} + \eta_n$. Assume first that $\theta_0 = 1$, $\alpha = 1$ and $Y_{n,i} \in \{1,2\}$. Then $\sigma^2(Z_{n-1}) = \nu_0 Z_{n-1}^{-1}(1 - \nu_0 Z_{n-1}^{-1})$ and $g(.) = 1$, implying $D_n^{(2)} = n$. Therefore $\widehat{\nu}_{0,n}$ is strongly consistent. This case has been studied in [17]. Assume now that $\theta_0 > 1$ with $g(Z) = Z^{-1}$. Then $Z_n \theta_0^{-n}$ converges a.s. ([11]), $D_n^{(1)} = \sum_{k=1}^n Z_{k-1}$ and $\widehat{\theta}_{0,n}$ is strongly consistent. But the consistency of $\widehat{\nu}_{0,n}$ depends on the value of α since $D_n^{(2)} = \sum_k Z_{k-1}^{1-2\alpha}$, $D_n^{(1,2)} = \sum_k Z_{k-1}^{1-\alpha}$. When $\alpha = 1$, $\lim_n D_n^{(2)} < \infty$ a.s., and $\widehat{\nu}_{0,n}$ cannot be consistent, whereas when $\alpha = 1/2$, the conditions of proposition 8 are fulfilled in probability with $\phi_n = \theta_0^{n/2}$.

Remarks.

1. In the case $\nu = 0$, $h = 0$, $g(.) = 1$, formula (4) is reduced to the classical formula $\widehat{\theta}_n^T = \mathcal{Z}_n^T \mathcal{W}_n[\mathcal{W}_n^T \mathcal{W}_n]^{-1}$, where $\mathcal{Z}_n^T = (Z_1, \ldots, Z_n)$, $\mathcal{W}_n[i,j] = W_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, and $\tilde{B}1$: $\overline{\lim}_n \lambda_{min}(\mathcal{W}_n^T \mathcal{W}_n) D_n^{-1} \overset{a.s.}{>} 0$.

2. Let $Z_n = \theta_0 W_n + \eta_n$ with $d = 1$, $\{\eta_n\}_n$ independent of $\{W_n\}_n$ and $\underline{\lim}_n E(\eta_k^2)g(W_k) \overset{a.s.}{>} 0$. Then $\widehat{\theta}_{0,n} = [\sum_{k=1}^n Z_k W_k g(W_k)]D_n^{-1}$, and therefore $Var(\widehat{\theta}_{0,n} - \theta_0|\{W_n\}_n) \geq \inf_k E(\eta_k^2)g(W_k)D_n^{-1}$, which does not converge to 0, as $n \to \infty$, if $\lim_n D_n < \infty$. Consequently in the general case, $B4$ is a necessary and sufficient condition for the strong consistency in the meaning that if $B4$ is not checked, then there exist some models in which the estimators are not consistent. But we saw in proposition 8 that we may have the consistency even if the simultaneous identifiability is not ensured.

Proof.

1.Use proposition 3 and lemma 7. Denote now $\overline{W}_.^{(i)} = W_.^{(i)}[||W_.^{(i)}||]^{-1}$, $i = 1, 2$. Concerning θ_0, $v(W_k) = ||W_k^{(1)}||^{-1}$ and conditions of proposition 3 are the following:

$B1^{(1)} : \underline{\lim}_n \inf_{\theta \in B_\delta^c} ||(\theta_0 - \theta)^T \overline{W}_.^{(1)}||_n \overset{a.s.}{>} 0$

$B2s^{(1)} : \overline{\lim}_n \sup_{\theta \in B_\delta^c} |(\theta_0 - \theta)^T \overline{W}_n^{(1)}| \overset{a.s.}{<} \infty$

$B3^{(1)} : \overline{\lim}_n ||(\nu_0 - \widehat{\nu}_{h,n})^T W_.^{(2)}[||W_.^{(1)}||]^{-1}||_n \overset{a.s.}{=} 0$

$B4^{(1)} : \lim_n D_n^{(1)} \overset{a.s.}{=} \infty$

$B5^{(1)} : \forall \delta > 0, \forall W_k^{(1)}, \sup_{\theta \in B_\delta^c} \theta^T W_k^{(1)}$ (resp. $\inf_{\theta \in B_\delta^c} \theta^T W_k^{(1)}$) is attained at some $\theta_{W_k}^{(1)sup}$ (resp. at some $\theta_{W_k}^{(1)inf}$).

Consider $B1^{(1)}$. Let $A_n^{(1)} = \sum_{k=h+1}^n W_k^{(1)} W_k^{(1)T} g(F_{k-1}, E_k)$. Since $A_n^{(1)}$ is a semi-definite matrix, there exists an orthogonal matrix U_n such that $A_n^{(1)} = U_n \Lambda_n U_n^T$, Λ_n being the diagonal matrix of the eigen values of $A_n^{(1)}$. Therefore

$$\underline{\lim}_n \inf_{\theta \in B_\delta^c} ||(\theta_0 - \theta)^T \overline{W}_.^{(1)}||_n^2 =$$
$$\underline{\lim}_n [\inf_{\theta \in B_\delta^c} (\theta_0 - \theta)^T A_n^{(1)} (\theta_0 - \theta)][D_n^{(1)}]^{-1} \geq$$
$$\underline{\lim}_n [\inf_{\theta \in B_\delta^c} [(\theta_0 - \theta)^T U_n \Lambda_n U_n^T (\theta_0 - \theta)]][D_n^{(1)}]^{-1} \geq$$
$$\delta \, \underline{\lim}_n [\lambda_{min}(A_n^{(1)})][D_n^{(1)}]^{-1}$$

and therefore $B1^{(1)}$ is satisfied under $\tilde{B}1^{(1)}$.

