Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models.

Cyrille Dubarry, Sylvain Le Corff

Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models.

Cyrille Dubarry* and Sylvain Le Corff†

April 25, 2012

Abstract

The approximation of fixed-interval smoothing distributions is a key issue in inference for general state-space hidden Markov models (HMM). This contribution establishes non-asymptotic bounds for the Forward Filtering Backward Smoothing (FFBS) and the Forward Filtering Backward Simulation (FFBSi) estimators of fixed-interval smoothing functionals. We show that the rate of convergence of the L_q-mean errors of both methods depends on the number of observations T and the number of particles N only through the ratio T/N for additive functionals. In the case of the FFBS, this improves recent results providing bounds depending on $T/\sqrt{N}$.

1 Introduction

State-space models play a key role in statistics, engineering and econometrics; see [2, 11, 18]. Consider a process $\{X_t\}_{t\geq 0}$ taking values in a general state-space $\mathbb{X}$. This hidden process can be observed only through the observation process $\{Y_t\}_{t\geq 0}$ taking values in $\mathbb{Y}$. Statistical inference in general state-space models involves the computation of expectations of additive functionals of the form

$$S_T = \sum_{t=1}^{T} h_t(X_{t-1}, X_t) \; ,$$

conditionally to $\{Y_t\}_{t=0}^T$, where T is a positive integer and $\{h_t\}_{t=1}^T$ are functions defined on $\mathbb{X}^2$. These smoothed additive functionals appear naturally for maximum likelihood parameter inference in hidden Markov models. The computation of the gradient of the log-likelihood function (Fisher score) or of the

*SAMOVAR, Institut Télécom/Télécom SudParis, France, cyrille.dubarry@it-sudparis.eu

†LTCI, Institut Télécom/Télécom ParisTech, France, sylvain.lecorff@telecom-paristech.fr

‡This work was supported by the French National Research Agency, under the program ANR-07 ROBO 0002

intermediate quantity of the Expectation Maximization algorithm involves the estimation of such smoothed functionals, see [2, Chapter 10 and 11] and [10].

Except for linear Gaussian state-spaces or for finite state-spaces, these smoothed additive functionals cannot be computed explicitly. In this paper, we consider Sequential Monte Carlo algorithms, henceforth referred to as particle methods, to approximate these quantities. These methods combine sequential importance sampling and sampling importance resampling steps to produce a set of random particles with associated importance weights to approximate the fixed-interval smoothing distributions.

The most straightforward implementation is based on the so-called path-space method. The complexity of this algorithm per time-step grows only linearly with the number N of particles, see [4]. However, a well-known shortcoming of this algorithm is known in the literature as the path degeneracy; see [10] for a discussion.

Several solutions have been proposed to solve this degeneracy problem. In this paper, we consider the Forward Filtering Backward Smoothing algorithm (FFBS) and the Forward Filtering Backward Simulation algorithm (FFBSi) introduced in [9] and further developed in [12]. Both algorithms proceed in two passes. In the forward pass, a set of particles and weights is stored. In the Backward pass of the FFBS the weights are modified but the particles are kept fixed. The FFBSi draws independently different particle trajectories among all possible paths. Since they use a backward step, these algorithms are mainly adapted for batch estimation problems. However, as shown in [5], when applied to additive functionals, the FFBS algorithm can be implemented forward in time, but its complexity grows quadratically with the number of particles. As shown in [8], it is possible to implement the FFBSi with a complexity growing only linearly with the number of particles.

The control of the L_q-norm of the deviation between the smoothed additive functional and its particle approximation has been studied recently in [5, 6]. In an unpublished paper by [6], it is shown that the FFBS estimator variance of any smoothed additive functional is upper bounded by terms depending on T and N only through the ratio T/N. Furthermore, in [5], for any $q > 2$, a L_q-mean error bound for smoothed functionals computed with the FFBS is established. When applied to strongly mixing kernels, this bound amounts to be of order $T/\sqrt{N}$ either for

(i) uniformly bounded in time general path-dependent functionals,

(ii) unnormalized additive functionals (see [5, Eq. (3.8), pp. 957]).

In this paper, we establish L_q-mean error and exponential deviation inequalities of both the FFBS and FFBSi smoothed functionals estimators. We show that, for any $q \geq 2$, the L_q-mean error for both algorithms is upper bounded by terms depending on T and N only through the ratio T/N under the strong mixing conditions for (i) and (ii). We also establish an exponential deviation inequality with the same functional dependence in T and N.

This paper is organized as follows. Section 2 introduces further definitions and notations and the FFBS and FFBSi algorithms. In Section 3, upper bounds for the L_q-mean error and exponential deviation inequalities of these two algorithms are presented. In Section 4, some Monte Carlo experiments are presented to support our theoretical claims. The proofs are presented in Sections 5 and 6.

2 Framework

Let $\mathbb{X}$ and $\mathbb{Y}$ be two general state-spaces endowed with countably generated σ-fields $\mathcal{X}$ and $\mathcal{Y}$. Let M be a Markov transition kernel defined on $\mathbb{X} \times \mathcal{X}$ and $\{g_t\}_{t\geq 0}$ a family of functions defined on $\mathbb{X}$. It is assumed that, for any $x \in \mathbb{X}$, $M(x,\cdot)$ has a density $m(x,\cdot)$ with respect to a reference measure λ on $(\mathbb{X},\mathcal{X})$. For any integers $T \geq 0$ and $0 \leq s \leq t \leq T$, any measurable function h on $\mathbb{X}^{t-s+1}$, and any probability distribution χ on $(\mathbb{X},\mathcal{X})$, define

$$\phi_{s:t|T}[h] \stackrel{\text{def}}{=} \frac{\int \chi(\mathrm{d}x_0) g_0(x_0) \prod_{u=1}^{T} M(x_{u-1}, \mathrm{d}x_u) g_u(x_u) h(x_{s:t})}{\int \chi(\mathrm{d}x_0) g_0(x_0) \prod_{u=1}^{T} M(x_{u-1}, \mathrm{d}x_u) g_u(x_u)} , \qquad (1)$$

where $a_{u:v}$ is a short-hand notation for $\{a_s\}_{s=u}^{v}$. The dependence on $g_{0:T}$ is implicit and is dropped from the notations.