Consider now $B2s^{(1)}$. According to Hölder's inequality with $p^{-1} + q^{-1} = 1$,

$$\overline{\lim}_n \sup_{\theta \in B_\delta^c} |(\theta_0 - \theta)^T \overline{W}_n^{(1)}| \leq \overline{\lim}_n \sup_{\theta \in B_\delta^c} ||\theta_0 - \theta||_{L_q} ||\overline{W}_n^{(1)}||_{L_p}$$

which is finite since B_δ^c is compact and $||\overline{W}_n^{(1)}||_{L_p} = 1$.

Next, consider $B3^{(1)}$, using again Hölder's inequality,

$$||(\nu_0 - \widehat{\nu}_{h,n})^T W_{\cdot}^{(2)}[||W_{\cdot}^{(1)}||_{L_p}]^{-1}||_n^2 \leq ||\nu_0 - \widehat{\nu}_{h,n}||_{L_q}^2 \, || \, ||W_{\cdot}^{(2)}||_{L_p}[||W_{\cdot}^{(1)}||_{L_p}]^{-1} \, ||_n^2$$
$$\leq ||\nu_0 - \widehat{\nu}_{h,n}||_{L_q}^2 [D_n^{(2)}][D_n^{(1)}]^{-1} \tag{6}$$

the limit of which is 0 since $\widehat{\nu}_{h,n}$ belongs to the compact set N.

Next concerning $B5^{(1)}$, it is automatically satisfied.

2. In the same way as previously, $B1^{(2)}$, $B2^{(2)}$, $B4^{(2)}$, $B5^{(2)}$ are satisfied. It remains to prove $B3^{(2)} : \overline{\lim}_n \, ||(\theta_0 - \widehat{\theta}_{h,n})^T W_{\cdot}^{(1)}[||W_{\cdot}^{(2)}||]^{-1}||_n \overset{a.s.}{=} 0$. We have, in the same way as for (6),

$$||(\theta_0 - \widehat{\theta}_{h,n})^T W_{\cdot}^{(1)}[||W_{\cdot}^{(2)}||]^{-1}||_n \leq ||\theta_0 - \widehat{\theta}_{h,n}||_{L_q}[D_n^{(1)}[D_n^{(2)}]^{-1}]^{1/2}$$
$$\leq \phi_n||\theta_0 - \widehat{\theta}_{h,n,\nu_0}||_{L_q}[D_n^{(1)}[\phi_n^2 D_n^{(2)}]^{-1}]^{1/2} +$$
$$\phi_n||\widehat{\theta}_{h,n,\nu_0} - \widehat{\theta}_{h,n}||_{L_q}[D_n^{(1)}[\phi_n^2 D_n^{(2)}]^{-1}]^{1/2}. \tag{7}$$

The first term converges in probability to 0 by Billingsley convergence results (known also as Slutzky theorem) [2]. Concerning the second term, using the fact that there exists $C < \infty$ such that $||\widehat{\nu}_{h,n} - \nu_0|| \leq C$ and using Hölder's inequality and (4),

$$\phi_n||\widehat{\theta}_{h,n,\nu_0} - \widehat{\theta}_{h,n}||_{L_q} = \phi_n||(\widehat{\nu}_{h,n} - \nu_0)^T \sum_{k=h+1}^{n} W_k^{(2)} W_k^{(1)T} g(W_k)[A_n^{(1)}]^{-1}||_{L_q}$$
$$\leq \phi_n||(\widehat{\nu}_{h,n} - \nu_0)^T 1 D_n^{(1,2)} 1^T U_n \Lambda_n^{-1} U_n^T||_{L_q}$$
$$\leq \phi_n||\widehat{\nu}_{h,n} - \nu_0||_{L_q} D_n^{(1,2)}||1^T U_n \Lambda_n^{-1} U_n^T||_{L_q}$$
$$\leq \phi_n D_n^{(1,2)} C||1^T U_n \Lambda_n^{-1} U_n^T||_{L_q}. \tag{8}$$

For $q = 2$, $||1^T U_n \Lambda_n^{-1} U_n^T||_{L_q}^2 = 1^T U_n \Lambda_n^{-2} U_n^T 1 \leq [\lambda_{min}(A_n^{(1)}]^{-2} \sum_j (\sum_i U_n[i,j])^2 \leq [\lambda_{min}(A_n^{(1)}]^{-2} d^3$. This leads to $\tilde{B}3^{(2)}$, using $\tilde{B}1^{(1)}$, (7) and (8). □

Corollary 9. *Assume the particular case of W_k orthogonal, for all k. Let $D_{n,i} = \sum_{k=h+1}^{n} |W_{k,i}|^2 g(F_{k-1}, E_k)$. Then*
1.$B1 \iff \underline{\lim}_n \min_{1 \leq i \leq d} ||\delta_{|W_{\cdot,i}|}||_n \overset{a.s.}{>} 0 \iff: \underline{\lim}_n \min_{1 \leq i \leq d} D_{n,i}[D_n]^{-1} \overset{a.s.}{>} 0$.
Under $B1$ and $B4$, $\lim_n \widehat{\theta}_{h,n,\nu} \overset{a.s.}{=} \theta_0$.
2. Under $B4_i : D_{n,i}$ increases to ∞, then $\lim_n \widehat{\theta}_{h,n,\nu,i} \overset{a.s.}{=} \theta_{0,i}$.

Remark. $\underline{\lim}_n \min_{1 \leq i \leq d} D_{n,i}[D_n]^{-1} > 0$ a.s. means that the amounts of information relative to each component of θ_0 are balanced.

P r o o f. Since W_k is orthogonal, for all k, $W_k W_k^T$ is diagonal, for all k, implying the first result. The other results are direct consequence of proposition 8. □

Assume now that W_k is of dimension $d = 2$ and $||.||_{L_p} = ||.||_{L_1}$. Let $D_{n,12} = \sum_{k=h+1}^{n} W_{k,1}W_{k,2}g(F_{k-1}, E_k)$, $D_{n,|12|} = \sum_{k=h+1}^{n} |W_{k,1}W_{k,2}|g(F_{k-1}, E_k)$. According to Hölder's inequality, $D_{n,12}^2 - D_{n,1}D_{n,2} \leq 0$, for all n. $D_{n,12}$ represents the information which is common to $\{W_{k,1}\}_k$ and $\{W_{k,2}\}_k$, whereas $D_{n,1}$ and $D_{n,2}$ represent the individual informations. Notice that when W_k is orthogonal, for all k, then $D_n = D_{n,1} + D_{n,2}$.