Remark 1. Note that this equation has a simple interpretation in the particular case of hidden Markov models. Indeed, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\{X_t\}_{t\geq 0}$ a Markov chain on $(\Omega,\mathcal{F},\mathbb{P})$ with transition kernel M and initial distribution χ (which we denote $X_0 \sim \chi$). Let $\{Y_t\}_{t\geq 0}$ be a sequence of observations on $(\Omega,\mathcal{F},\mathbb{P})$ conditionally independent given $\sigma(X_t, t \geq 0)$ and such that the conditional distribution of Y_u given $\sigma(X_t, t \geq 0)$ has a density given by $g(X_u,\cdot)$ with respect to a reference measure on $\mathcal{Y}$ and set $g_u(x) = g(x, Y_u)$. Then, the quantity $\phi_{s:t|T}[h]$ defined by (1) is the conditional expectation of $h(X_{s:t})$ given $Y_{0:T}$:

$$\phi_{s:t|T}[h] = \mathbb{E}\left[h(X_{s:t}) | Y_{0:T}\right] , \quad X_0 \sim \chi .$$

In its original version, the FFBS algorithm proceeds in two passes. In the forward pass, each filtering distribution $\phi_t \stackrel{\text{def}}{=} \phi_{t:t}$, for any $t \in \{0,\ldots,T\}$, is approximated using weighted samples $\left\{(\omega_t^{N,\ell}, \xi_t^{N,\ell})\right\}_{\ell=1}^{N}$, where T is the number of observations and N the number of particles: all sampled particles and weights are stored. In the backward pass of the FFBS, these importance weights are then modified (see [9, 15, 16]) while the particle positions are kept fixed. The importance weights are updated recursively backward in time to obtain an approximation of the fixed-interval smoothing distributions $\left\{\phi_{s:T|T}\right\}_{s=0}^{T}$. The particle approximation is constructed as follows.

Forward pass Let $\{\xi_0^{N,\ell}\}_{\ell=1}^{N}$ be i.i.d. random variables distributed according to the instrumental density ρ_0 and set the importance weights $\omega_0^{N,\ell} \stackrel{\text{def}}{=} \mathrm{d}\chi/\mathrm{d}\rho_0(\xi_0^{N,\ell})\, g_0(\xi_0^{N,\ell})$. The weighted sample $\{(\xi_0^{N,\ell}, \omega_0^{N,\ell})\}_{\ell=1}^{N}$ then targets the

initial filter ϕ_0 in the sense that $\phi_0^N[h] \stackrel{\text{def}}{=} \sum_{\ell=1}^N \omega_0^{N,\ell} h(\xi_0^{N,\ell}) / \sum_{\ell=1}^N \omega_0^{N,\ell}$ is a consistent estimator of $\phi_0[h]$ for any bounded and measurable function h on $\mathbb{X}$. Let now $\{(\xi_{s-1}^{N,\ell}, \omega_{s-1}^{N,\ell})\}_{\ell=1}^N$ be a weighted sample targeting ϕ_{s-1}. We aim at computing new particles and importance weights targeting the probability distribution ϕ_s. Following [17], this may be done by simulating pairs $\{(I_s^{N,\ell}, \xi_s^{N,\ell})\}_{\ell=1}^N$ of indices and particles from the instrumental distribution:

$$\pi_{s|s}(\ell, h) \propto \omega_{s-1}^{N,\ell} \vartheta_s(\xi_{s-1}^{N,\ell}) P_s(\xi_{s-1}^{N,\ell}, h) \ ,$$

on the product space $\{1, \ldots, N\} \times \mathbb{X}$, where $\{\vartheta_s(\xi_{s-1}^{N,\ell})\}_{\ell=1}^N$ are the adjustment multiplier weights and P_s is a Markovian proposal transition kernel. In the sequel, we assume that $P_s(x, \cdot)$ has, for any $x \in \mathbb{X}$, a density $p_s(x, \cdot)$ with respect to the reference measure λ. For any $\ell \in \{1, \ldots, N\}$ we associate to the particle $\xi_s^{N,\ell}$ its importance weight defined by:

$$\omega_s^{N,\ell} \stackrel{\text{def}}{=} \frac{m(\xi_{s-1}^{N,I_s^{N,\ell}}, \xi_s^{N,\ell}) g_s(\xi_s^{N,\ell})}{\vartheta_s(\xi_{s-1}^{N,I_s^{N,\ell}}) p_s(\xi_{s-1}^{N,I_s^{N,\ell}}, \xi_s^{N,\ell})} \ .$$

Backward smoothing For any probability measure η on $(\mathbb{X}, \mathcal{X})$, denote by B_η the backward smoothing kernel given, for all bounded measurable function h on $\mathbb{X}$ and for all $x \in \mathbb{X}$, by:

$$\mathrm{B}_\eta(x, h) \stackrel{\text{def}}{=} \frac{\int \eta(\mathrm{d}x') \ m(x', x) h(x')}{\int \eta(\mathrm{d}x') \ m(x', x)} \ ,$$

For all $s \in \{0, \ldots, T-1\}$ and for all bounded measurable function h on $\mathbb{X}^{T-s+1}$, $\phi_{s:T|T}[h]$ may be computed recursively, backward in time, according to

$$\phi_{s:T|T}[h] = \int \mathrm{B}_{\phi_s}(x_{s+1}, \mathrm{d}x_s) \, \phi_{s+1:T|T}(\mathrm{d}x_{s+1:T}) \, h(x_{s:T}) \ .$$

2.1 The forward filtering backward smoothing algorithm

Consider the weighted samples $\left\{(\xi_t^{N,\ell}, \omega_t^{N,\ell})\right\}_{\ell=1}^N$, drawn for any $t \in \{0, \ldots, T\}$ in the forward pass. An approximation of the fixed-interval smoothing distribution can be obtained using

$$\phi_{s:T|T}^N[h] = \int \mathrm{B}_{\phi_s^N}(x_{s+1}, \mathrm{d}x_s) \, \phi_{s+1:T|T}^N(\mathrm{d}x_{s+1:T}) \, h(x_{s:T}) \ , \tag{2}$$

and starting with $\phi_{T:T|T}^N[h] = \phi_T^N[h]$. Now, by definition, for all $x \in \mathbb{X}$ and for all bounded measurable function h on $\mathbb{X}$,

$$\mathrm{B}_{\phi_s^N}(x, h) = \sum_{i=1}^N \frac{\omega_s^{N,i} m(\xi_s^{N,i}, x)}{\sum_{\ell=1}^N \omega_s^{N,\ell} m(\xi_s^{N,\ell}, x)} h\left(\xi_s^{N,i}\right) \ ,$$