Proposition 10. *1. In the general case*

$$B1 \Longleftrightarrow \tilde{B}1 \Longleftrightarrow \underline{\lim}_n (-D_{n,12}^2 + D_{n,1}D_{n,2})(D_{n,1} + D_{n,2})^{-2} \overset{a.s.}{>} 0.$$

2. Assume W_k orthogonal, for all k. Then

$$B1 \Longleftrightarrow \tilde{B}1 : 0 \overset{a.s.}{<} \underline{\lim}_n \frac{D_{n,1}}{D_{n,2}} \leq \overline{\lim}_n \frac{D_{n,1}}{D_{n,2}} \overset{a.s.}{<} \infty. \tag{9}$$

Remark. Assume the particular case W_k orthogonal, for all k, $D_{n,1} > D_{n,2}$, $\lim_n D_{n,i} \overset{a.s.}{=} \infty$, $\lim_n D_{n,2}D_{n,1}^{-1} \overset{a.s.}{=} 0$. Then $\{\widehat{\theta}_{h,n,i}\}_i$ are separately strongly consistent but not simultaneously consistent. Moreover if we assume that $E(\eta_k^2|F_{k-1}, E_k)g(F_{k-1}, E_k) = \sigma^2$ and $\{W_{k,i}^2 g(F_{k-1}, E_k)\}_k$ is deterministic, then $Var(\widehat{\theta}_{h,n,1} - \theta_{0,1})[Var(\widehat{\theta}_{h,n,2} - \theta_{0,2})]^{-1} = D_{n,2}D_{n,1}^{-1}$, which tends to 0. Therefore $\widehat{\theta}_{h,n,1}$ converges infinitely more rapidly than $\widehat{\theta}_{h,n,2}$.

P r o o f. 1. We have

$$\lambda_{min}\left(\sum_{k=h+1}^{n} W_k W_k^T g(F_{k-1}, E_k)\right) =$$
$$2^{-1}[D_{n,1} + D_{n,2} - \sqrt{(D_{n,1} + D_{n,2})^2 + 4(D_{n,12}^2 - D_{n,1}D_{n,2})}].$$

$\tilde{B}1$ is therefore satisfied if and only if

$$\underline{\lim}_n \frac{1 - \sqrt{1 + 4(D_{n,12}^2 - D_{n,1}D_{n,2})(D_{n,1} + D_{n,2})^{-2}}}{D_n(D_{n,1} + D_{n,2})^{-1}} > 0, a.s.$$

For $L_p = L_1$, $D_n = D_{n,1} + D_{n,2} + 2D_{n,|12|}$. But according to Hölder's inequality, $D_{n,|12|}^2 \leq D_{n,1}D_{n,2}$, leading to

$$\overline{\lim}_n \frac{|D_{n,|12|}|}{D_{n,1} + D_{n,2}} \leq \overline{\lim}_n \frac{1}{[D_{n,1}D_{n,2}^{-1}]^{1/2} + [D_{n,2}D_{n,1}^{-1}]^{1/2}} < \infty, a.s.$$

Therefore $\tilde{B}1$ is checked if and only if

$$\underline{\lim}_n 1-\sqrt{1+4(D_{n,12}^2-D_{n,1}D_{n,2})(D_{n,1}+D_{n,2})^{-2}}>0 a.s.$$

which leads to the result.

Next, since $\tilde{B}1$ implies $B1$, it remains to prove that $\tilde{B}1^c$ implies $B1^c$.

$$B1^c : \underline{\lim}_n \sum_{k=h+1}^{n} [(\theta_0-\theta)^T W_k]^2 g(F_{k-1},E_k))[D_n]^{-1} \overset{a.s.}{=} 0.$$

Since $d=2$,

$$\sum_{k=h+1}^{n} [(\theta_0-\theta)^T W_k]^2 g(F_{k-1},E_k) =$$
$$(\theta_{0,1}-\theta_1)^2 D_{n,1}+(\theta_{0,2}-\theta_2)^2 D_{n,2}+2(\theta_{0,1}-\theta_1)(\theta_{0,2}-\theta_2)D_{n,12}.$$

Assume $\tilde{B}1^c$ or equivalently, there exists an infinite subsequence $\{n_j\}_j$ such that

$$\lim_{n_j}[\frac{(\theta_{0,1}-\theta_1)^2 D_{n_j,1}+(\theta_{0,2}-\theta_2)^2 D_{n_j,2}+2(\theta_{0,1}-\theta_1)(\theta_{0,2}-\theta_2)D_{n_j,12}}{D_{n_j}} -$$
$$\frac{(\theta_{0,1}-\theta_1)^2 D_{n_j,1}+(\theta_{0,2}-\theta_2)^2 D_{n_j,2}+2(\theta_{0,1}-\theta_1)(\theta_{0,2}-\theta_2)D_{n_j,1}^{1/2}D_{n_j,2}^{1/2}}{D_{n_j}}] \overset{a.s.}{=} 0.$$

But $(\theta_{0,1}-\theta_1)^2 D_{n_j,1}+(\theta_{0,2}-\theta_2)^2 D_{n_j,2}+2(\theta_{0,1}-\theta_1)(\theta_{0,2}-\theta_2)D_{n_j,1}^{1/2}D_{n_j,2}^{1/2} = [(\theta_{0,1}-\theta_1)D_{n_j,1}^{1/2}+(\theta_{0,2}-\theta_2)D_{n_j,2}^{1/2}]^2$ which is null for some θ. Therefore $B1^c$ is checked.

2. Result (9) is directly deduced from item 1 since $D_{n,12}=0$. □

6. Asymptotic convergence rate of $\widehat{\theta}_{h,n}-\theta_0$ in the linear orthogonal model

In this section, we assume that $\nu_0=0$ and we write $\widehat{\theta}_{h,n}$ instead of $\widehat{\theta}_{h,n,\nu}$.

The asymptotic law of the estimator could be obtained in the general case under some suitable assumptions using central limit theorems for martingales and the classical Taylor's decomposition at the first order of $\partial S_{h,n}/\partial\theta$ at θ_0:

$$-\frac{\partial^2 S_{h,n}}{\partial\theta\partial\theta^T}(\tilde{\theta}_n)(\widehat{\theta}_{h,n}-\theta_0) = -\frac{\partial S_{h,n}}{\partial\theta}(\theta_0),$$

where $\tilde{\theta_n}$ lies between θ_0 and $\widehat{\theta}_{h,n}$. But the assumptions used in these theorems (see for example theorem 7.4.28 in [7]) being difficult to check in the general case, we study here only the linear model with W_k orthogonal, $W_k^{(2)} = 0$ and $\delta(F_{k-1}, E_k) = 1$, for all k.

Since $S_{h,n}(\theta) = \sum_{k=h+1}^n (Z_k - \theta^T W_k)^2 g(F_{k-1}, E_k)$, we have

$$\begin{aligned}
\frac{\partial S_{h,n}}{\partial \theta_i}(\theta_0) &= -2 \sum_{k=h+1}^{n} \eta_k W_{k,i} g(F_{k-1}, E_k) \\
\frac{\partial^2 S_{h,n}}{\partial \theta_i^2}(\theta_0) &= 2 \sum_{k=h+1}^{n} W_{k,i}^2 g(F_{k-1}, E_k) = 2D_{n,i} \\
\frac{\partial^2 S_{h,n}}{\partial \theta_i \partial \theta_j}(\theta_0) &= 2 \sum_{k=h+1}^{n} W_{k,i} W_{k,j} g(F_{k-1}, E_k) = 0
\end{aligned}$$

Then $2^{-1}[\partial^2 S_{h,n}][\partial\theta\partial\theta^T]^{-1}(\theta)$ is a diagonal matrix, independent of θ, with $D_{n,1}$, $\ldots, D_{n,d}$ on the diagonal, and that we denote $\Lambda_{D_{n,.}}$.

Here $\widehat{\theta}_{h,n,i} - \theta_0 = [\sum_k \eta_k W_{k,i} g(F_{k-1}, E_k)] D_{n,i}^{-1}$. Therefore

$$\Lambda_{D_{n,.}}(\widehat{\theta}_{h,n} - \theta_0) = -\frac{1}{2}\frac{\partial S_{h,n}}{\partial \theta}(\theta_0). \tag{10}$$

Lemma 11. $E(\frac{\partial S_{h,n}}{\partial \theta_i}(\theta_0)\frac{\partial S_{h,n}}{\partial \theta_j}(\theta_0)) = 0, \forall i \neq j.$

Proof. Write G_k for $g(F_{k-1}, E_k)$. For $i \neq j$, we have

$$\frac{1}{4}E(\frac{\partial S_{h,n}}{\partial \theta_i}(\theta_0)\frac{\partial S_{h,n}}{\partial \theta_j}(\theta_0)) = 2\sum_{l>k} E[\eta_l \eta_k W_{l,i} W_{k,j} G_l G_k] + \sum_k E[\eta_k^2 W_{k,i} W_{k,j} G_k^2].$$

The quadratic term obtained for $l = k$ is null since $W_{k,i}W_{k,j} = 0$ (orthogonality of W_k), for all k. Moreover, using the σ-(F_{l-1}, E_l) measurability of $\{W_k, G_k\}_{k\leq l}$ and $\{\eta_k\}_{k<l}$, we have

$$\begin{aligned}
E[\eta_l \eta_k W_{l,i} W_{k,j} G_l G_k] &= E[E[\eta_l \eta_k W_{l,i} W_{k,j} G_l G_k] | F_{l-1}, E_l] \\
&= E[\eta_k W_{k,j} G_k W_{l,i} G_l E[\eta_l | F_{l-1}, E_l]]
\end{aligned}$$

which is null since $E[\eta_l | F_{l-1}, E_l] = 0$. □

Proposition 12. *Assume that there exists a deterministic sequence* $\{\phi_n\}_n$ *such that, for* $i = 1, \ldots, d$, *i)* $\lim_n D_{n,i}\phi_n^{-2} \stackrel{P}{=} d_{*,i} < \infty$,

ii) $\lim_n \phi_n^{-2} \sum_{k=h+1}^n E(\eta_k^2|F_{k-1},E_k) g^2(F_{k-1},E_k) W_{k,i}^2 \overset{a.s.}{=} \sigma_i^2$,
iii) $\lim_n \sum_{k=h+1}^n P(|\eta_k| \geq \phi_n \epsilon [g(F_{n-1},E_n)]^{-1} W_{k,i}^{-1} |F_{k-1},E_k) \overset{P}{=} 0, \forall \epsilon > 0$.
Then $\lim_n \phi_n(\widehat{\theta}_{h,n} - \theta_0) \overset{d}{=} \mathcal{N}(0, \Lambda_{\sigma_{\cdot}^2 d_{*,\cdot}^{-2}})$.