and inserting this expression into (2) gives the following particle approximation of the fixed-interval smoothing distribution $\phi_{0:T|T}[h]$

$$\phi_{0:T|T}^N[h] = \sum_{i_0=1}^N \cdots \sum_{i_T=1}^N \left(\prod_{u=1}^T \Lambda_u^N(i_u, i_{u-1}) \right) \times \frac{\omega_T^{N,i_T}}{\Omega_T^N} h\left(\xi_0^{N,i_0}, \ldots, \xi_T^{N,i_T}\right) , \quad (3)$$

where h is a bounded measurable function on $\mathbb{X}^{T+1}$,

$$\Lambda_t^N(i,j) \stackrel{\text{def}}{=} \frac{\omega_t^{N,j} m(\xi_t^{N,j}, \xi_{t+1}^{N,i})}{\sum_{\ell=1}^N \omega_t^{N,\ell} m(\xi_t^{N,\ell}, \xi_{t+1}^{N,i})} , \quad (i,j) \in \{1, \ldots, N\}^2 , \quad (4)$$

and

$$\Omega_t^N \stackrel{\text{def}}{=} \sum_{\ell=1}^N \omega_t^{N,\ell} . \quad (5)$$

The estimator of the fixed-interval smoothing distribution $\phi_{0:T|T}^N$ might seem impractical since the cardinality of its support is N^{T+1}. Nevertheless, for additive functionals of the form

$$S_{T,r}(x_{0:T}) = \sum_{t=r}^T h_t(x_{t-r:t}) , \quad (6)$$

where r is a non negative integer and $\{h_t\}_{t=r}^T$ is a family of bounded measurable functions on $\mathbb{X}^{r+1}$, the complexity of the FFBS algorithm is reduced to $O(N^{r+2})$. Furthermore, the smoothing of such functions can be computed forward in time as shown in [5]. This forward algorithm is exactly the one presented in [10] as an alternative to the use of the path-space method. Therefore, the results outlined in Section 3 hold for this method and confirm the conjecture mentioned in [10].

2.2 The forward filtering backward simulation algorithm

We now consider an algorithm whose complexity grows only linearly with the number of particles for any functional on $\mathbb{X}^{T+1}$. For any $t \in \{1, \ldots, T\}$, we define

$$\mathcal{F}_t^N \stackrel{\text{def}}{=} \sigma\left\{(\xi_s^{N,i}, \omega_s^{N,i}); 0 \le s \le t, 1 \le i \le N\right\} .$$

The transition probabilities $\{\Lambda_t^N\}_{t=0}^{T-1}$ defined in (4) induce an inhomogeneous Markov chain $\{J_u\}_{u=0}^T$ evolving backward in time as follows. At time T, the random index J_T is drawn from the set $\{1, \ldots, N\}$ with probability proportional to $(\omega_T^{N,1}, \ldots, \omega_T^{N,N})$. For any $t \in \{0, \ldots, T-1\}$, the index J_t is sampled in the set $\{1, \ldots, N\}$ according to $\Lambda_t^N(J_{t+1}, \cdot)$. The joint distribution of $J_{0:T}$ is therefore given, for $j_{0:T} \in \{1, \ldots, N\}^{T+1}$, by

$$\mathbb{P}\left[J_{0:T} = j_{0:T} \,\middle|\, \mathcal{F}_T^N\right] = \frac{\omega_T^{N,j_T}}{\Omega_T^N} \Lambda_{T-1}^N(j_T, j_{T-1}) \ldots \Lambda_0^N(j_1, j_0) . \quad (7)$$

Thus, the FFBS estimator (3) of the fixed-interval smoothing distribution may be written as the conditional expectation

$$\phi_{0:T|T}^N[h] = \mathbb{E}\left[h\left(\xi_0^{N,J_0}, \ldots, \xi_T^{N,J_T}\right) \middle| \mathcal{F}_T^N \right] ,$$

where h is a bounded measurable function on $\mathbb{X}^{T+1}$. We may therefore construct an unbiased estimator of the FFBS estimator given by

$$\widetilde{\phi}_{0:T|T}^N[h] = N^{-1} \sum_{\ell=1}^{N} h\left(\xi_0^{N,J_0^\ell}, \ldots, \xi_T^{N,J_T^\ell}\right) , \tag{8}$$

where $\{J_{0:T}^\ell\}_{\ell=1}^N$ are N paths drawn independently given $\mathcal{F}_T^N$ according to (7) and where h is a bounded measurable function on $\mathbb{X}^{T+1}$. This practical estimator was introduced in [12] (Algorithm 1, p. 158). An implementation of this estimator whose complexity grows linearly in N is introduced in [8].

3 Non-asymptotic deviation inequalities

In this Section, the L_q-mean error bounds and exponential deviation inequalities of the FFBS and FFBSi algorithms are established for additive functionals of the form (6). Our results are established under the following assumptions.

A1 (i) There exists $(\sigma_-, \sigma_+) \in (0,\infty)^2$ such that $\sigma_- < \sigma_+$ and for any $(x,x') \in \mathbb{X}^2$, $\sigma_- \leq m(x,x') \leq \sigma_+$ and we set $\rho \stackrel{\text{def}}{=} 1 - \sigma_-/\sigma_+$.

(ii) There exists $c_- \in \mathbb{R}_+^*$ such that $\int \chi(\mathrm{d}x) g_0(x) \geq c_-$ and for any $t \in \mathbb{N}^*$, $\inf_{x\in\mathbb{X}} \int M(x,\mathrm{d}x') g_t(x') \geq c_-$.

A2 (i) For all $t \geq 0$ and all $x \in \mathbb{X}$, $g_t(x) > 0$.

(ii) $\sup_{t\geq 0} |g_t|_\infty < \infty$.