Condition ii) may be replaced by the stronger condition: for all $\epsilon > 0$,

$$\lim_n \phi_n^{-2} \sum_{k=h+1}^n E(\eta_k^2|F_{k-1},E_k) g^2(F_{k-1},E_k) W_{k,i}^2 1_{\{|\eta_k g(F_{n-1},E_n) W_{k,i}| \geq \phi_n \epsilon\}} |F_{k-1},E_k) \overset{P}{=} 0.$$

Proof. First, according to (10),

$$\frac{1}{\phi_n} \Lambda_{D_{n,\cdot}} (\widehat{\theta}_{h,n} - \theta_0) = -\frac{1}{\phi_n} \frac{1}{2} \frac{\partial S_{h,n}}{\partial \theta}(\theta_0), \tag{11}$$

where the ith term of this vector is

$$\frac{1}{\phi_n} D_{n,i}(\widehat{\theta}_{h,n,i} - \theta_{0,i}) = \frac{\sum_{k=h+1}^n \eta_k W_{k,i} G_k}{\phi_n}. \tag{12}$$

Moreover according to lemma 11 and (11),

$$E(D_{n,i}(\widehat{\theta}_{h,n,i} - \theta_{0,i}) D_{n,j}(\widehat{\theta}_{h,n,j} - \theta_{0,j})) = 0, i \neq j.$$

Then, for each $i = 1, \ldots, d$, we apply the central limit theorem for martingales (see for example theorem 7.4.28, [7]) to (12) and we obtain

$$\lim_n \frac{\sum_{k=h+1}^n \eta_k W_{k,i} G_k}{\phi_n} \overset{d}{=} \mathcal{N}(0, \sigma_i^2).$$

Finally using Slutzky theorem and the convergence in probability of $\phi_n^{-2} \Lambda_{D_{n,\cdot}}$, we obtain the asymptotic distribution $\mathcal{N}(0, \Lambda_{\sigma_{\cdot}^2 d_{*,\cdot}^{-2}})$ of $\phi_n(\widehat{\theta}_{h,n} - \theta_0) = \phi_n^{-1} \Lambda_{D_{n,\cdot}} (\widehat{\theta}_{h,n} - \theta_0) \phi_n^2 \Lambda_{D_{n,\cdot}}^{-1}$. □

7. Nonparametric estimation

Let $Z_n = \theta_{n0}^T W_n + \eta_n$, where η_n satisfies the assumption of model (1) and θ_{n0} is of finite dimension d, for all n. Then, if $\tilde{B}1$ and $B4$ are checked with $h = n-1$, the conditional least squares estimator $\widehat{\theta}_n^T = [Z_n W_n^T][W_n W_n^T]^{-1}$ is strongly consistent, *i.e.* $\lim_n \widehat{\theta}_n - \theta_{n0} \overset{a.s.}{=} 0$. The proof is the same one as the proof used in the parametric case (Wu's lemma [20]), where $\theta \in B_\delta^c$ is replaced by $\theta_n - \theta_{n0} \in B_{n,\delta}^c(0)$, $B_{n,\delta}^c(0) = \{\theta' : \sum_{j=1}^d |\theta_j' - \theta_{n0,j}| \geq \delta\}$. For example, for $d = 1$ with

$W_n = Z_{n-1}$, then $\widehat{\theta}_n = Z_n Z_{n-1}^{-1}$. In this case, $\tilde{B}1$ is automatically satisfied on the set of nonnull observations implying that the only condition $B4$ "$W_n^2 g(F_{n-1}, E_n)$ increases a.s. to ∞" has to be checked. The supercritical Galton-Watson process is such an example. In that case, $W_n = Z_{n-1}$, $g(F_{n-1}, E_n) = Z_{n-1}^{-1}$ and therefore $B4$ ("Z_{n-1} increases a.s. to ∞") is satisfied on the nonextinction set.

8. Regenerative branching processes

Let a regenerative process (see for ex. [8], [18], [22]) defined by $\{\xi_j, (X_j(.), T_j)\}_j$, where $\{\xi_j\}_j$ is the process of waiting times between the successive working periods T_{j-1}, T_j, and $X_j(.)$ is the jth working process defined on the working period T_j. Let $\mathcal{T}_j = \xi_j + T_j$, $N(t) = \max\{J : \sum_{j=1}^{J} \mathcal{T}_j \leq t\}$; $N(t)$ is the number of periods $\mathcal{T}_j$ until t; and let $\sigma(t) = t - \sum_{j=1}^{N(t)} \mathcal{T}_j - \xi_{N(t)+1}$; $\sigma(t)$ is, when it is positive, the working time from $\sum_{j=1}^{N(t)} \mathcal{T}_j + \xi_{N(t)+1}$ until t. Then, the regenerative process $\{Z_t\}_t$ may be written as

$$\begin{aligned} Z_t &= X_{N(t)+1}(\sigma(t)), \text{ if } \sigma(t) \geq 0 \\ &= 0, \text{ if } \sigma(t) < 0. \end{aligned}$$

When the period $\mathcal{T}_j = \mathcal{T}$, for all j, with $\mathcal{T}$ deterministic, and independent $\{\xi_j, (X_j(.), T_j)\}_j$, for all j, then the different periods $\mathcal{T}_j$ may be considered as replications of the same process leading for example to classical regression models with replications or ANOVA models.

We assume here processes in discrete time with

$$\begin{aligned} X_j(l) &= U_{l+1}(X_j(l-1)) = \delta_{X_j(l-1)} \sum_{i=1}^{X_j(l-1)} Y_{j,i}(l) \\ X_j(0) &= I_j \delta_{j,0}^I, \end{aligned}$$

where the $\{Y_{j,i}(l)\}_i$ are i.i.d. given $\{\{X_{j'}(.), \xi_{j'}\}_{j'<j}, \xi_j\}$, the conditional law of the Bernoulli variable $\delta_{j,0}^I$ may depend on $\{X_{j'}(.), \xi_{j'}\}_{j'<j}$. The $\{I_j\}_j$ are i.i.d. and independent of the past, given $\delta_{j,0}^I = 1$. The $\{T_j\}_j$ are the survival times of the branching processes $\{X_j(.)\}_j$. Therefore Z_n is recursively defined from the past of the process by

$$Z_n = \delta_{Z_{n-1}} \sum_{i=1}^{Z_{n-1}} Y_{n,i} + I_n \delta_n^I,$$

where the $\{Y_{n,i}\}_i$ are i.i.d. with the same conditional laws as the $\{Y_{j,i}(l)\}_i$, and $I_n\delta_n^I$ has the same conditional law as $I_{N(n)+1}\delta_{N(n)+1}^I$. The process $\{Z_n\}_n$ is a general branching process with immigration.