A3 $\sup_{t\geq 1} |\vartheta_t|_\infty < \infty$, $\sup_{t\geq 0} |p_t|_\infty < \infty$ and $\sup_{t\geq 0} |\omega_t|_\infty < \infty$ where

$$\omega_0(x) \stackrel{\text{def}}{=} \frac{d\chi}{d\rho_0}(x) g_0(x), \quad \omega_t(x,x') \stackrel{\text{def}}{=} \frac{m(x,x') g_t(x')}{\vartheta_t(x) p_t(x,x')}, \forall t \geq 1 .$$

Assumptions A1 and A2 give bounds for the model and assumption A3 for quantities related to the algorithm. A1(i), referred to as the strong mixing condition, is crucial to derive time-uniform exponential deviation inequalities and a time-uniform bound of the variance of the marginal smoothing distribution (see [7] and [8]). For all function h from a space E to $\mathbb{R}$, $\mathrm{osc}(h)$ is defined by:

$$\mathrm{osc}(h) \stackrel{\text{def}}{=} \sup_{(z,z')\in \mathrm{E}^2} |h(z) - h(z')| .$$

Theorem 1. *Assume A1–3. For all $q \geq 2$, there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T < \infty$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$,*

$$\left\|\phi_{0:T|T}^N\left[S_{T,r}\right] - \phi_{0:T|T}\left[S_{T,r}\right]\right\|_q \leq \frac{C}{\sqrt{N}}\Upsilon_{r,T}^N\left(\sum_{s=r}^T \mathrm{osc}(h_s)^2\right)^{1/2} ,$$

where $S_{T,r}$ is defined by (6)*, $\phi_{0:T|T}^N$ is defined by* (3) *and where*

$$\Upsilon_{r,T}^N \stackrel{\mathrm{def}}{=} \sqrt{r+1}\left(\sqrt{1+r} \wedge \sqrt{T-r+1} + \frac{\sqrt{1+r}\sqrt{T-r+1}}{\sqrt{N}}\right) .$$

Similarly,

$$\left\|\widetilde{\phi}_{0:T|T}^N\left[S_{T,r}\right] - \phi_{0:T|T}\left[S_{T,r}\right]\right\|_q \leq \frac{C}{\sqrt{N}}\Upsilon_{r,T}^N\left(\sum_{s=r}^T \mathrm{osc}(h_s)^2\right)^{1/2} ,$$

where $\widetilde{\phi}_{0:T|T}^N$ is defined by (8)*.*

Remark 2. In the particular cases where $r = 0$ and $r = T$, $\Upsilon_{0,T}^N = 1+\sqrt{T+1/N}$ and $\Upsilon_{T,T}^N = \sqrt{T+1}(1+\sqrt{T+1/N})$. Then, Theorem 1 gives

$$\left\|\phi_{0:T|T}^N\left[S_{T,0}\right] - \phi_{0:T|T}\left[S_{T,0}\right]\right\|_q \leq C\frac{\left(\sum_{s=0}^T \mathrm{osc}(h_s)^2\right)^{1/2}}{\sqrt{N}}\left(1+\sqrt{\frac{T+1}{N}}\right) ,$$

and

$$\left\|\phi_{0:T|T}^N\left[S_{T,T}\right] - \phi_{0:T|T}\left[S_{T,T}\right]\right\|_q \leq C\sqrt{\frac{T+1}{N}}\left(1+\sqrt{\frac{T+1}{N}}\right)\mathrm{osc}(h_T)^2 .$$

As stated in Section 1, theses bounds improve the results given in [5] for the FFBS estimator.

Remark 3. The dependence on $1/\sqrt{N}$ is hardly surprising. Under the stated strong mixing condition, it is known that the L_q-norm of the marginal smoothing estimator $\phi_{t-r:t|T}^N[h]$, $t \in \{r, \ldots, T\}$ is uniformly bounded in time by $\left\|\phi_{t-r:t|T}^N[h]\right\|_q \leq C\mathrm{osc}(h)N^{-1/2}$ (where C depends only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$). The dependence in $\sqrt{T}$ instead of T reflects the forgetting property of the filter and the backward smoother. As for $r \leq s < t \leq T$, the estimators $\phi_{s-r:s|T}^N[h_s]$ and $\phi_{t-r:t|T}^N[h_t]$ become asymptotically independent as $(t-s)$ gets large, the L_q-norm of the sum $\sum_{t=r}^T \phi_{t-r:t|T}^N[h_t]$ scales as the sum of a mixing sequence (see [3]).

Remark 4. It is easy to see that the scaling in $\sqrt{T/N}$ cannot in general be improved. Assume that the kernel m satisfies $m(x, x') = m(x')$ for all $(x, x') \in \mathbb{X} \times \mathbb{X}$. In this case, for any $t \in \{0, \ldots, T\}$, the filtering distribution is

$$\phi_t[h_t] = \frac{\int m(x) g_t(x) h_t(x) \mathrm{d}x}{\int m(x) g_t(x) \mathrm{d}x} ,$$

and the backward kernel is the identity kernel. Hence, the fixed-interval smoothing distribution coincides with the filtering distribution. If we assume that we apply the bootstrap filter for which $p_s(x, x') = m(x')$ and $\vartheta_s(x) = 1$, the estimators $\{\phi^N_{t|T}[h_t]\}_{t\in\{0,\ldots,T\}}$ are independent random variables corresponding to importance sampling estimators. It is easily seen that

$$\left\| \sum_{t=0}^{T} \phi^N_t[h_t] - \phi_t[h_t] \right\|_q \leq C \max_{0\leq t\leq T} \{\mathrm{osc}(h_t)\} \sqrt{\frac{T}{N}} .$$

Remark 5. The independent case also clearly illustrates why the path-space methods are sub-optimal (see also [1] for a discussion). When applied to the independent case (for all $(x, x') \in \mathbb{X} \times \mathbb{X}$, $m(x, x') = m(x')$ and $p_s(x, x') = m(x')$), the asymptotic variance of the path-space estimators is given in [4] by

$$\begin{aligned}
&\Gamma_{0:T|T}[S_{T,0}] \\
&\quad \stackrel{\mathrm{def}}{=} \sum_{t=0}^{T-1} \frac{m(g_T^2)}{m(g_T)^2} \frac{m(g_t[h_t - \phi_t(h_t)]^2)}{m(g_t)} + \frac{m(g_T^2[h_T - \phi_T(h_T)]^2)}{m(g_T)^2} \\
&\qquad + \sum_{t=1}^{T-1} \left\{ \sum_{s=0}^{t-1} \frac{m(g_t^2)}{m(g_t)^2} \frac{m(g_s[h_s - \phi_s(h_s)]^2)}{m(g_s)} + \frac{m(g_t^2[h_t - \phi_t(h_t)]^2)}{m(g_t)^2} \right\} \\
&\qquad + \frac{\chi(g_0^2[h_0 - \phi_0(h_0)]^2)}{\chi(g_0)^2} .
\end{aligned}$$

The asymptotic variance thus increases as T^2 and hence, under the stated assumptions, the variance of the path-space methods is of order T^2/N. It is believed (and proved in some specific scenarios) that the same scaling holds for path-space methods for non-degenerated Markov kernel (the result has been formally established for strongly mixing kernel under the assumption that σ_-/σ_+ is sufficiently close to 1).