Let $F_{n-1} = \{Z_k\}_{k\leq n-1}$ and E_n any subset of $\{\delta_k^I\}_k, \{C_k\}_k$, C_k being any environmental variable at time k. Assume

$$E(Y_{n,i}|F_{n-1}, E_n) = m\,\alpha(F_{n-1}, E_n);\ Var(Y_{n,i}|F_{n-1}, E_n) = \sigma^2\beta(F_{n-1}, E_n)$$
$$E(I_n|\delta_n^I = 1, F_{n-1}, E_n) = \lambda;\ Var(I_n|\delta_n^I = 1, F_{n-1}, E_n) = b^2.$$

These assumptions mean that the process may be size-dependent. It is the case for population dynamics depending on a limited environment (nearly any biological populations, for example infectious diseases, ...).

Denote $p(F_{n-1}, E_n) = E(\delta_n^I|F_{n-1}, E_n)$. When $\delta_n^I \in E_n$, then $p(F_{n-1}, E_n) = \delta_n^I$. We have

$$E(Z_n|F_{n-1}, E_n) = m\,\alpha(F_{n-1}, E_n)Z_{n-1} + \lambda p(F_{n-1}, E_n). \tag{13}$$

Our aim is the estimation of the parameters of $E(Z_n|F_{n-1}, E_n)$ given by (13). Assuming $\alpha(.)$, $\beta(.)$ and $p(.)$ known, the model is linear in $\theta_0 = (m, \lambda)$. Here $W_k = (\alpha(F_{k-1}, E_k)Z_{k-1}, p(F_{k-1}, E_k))$. We assume $\delta_{Z_{k-1}}\delta_{p(F_{k-1},E_k)} = 0$, for all k, that is W_k is orthogonal, for all k, implying $D_{n,12} = 0$, for all n. It is the case when $p(F_{n-1}, E_n) = \delta_n^I$ with $\delta_n^I = 0$ when $Z_{n-1} > 0$.

Let $\eta_n = Z_n - E(Z_n|F_{n-1}, E_n)$. Then η_n is a martingale difference and we have $E(\eta_n^2|F_{n-1}, E_n) = \sigma^2\beta(F_{n-1}, E_n)Z_{n-1} + (b^2 + \lambda^2(1 - p(F_{n-1}, E_n)))p(F_{n-1}, E_n)$. Let

$$\begin{aligned}
v(F_{n-1}, E_n) &= \alpha^{-1}(F_{n-1}, E_n)Z_{n-1}^{-1}\delta_{Z_{n-1}} + p^{-1}(F_{n-1}, E_n)\delta_{p(F_{n-1},E_n)} \\
g(F_{n-1}, E_n) &= \beta^{-1}(F_{n-1}, E_n)Z_{n-1}^{-1}\delta_{Z_{n-1}} + p^{-1}(F_{n-1}, E_n)\delta_{p(F_{n-1},E_n)} \\
D_{n,1} &= \sum_{k=h+1}^{n} \alpha^2(F_{n-1}, E_n)\beta^{-1}(F_{k-1}, E_k)Z_{k-1} \overset{notation}{=} S_n \\
D_{n,2} &= \sum_{k=h+1}^{n} p(F_{k-1}, E_k) \overset{notation}{=} V_n
\end{aligned}$$

According to (5), the estimators are

$$\widehat{m}_{h,n} = \frac{\sum_{k=h+1}^{n} Z_k\alpha(F_{k-1}, E_k)\beta^{-1}(F_{k-1}, E_k)\delta_{Z_{k-1}}}{\sum_{k=h+1}^{n} Z_{k-1}\alpha^2(F_{k-1}, E_k)\beta^{-1}(F_{k-1}, E_k)\delta_{Z_{k-1}}} \tag{14}$$

$$\widehat{\lambda}_{h,n} = \frac{\sum_{k=h+1}^{n} Z_k\delta_{p(F_{k-1},E_k)}}{\sum_{k=h+1}^{n} p(F_{k-1}, E_k)}. \tag{15}$$

8.1. Identifiability and consistency

8.1.1. General size-dependent case

According to proposition 10, we have the following result

Proposition 13. *1. (m, λ) is identifiable if and only if*

$$0 < \underline{\lim}_n \frac{S_n}{V_n} \leq \overline{\lim}_n \frac{S_n}{V_n} < \infty, a.s. \tag{16}$$

2, If $\overline{\lim}_n V_n S_n^{-1} \overset{a.s.}{=} 0$ (resp. $\overline{\lim}_n S_n V_n^{-1} \overset{a.s.}{=} 0$), then m is identifiable alone with an asymptotically negligible information $\{\delta_k^I\}$ concerning the immigration process, (resp. λ is identifiable alone with an asymptotically negligible information $\{\delta_{Z_{k-1}}\}$ concerning the branching process).

Remark. Since $W_{k,1}v(F_{k-1}, E_k) = \delta_{Z_{k-1}}$, $W_{k,2}v(F_{k-1}, E_k) = \delta_{p(F_{k-1},E_k)}$, then, for $\delta_{p(F_{k-1},E_k)} = \delta_k^I$, $S_n V_n^{-1} = ||\delta_{Z_{.-1}}||_n ||\delta_.^I||_n^{-1}$ is the ratio between the information relative to the presence of the branching process and that relative to the presence of the immigration process. When the presence of these two processes is balanced (16), the parameters are simultaneously identifiable.

The following corollary is a direct application of corollary 9 and proposition 13.

Corollary 14. *1. Assume $B4_1$: S_n increases a.s. to ∞ (resp. $B4_2$: V_n increases a.s. to ∞). Then $\widehat{m}_{h,n}$ (resp. $\widehat{\lambda}_{h,n}$) is strongly consistent.*

2. Assume B4: $S_n + V_n$ increases a.s. to ∞ with

$$0 < \underline{\lim}_n S_n V_n^{-1} \leq \overline{\lim}_n S_n V_n^{-1} < \infty, a.s..$$

Then $(\widehat{m}_{h,n}, \widehat{\lambda}_{h,n})$ is strongly consistent.