We provide below a brief outline of the main steps of the proofs (a detailed proof is given in Section 5). Following [8], the proofs rely on a decomposition of the smoothing error. For all $0 \leq t \leq T$ and all bounded and measurable function h on $\mathbb{X}^{T+1}$ define the kernel $\mathrm{L}_{t,T} : \mathbb{X}^{t+1} \times \mathcal{X}^{\otimes T+1} \to [0, 1]$ by

$$\mathrm{L}_{t,T} h(x_{0:t}) \stackrel{\mathrm{def}}{=} \int \prod_{u=t+1}^{T} M(x_{u-1}, \mathrm{d}x_u) g_u(x_u) h(x_{0:T}) .$$

The fixed-interval smoothing distribution may then be expressed, for all bounded and measurable function h on $\mathbb{X}^{T+1}$, by

$$\phi_{0:T|T}[h] = \frac{\phi_{0:t|t}\left[\mathrm{L}_{t,T}h\right]}{\phi_{0:t|t}\left[\mathrm{L}_{t,T}\mathbf{1}\right]} ,$$

and this suggests to decompose the smoothing error as follows

$$\begin{aligned}\Delta_T^N[h] &\stackrel{\text{def}}{=} \phi_{0:T|T}^N[h] - \phi_{0:T|T}[h] \\ &= \sum_{t=0}^{T} \frac{\phi_{0:t|t}^N\left[\mathrm{L}_{t,T}h\right]}{\phi_{0:t|t}^N\left[\mathrm{L}_{t,T}\mathbf{1}\right]} - \frac{\phi_{0:t-1|t-1}^N\left[\mathrm{L}_{t-1,T}h\right]}{\phi_{0:t-1|t-1}^N\left[\mathrm{L}_{t-1,T}\mathbf{1}\right]} ,\end{aligned} \tag{9}$$

where we used the convention

$$\frac{\phi_{0:-1|-1}^N\left[\mathrm{L}_{-1,T}h\right]}{\phi_{0:-1|-1}^N\left[\mathrm{L}_{-1,T}\mathbf{1}\right]} = \frac{\phi_0\left[\mathrm{L}_{0,T}h\right]}{\phi_0\left[\mathrm{L}_{0,T}\mathbf{1}\right]} = \phi_{0:T|T}[h] .$$

Furthermore, for all $0 \leq t \leq T$,

$$\begin{aligned}\phi_{0:t|t}^N\left[\mathrm{L}_{t,T}h\right] &= \int \phi_{0:t|t}^N(\mathrm{d}x_{0:t})\mathrm{L}_{t,T}h(x_{0:t}) \\ &= \int \phi_t^N(\mathrm{d}x_t)\mathrm{B}_{\phi_{t-1}^N}(x_t, \mathrm{d}x_{t-1}) \cdots \mathrm{B}_{\phi_0^N}(x_1, \mathrm{d}x_0)\mathrm{L}_{t,T}h(x_{0:t}) \\ &= \int \phi_t^N(\mathrm{d}x_t)\mathcal{L}_{t,T}^N h(x_t) ,\end{aligned}$$

where $\mathcal{L}_{t,T}^N$ and $\mathcal{L}_{t,T}$ are two kernels on $\mathbb{X} \times \mathcal{X}^{\otimes(T+1)}$ defined for all $x_t \in \mathbb{X}$ by

$$\mathcal{L}_{t,T}h(x_t) \stackrel{\text{def}}{=} \int \mathrm{B}_{\phi_{t-1}}(x_t, \mathrm{d}x_{t-1}) \cdots \mathrm{B}_{\phi_0}(x_1, \mathrm{d}x_0)\mathrm{L}_{t,T}h(x_{0:t}) \tag{10}$$

$$\mathcal{L}_{t,T}^N h(x_t) \stackrel{\text{def}}{=} \int \mathrm{B}_{\phi_{t-1}^N}(x_t, \mathrm{d}x_{t-1}) \cdots \mathrm{B}_{\phi_0^N}(x_1, \mathrm{d}x_0)\mathrm{L}_{t,T}h(x_{0:t}) . \tag{11}$$

For all $1 \leq t \leq T$ we can write

$$\begin{aligned}\frac{\phi_{0:t|t}^N[\mathrm{L}_{t,T}h]}{\phi_{0:t|t}^N[\mathrm{L}_{t,T}\mathbf{1}]} - \frac{\phi_{0:t-1|t-1}^N[\mathrm{L}_{t-1,T}h]}{\phi_{0:t-1|t-1}^N[\mathrm{L}_{t-1,T}\mathbf{1}]} &= \frac{\phi_t^N[\mathcal{L}_{t,T}^N h]}{\phi_t^N[\mathcal{L}_{t,T}^N\mathbf{1}]} - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N\mathbf{1}]} \\ &= \frac{1}{\phi_t^N[\mathcal{L}_{t,T}^N\mathbf{1}]}\left(\phi_t^N[\mathcal{L}_{t,T}^N h] - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N\mathbf{1}]}\phi_t^N[\mathcal{L}_{t,T}^N\mathbf{1}]\right) ,\end{aligned}$$

and then,

$$\Delta_T^N[h] = \sum_{t=0}^{T} \frac{N^{-1}\sum_{\ell=1}^N \omega_t^{N,\ell} G_{t,T}^N h(\xi_t^{N,\ell})}{N^{-1}\sum_{\ell=1}^N \omega_t^{N,\ell}\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})} , \tag{12}$$

with $G^N_{t,T}$ is a kernel on $\mathbb{X} \times \mathcal{X}^{\otimes(T+1)}$ defined, for all $x_t \in \mathbb{X}$ and all bounded and measurable function h on $\mathbb{X}^{T+1}$, by

$$G^N_{t,T} h(x_t) \stackrel{\text{def}}{=} \mathcal{L}^N_{t,T} h(x_t) - \frac{\phi^N_{t-1}[\mathcal{L}^N_{t-1,T} h]}{\phi^N_{t-1}[\mathcal{L}^N_{t-1,T}\mathbf{1}]} \mathcal{L}^N_{t,T}\mathbf{1}(x_t) \; ,$$

where, by the same convention as above,

$$G^N_{0,T} h(x_0) \stackrel{\text{def}}{=} \mathrm{L}_{0,T} h(x_0) - \frac{\phi_0[\mathcal{L}_{0,T} h]}{\phi_0[\mathcal{L}_{0,T}\mathbf{1}]} \mathcal{L}_{0,T}\mathbf{1}(x_0) \; .$$