8.1.2. Bienaymé-Galton-Watson case with immigrations allowed in the 0 state. Assume that the $\{X_k(.)\}_k$ are i.i.d. Galton-Watson processes, independent of the waiting periods $\{\xi_k\}_k$ which are assumed i.i.d.. Therefore $\alpha(.) = \beta(.) = 1$. Assume $p(F_{k-1}, E_k) = \delta_k^I$, for all k. Then $S_n = \sum_{k=h+1}^n Z_{k-1}$, $V_n = \sum_{k=h+1}^n \delta_k^I$, and $S_n \geq V_n$. According to (14) and (15), the estimators are

$$\widehat{m}_{h,n} = \frac{\sum_{k=h+1}^n Z_k \delta_{Z_{k-1}}}{\sum_{k=h+1}^n Z_{k-1} \delta_{Z_{k-1}}}; \ \widehat{\lambda}_{h,n} = \frac{\sum_{k=h+1}^n I_k \delta_k^I}{\sum_{k=h+1}^n \delta_k^I}.$$

Notice first that when $m < 1$, then $\{Z_n\}_n$ is stationary, implying that the asymptotic information given by $\lim_n S_n + V_n$ is stationary when $n - h$ is fixed, and therefore it cannot increase to ∞. The consistency of the estimators cannot be ensured in that case. So we will assume that h is fixed. Let $s_* = E(\sum_{l=1}^{T_j} X_j(l)) = \lambda(1-m)^{-1}$.

Proposition 15. *If $E(\mathcal{T}_1) < \infty$, and therefore $m < 1$, then on the set $\{\lim_n V_n = \infty\}$, we have $\lim_n S_n V_n^{-1} \overset{a.s.}{=} s_*$, $\lim_n V_n n^{-1} \overset{a.s.}{=} [E(\mathcal{T}_1)]^{-1}$, and $(\widehat{m}_{h,n}, \widehat{\lambda}_{h,n})$ is strongly consistent.*

Remarks
1. If $E(\xi_1) = \infty$ with $\lim_n P(\xi_1 > n)[P(T_1 > n)]^{-1} = \infty$, then $\lim_n P(\sigma(n) \geq 0) = 0$. Therefore on $\{\lim_n S_n < \infty\}$, the information is not sufficient for estimating m nor λ. 2. If $E(T_1) = \infty$ *i.e.* $m \geq 1$ with $\lim_n P(\xi_1 > n)[P(T_1 > n)]^{-1} = 0$, then $\lim_n P(\sigma(n) \leq 0) = 0$, which implies that on $\{\lim_n S_n = \infty\}$, $\widehat{m}_{h,n}$ is strongly consistent whereas on $\lim_n V_n < \infty$, the information concerning λ is not sufficient for its estimation.

Proof. Use the results in the general case together with results of the renewal theory and those of regenerative processes for branching processes. First note that $\lim_n V_n n^{-1} \overset{a.s.}{=} [E(\mathcal{T}_1)]^{-1}$ by the classical renewal theory (cf [8]). Moreover

$$\frac{\sum_{k=1}^{V_n} U_{k-1}}{V_n} \leq \frac{S_n}{V_n} \leq \frac{\sum_{k=1}^{V_n+1} U_{k-1}}{V_n+1} \frac{V_n+1}{V_n}, \tag{17}$$

where the $\{U_k\}_k$ are i.i.d. as $U_1 = \sum_{k=1}^{T_1} Z_{k-1}$ with $E(U_1) = \lambda(1-m)^{-1}$, $Var(U_1) < \infty$. Therefore according to the SLLN for i.i.d. variables and using (17), we get $\lim_n S_n V_n^{-1} \overset{a.s.}{=} E(U_1)$. This implies according to corollary 14, that $(\widehat{m}_{h,n}, \widehat{\lambda}_{h,n})$ is (simultaneously) strongly consistent. □

Notice that we could also use here directly the SLLN for i.i.d. variables for obtaining the strong consistency of $\widehat{\lambda}_{h,n}$, and that of $\widehat{m}_{h,n}$ may be obtained directly by using $\widehat{m}_{h,n} \simeq [S_n V_n^{-1} - \widehat{\lambda}_{h,n}][S_n V_n^{-1}]^{-1}$ which converges a.s. to m.

8.2. Asymptotic convergence rate of $\widehat{\theta}_{h,n} - \theta_0$; $\theta_0 = (m, \lambda)^T$

Proposition 16. *Assume the frame $m < 1$ of proposition 15. Then*

$$\lim_n \sqrt{n}(\widehat{\theta}_{h,n} - \theta_0) \overset{d}{=} \mathcal{N}(0, \Lambda), \tag{18}$$

where Λ is a diagonal matrix with $\sigma^2 s_*^{-1} E(\mathcal{T}_1)$, $b^2 E(\mathcal{T}_1)$ on the diagonal.

Proof. The result follows from proposition 12 with $\phi_n = n^{1/2}$. Here we have $\eta_n g(F_{n-1}, E_n) W_{n,1} = (Z_n - mZ_{n-1})\delta_{Z_{n-1}}$, $\eta_n g(F_{n-1}, E_n) W_{n,2} = (I_n - \lambda)\delta_n^I$,

$d_{*,1} = s_*[E(\mathcal{T}_1)]^{-1}$, $d_{*,2} = [E(\mathcal{T}_1)]^{-1}$ and $\sigma_1^2 = \sigma^2 s_*[E(\mathcal{T}_1)]^{-1}$, $\sigma_2^2 = b^2[E(\mathcal{T}_1)]^{-1}$ since iii) becomes

$$\begin{aligned} n^{-1}\sum_{k=h+1}^{n} E(\eta_k^2|F_{k-1},E_k)g^2(F_{k-1},E_k)W_{k,1}^2 &= n^{-1}\sum_{k=h+1}^{n} Z_{k-1}\sigma^2 \\ &= n^{-1}S_n\sigma^2 \\ n^{-1}\sum_{k=h+1}^{n} E(\eta_k^2|F_{k-1},E_k)g^2(F_{k-1},E_k)W_{k,2}^2 &= n^{-1}\sum_{k=h+1}^{n} b^2\delta_k^I \\ &= n^{-1}V_n b^2. \end{aligned}$$