Two families of random variables $\left\{C^N_{t,T}(f)\right\}_{t=0}^T$ and $\left\{D^N_{t,T}(f)\right\}_{t=0}^T$ are now introduced to transform (12) into a suitable decomposition to compute an upper bound for the L_q-mean error. As shown in Lemma 1, the random variables $\{\omega^{N,\ell}_t G^N_{t,T} f(\xi^{N,\ell}_t)\}_{\ell=1}^N$ are centered given $\mathcal{F}^N_{t-1}$. The idea is to replace $N^{-1}\sum_{\ell=1}^N \omega^{N,\ell}_t \mathcal{L}_{t,T}\mathbf{1}(\xi^{N,\ell}_t)$ in (12) by its conditional expectation given $\mathcal{F}^N_{t-1}$ to get a martingale difference. This conditional expectation is computed using the following intermediate result. For any measurable function h on $\mathbb{X}$ and any $t \in \{0, \ldots, T\}$,

$$\mathbb{E}\left[\omega^{N,1}_t h(\xi^{N,1}_t) \middle| \mathcal{F}^N_{t-1}\right] = \frac{\phi^N_{t-1}\left[M g_t h\right]}{\phi^N_{t-1}[\vartheta_t]} \; . \tag{13}$$

Indeed,

$$\begin{aligned}
&\mathbb{E}\left[\omega^{N,1}_t h(\xi^{N,1}_t) \middle| \mathcal{F}^N_{t-1}\right] \\
&= \mathbb{E}\left[\frac{m(\xi^{N,I^{N,1}_t}_{t-1}, \xi^{N,1}_t) g_t(\xi^{N,1}_t)}{\vartheta_t(\xi^{N,I^{N,1}_t}_{t-1}) p_t(\xi^{N,I^{N,1}_t}_{t-1}, \xi^{N,1}_t)} h(\xi^{N,1}_t) \middle| \mathcal{F}^N_{t-1}\right] \\
&= \left(\sum_{i=1}^N \omega^{N,i}_{t-1} \vartheta_t(\xi^{N,i}_{t-1})\right)^{-1} \sum_{i=1}^N \int \omega^{N,i}_{t-1} \vartheta_t(\xi^{N,i}_{t-1}) p_t(\xi^{N,i}_{t-1}, x) \frac{M(\xi^{N,i}_{t-1}, \mathrm{d}x) g_t(x)}{\vartheta_t(\xi^{N,i}_{t-1}) p_t(\xi^{N,i}_{t-1}, x)} h(x) \\
&= \left(\sum_{i=1}^N \omega^{N,i}_{t-1} \vartheta_t(\xi^{N,i}_{t-1})\right)^{-1} \sum_{i=1}^N \int \omega^{N,i}_{t-1} M(\xi^{N,i}_{t-1}, \mathrm{d}x) g_t(x) h(x) \\
&= \frac{\phi^N_{t-1}\left[M g_t h\right]}{\phi^N_{t-1}[\vartheta_t]} \; .
\end{aligned}$$

This result, applied with the function $h = \mathcal{L}_{t,T}\mathbf{1}$, yields

$$\mathbb{E}\left[\omega^{N,1}_t \mathcal{L}_{t,T}\mathbf{1}(\xi^{N,1}_t) \middle| \mathcal{F}^N_{t-1}\right] = \frac{\phi^N_{t-1}\left[M g_t \mathcal{L}_{t,T}\mathbf{1}\right]}{\phi^N_{t-1}[\vartheta_t]} = \frac{\phi^N_{t-1}\left[\mathcal{L}_{t-1,T}\mathbf{1}\right]}{\phi^N_{t-1}[\vartheta_t]} \; .$$

For any $0 \leq t \leq T$, define for all bounded and measurable function h on $\mathbb{X}^{T+1}$,

$$D_{t,T}^N(h) \stackrel{\text{def}}{=} \mathbb{E}\left[\omega_t^{N,1}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\middle|\mathcal{F}_{t-1}^N\right]^{-1} N^{-1}\sum_{\ell=1}^N \omega_t^{N,\ell}\frac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \tag{14}$$
$$= \frac{\phi_{t-1}^N[\vartheta_t]}{\phi_{t-1}^N\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]} N^{-1}\sum_{\ell=1}^N \omega_t^{N,\ell}\frac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \ ,$$

$$C_{t,T}^N(h) \stackrel{\text{def}}{=} \left[\frac{1}{N^{-1}\sum_{i=1}^N \omega_t^{N,i}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,i})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}} - \frac{\phi_{t-1}^N[\vartheta_t]}{\phi_{t-1}^N\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]}\right]$$
$$\times N^{-1}\sum_{\ell=1}^N \omega_t^{N,\ell}\frac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \ . \tag{15}$$

Using these notations, (12) can be rewritten as follows:

$$\Delta_T^N[h] = \sum_{t=0}^T D_{t,T}^N(h) + \sum_{t=0}^T C_{t,T}^N(h) \ . \tag{16}$$

For any $q \geq 2$, the derivation of the upper bound relies on the triangle inequality:

$$\left\|\Delta_T^N[S_{T,r}]\right\|_q \leq \left\|\sum_{t=0}^T D_{t,T}^N(S_{T,r})\right\|_q + \sum_{t=0}^T \left\|C_{t,T}^N(S_{T,r})\right\|_q \ ,$$

where $S_{T,r}$ is defined in (6). The proof for the FFBS estimator $\phi_{0:T|T}^N$ is completed by using Proposition 1 and Proposition 2. According to (16), the smoothing error can be decomposed into a sum of two terms which are considered separately. The first one is a martingale whose L_q-mean error is upper-bounded by $\sqrt{(T+1)/N}$ as shown in Proposition 1. The second one is a sum of products, L_q-norm of which being bounded by $1/N$ in Proposition 2.

The end of this section is devoted to the exponential deviation inequality for the error $\Delta_T^N[S_{T,r}]$ defined by (9). We use the decomposition of $\Delta_T^N[S_{T,r}]$ obtained in (16) leading to a similar dependence on the ratio $(T+1)/N$. The martingale term $D_{t,T}^N(S_{T,r})$ is dealt with using the Azuma-Hoeffding inequality while the term $C_{t,T}^N(S_{T,r})$ needs a specific Hoeffding-type inequality for ratio of random variables.

Theorem 2. *Assume A1–3. There exists a constant C (depending only on σ_-, σ_+, r, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T < \infty$, any $N \geq 1$, any $\varepsilon > 0$, any integer r, and any bounded and measurable functions $\{h_s\}_{s=r}^T$,*

$$\mathbb{P}\left\{\left|\phi_{0:T|T}\left[S_{T,r}\right] - \phi_{0:T|T}^N\left[S_{T,r}\right]\right| > \varepsilon\right\}$$
$$\leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right) + 8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right) \ ,$$

where $S_{T,r}$ is defined by (6)*, $\phi^N_{0:T|T}$ is defined by* (3) *and where*

$$\Theta_{r,T} \stackrel{\text{def}}{=} (1+r)\left\{(1+r) \wedge (T-r+1)\right\} . \tag{17}$$

Similarly,

$$\mathbb{P}\left\{\left|\phi_{0:T|T}\left[S_{T,r}\right] - \widetilde{\phi}^N_{0:T|T}\left[S_{T,r}\right]\right| > \varepsilon\right\} \leq 4\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^{T}\operatorname{osc}(h_s)^2}\right) + 8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^{T}\operatorname{osc}(h_s)}\right) ,$$

where $\widetilde{\phi}^N_{0:T|T}$ is defined by (8)*.*

4 Monte-Carlo Experiments

In this section, the performance of the FFBSi algorithm is evaluated through simulations and compared to the path-space method.

4.1 Linear gaussian model

Let us consider the following model:

$$\begin{cases} X_{t+1} &= \phi X_t + \sigma_u U_t \ , \\ Y_t &= X_t + \sigma_v V_t \ , \end{cases}$$

where X_0 is a zero-mean random variable with variance $\frac{\sigma_u^2}{1-\phi^2}$, $\{U_t\}_{t\geq 0}$ and $\{V_t\}_{t\geq 0}$ are two sequences of independent and identically distributed standard gaussian random variables (independent from X_0). The parameters $(\phi, \sigma_u, \sigma_v)$ are assumed to be known. Observations were generated using $\phi = 0.9$, $\sigma_u = 0.6$ and $\sigma_v = 1$. Table 1 provides the empirical variance of the estimation of the unnormalized smoothed additive functional $\mathcal{I}_T \stackrel{\text{def}}{=} \sum_{t=0}^{T} \mathbb{E}\left[X_t | Y_{0:T}\right]$ given by the path-space and the FFBSi methods over 250 independent Monte Carlo experiments. We display in Figure 1 the empirical variance for different values of N as a function of T for both estimators. These estimates are represented by dots and a linear regression (resp. quadratic regression) is also provided for the FFBSi algorithm (resp. for the path-space method).

In Figure 2 the FFBSi algorithm is compared to the path-space method to compute the smoothed value of the empirical mean $(T+1)^{-1}\mathcal{I}_T$. For the purpose of comparison, this quantity is computed using the Kalman smoother. We display in Figure 2 the box and whisker plots of the estimations obtained with 100 independent Monte Carlo experiments. The FFBSi algorithm clearly outperforms the other method for comparable computational costs. In Table 2, the mean CPU times over the 100 runs of the two methods are given as a function of the number of particles (for $T = 500$ and $T = 1000$).

Table 1: Empirical variance for different values of T and N.

Path-space

T \ N	300	500	750	1000	1500	5000	10000	15000	20000
300	137.8	119.4	63.7	46.1	36.2	12.8	7.1	3.8	3.0
500	290.0	215.3	192.5	161.9	80.3	30.1	14.9	11.3	7.4
750	474.9	394.5	332.9	250.5	206.8	71.0	35.6	24.4	21.7
1000	673.7	593.2	505.1	483.2	326.4	116.4	70.8	37.9	34.6
1500	1274.6	1279.7	916.7	804.7	655.1	233.9	163.1	89.7	80.0

FFBSi

T \ N	300	500	750	1000	1500
300	5.1	3.1	2.3	1.4	1.0
500	9.7	5.1	3.7	2.6	2.2
750	11.2	7.1	4.9	3.7	2.6
1000	16.5	10.5	6.7	5.1	3.4
1500	25.6	14.1	7.8	6.8	5.1

Table 2: Average CPU time to compute the smoothed value of the empirical mean in the LGM

$T = 500$	FFBSi	Path-space method		
N	500	500	5000	10000
CPU time (s)	4.87	0.24	2.47	4.65

$T = 1000$	FFBSi	Path-space method		
N	1000	1000	10000	20000
CPU time (s)	16.5	0.9	8.5	17.2

4.2 Stochastic Volatility Model

Stochastic volatility models (SVM) have been introduced to provide better ways of modeling financial time series data than ARCH/GARCH models ([14]). We consider the elementary SVM model introduced by [14]:

$$\begin{cases} X_{t+1} = \phi X_t + \sigma U_{t+1} \ , \\ Y_t = \beta \mathrm{e}^{\frac{X_t}{2}} V_t \ , \end{cases}$$

where X_0 is a zero-mean random variable with variance $\frac{\sigma_u^2}{1-\phi^2}$, $\{U_t\}_{t\geq 0}$ and $\{V_t\}_{t\geq 0}$ are two sequences of independent and identically distributed standard

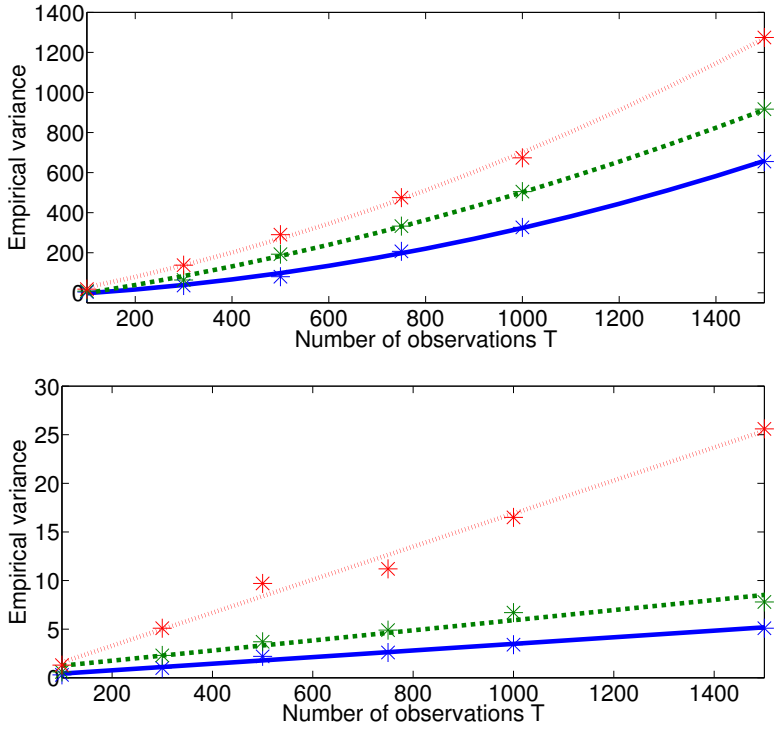


Figure 1: Empirical variance of the path-space (top) and FFBSi (bottom) for $N = 300$ (dotted line), $N = 750$ (dashed line) and $N = 1500$ (bold line).