We have, according to the stationarity of the process, the Lindeberg's condition:

$$\lim_n n^{-1}\sum_{k=h+1}^{n} E([Z_k - mZ_{k-1}]^2 1_{\{|Z_k-mZ_{k-1}|\geq \epsilon n^{1/2}\}}) \stackrel{P}{=} 0 \tag{19}$$

$$\lim_n n^{-1}\sum_{k=h+1}^{n} E((I_k-\lambda)^2\delta_k^I 1_{\{|(I_k-\lambda)\delta_k^I|\geq \epsilon n^{1/2}\}}) \stackrel{P}{=} 0. \tag{20}$$

Notice that in that case, we also have

$$-\frac{1}{2}\frac{1}{\sqrt{n}}\frac{\partial S_{h,n}}{\partial\theta^T}(\theta_0) = \left(\frac{\sum_{l=1}^{S_n}(Y_{l,i}-m)}{\sqrt{n}}, \frac{\sum_{l=1}^{V_n}(I_l-\lambda)}{\sqrt{n}}\right)$$

and therefore the convergence of each term of this vector may also directly follow from the central limit theorem for random sums together with $\lim_n S_n n^{-1} \stackrel{P}{=} s_*[E(\mathcal{T}_1)]^{-1}$, $\lim_n V_n n^{-1} \stackrel{P}{=} [E(\mathcal{T}_1)]^{-1}$. Finally (18) follows from lemma 11. □

REFERENCES

[1] Anderson T.W., J. B. Taylor Strong consistency of least squares estimates in dynamic models, *Ann. Statist.* **7** (1979) 484–489.

[2] Billingsley P. Convergence of probability measures. Wiley series in Probability and Mathematical Statistics. John Wiley 1968.

[3] Christopeit N., K. Helmes Strong consistency of least squares estimators in linear regression models, *Ann. Statist.* **8** (1980) 778-188.

[4] Dion, J.-P., N. M. Yanev Estimation theory for branching processes with or without immigration, *C. R. Acad. Bulgare Sci.* **44(4)** (1991) 19–22.

[5] Dion, J.-P., N. M. Yanev Central limit theorem for martingales in BGWR branching processes with some statistical applications, *Math. Methods Statist.* **4(3)** (1995) 344–358.

[6] Dion, J. P., N. M. Yanev Limit theorems and estimation theory for branching processes with an increasing random number of ancestors, *J. Appl. Probab.* **34(2)** (1997) 309–327.

[7] *Dacunha-Castelle D., M. Duflo* Probabilits et statistiques. 2. Problmes temps mobile. Masson 1993.

[8] Feller W. An introduction to probability theory and its applications, volume II of Probabiblity and Mathematical Statistics. John Wiley 1971.

[9] Hall P., C. C. Heyde Martingale limit theory and its application. Probability and Mathematical Statistics, New York 1980.

[10] Heyde C.C., E. Seneta Analogues of classical limit theorems for the supercritical Galton-Watson process with immigration, *Math. Biosci.* **11** (1971) 249–259.

[11] Klebaner, F. C. On population-size dependent branching processes. *Adv. in Appl. Probab.* **16** (1984) 30–55.

[12] Lai T. L., H. Robbins, C. Z. Wei Strong consistency of estimates in multiple regression II, *J. Multivariate Anal.* **9** (1979) 343–361.

[13] Lai T. L., C. Z. Wei Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Ann. Statist.* **10** (1982) 154–166.

[14] Lai T. L., C. Z. Wei Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters, *J. Multivariate Anal.* **13** (1983) 1–23.

[15] Lai T.L. Asymptotic properties of nonlinear least squares estimates in stochastic regression models, *Ann. Statist.* **22** (1994) 1917–1930.

[16] Lalam N., C. Jacob Estimation of the offspring mean in a supercritical or nearcritical size-dependent branching process, *Adv. Appl. Probab.* **36(2)** (2004) 582–601.

[17] Lalam N., C. Jacob, P. Jagers (2003) Modelling of the PCR amplification process by a size-dependent branching process and estimation of the efficiency. *Adv. Appl. Probab.*, **36(2)** (2004) 602–615.

[18] Yanev, N. M., K. V. Mitov, G. P. Yanev Renewal, regenerative, and branching processes with stable distributions. *J. Math. Sci.* (N. Y.) **122(4)** (2004) 3438–3448.

[19] Wei C.Z., J. Winnicki Estimation of the means in the branching process with immigration, *Ann. Statist.* **18** (1990) 1757–1773.

[20] Wu C.F. Asymptotic theory of nonlinear least squares estimation, *Ann. Statist.* **9** (1981) 501–513.

[21] Yanev, N. M., S. Tschoukova-Dantcheva On the statistics of branching processes with immigration. *C. R. Acad. Bulgare Sci.* **33(4)** (1980) 469–471.

[22] Mitov, K. V., N. M. Yanev Regenerative processes in the infinite mean cycle case. *J. Appl. Probab.* **38(1)** (2001) 165–179.

Christine Jacob
Applied Mathematics and
Informatics unity,
INRA,
78352 Jouy-en-Josas Cedex, France
e-mail: `cj@banian.jouy.inra.fr`

Nadia Lalam
EURANDOM,
P.O. Box 513
5600 MB Eindhoven,
The Netherlands

Nicolas Yanev
Institute of Mathematics
and Informatics,
Bulgarian Academy of Sciences,
8, G.Bonchev Str.,
1113 Sofia, Bulgaria
e-mail: `yanev@math.bas.bg`