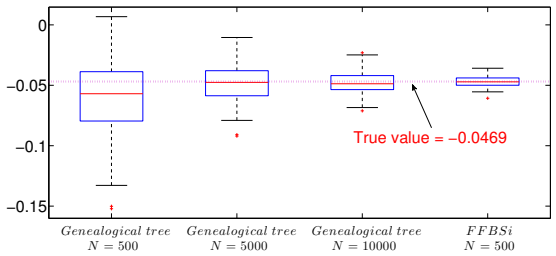
gaussian random variables (independent from X_0). This model was used to generate simulated data with parameters $(\phi = 0.3, \sigma = 0.5, \beta = 1)$ assumed to be known in the following experiments. The empirical variance of the estimation of $\mathcal{I}_T$ given by the path-space and the FFBSi methods over 250 independent Monte Carlo experiments is displayed in Table 3. We display in Figure 3 the empirical variance for different values of N as a function of T for both estimators.

5 Proof of Theorem 1

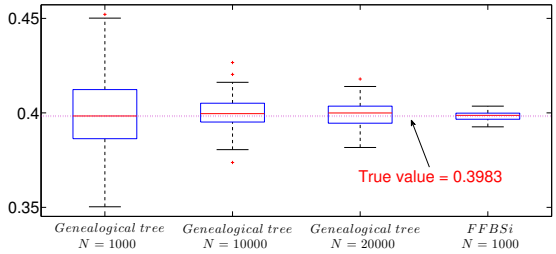
We preface the proof of Proposition 1 by the following Lemma:

Lemma 1. *Under assumptions A1–3, we have, for any $t \in \{0, \ldots, T\}$ and any measurable function h on $\mathbb{X}^{T+1}$:*

(i) The random variables $\left\{ \omega_t^{N,\ell} \dfrac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right\}_{\ell=1}^{N}$ *are, for all $N \in \mathbb{N}$:*



(a) Time $T = 500$



(b) Time $T = 1000$

Figure 2: Computation of smoothed additive functionals in a linear gaussian model. The variance of the estimation given by the FFBSi algorithm is the smallest one in both cases.

Table 3: Empirical variance for different values of T and N in the SVM.

Path-space method

T \ N	300	500	750	1000	1500	5000	10000	15000	20000
300	52.7	33.7	22.0	17.8	12.3	3.8	2.0	1.4	1.2
500	116.3	84.8	64.8	53.5	30.7	11.4	6.8	4.1	2.8
750	184.7	187.6	134.2	120.0	65.8	29.1	12.8	7.3	7.7
1000	307.7	240.4	244.7	182.8	133.2	43.6	24.5	15.6	11.6
1500	512.1	487.5	445.5	359.9	249.5	90.9	52.0	32.6	29.3

FFBSi

T \ N	300	500	750	1000	1500
300	1.2	0.6	0.5	0.4	0.2
500	2.1	1.2	0.8	0.6	0.4
750	3.7	1.8	1.4	0.9	0.6
1000	4.0	2.7	1.8	1.3	0.9
1500	7.3	3.8	3.1	1.6	1.4

(a) conditionally independent and identically distributed given $\mathcal{F}_{t-1}^N$,

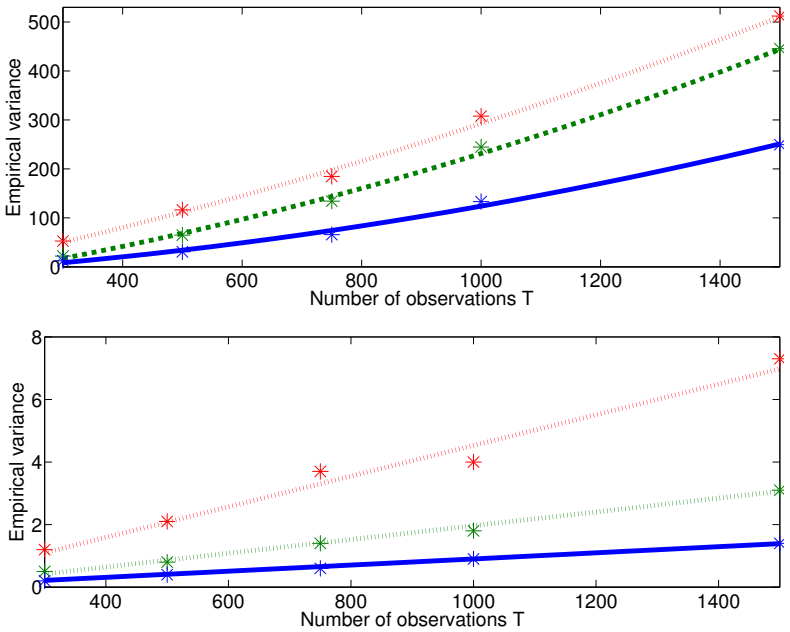


Figure 3: Empirical variance of the path-space (top) and FFBSi (bottom) for $N = 300$ (dotted line), $N = 750$ (dashed line) and $N = 1500$ (bold line) in the SVM.

(b) centered conditionally to $\mathcal{F}_{t-1}^N$.

where $G_{t,T}^N h$ *is defined in* (3) *and* $\mathcal{L}_{t,T}^N$ *is defined in* (11).

(ii) For any integers r, t *and* N*:*

$$\left| \frac{G_{t,T}^N S_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right| \leq \sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \mathrm{osc}(h_s) \ , \tag{18}$$

where $S_{T,r}$ *and* ρ *are respectively defined in* (6) *and in A1*(i).

(iii) For all $x \in \mathbb{X}$, $\dfrac{\mathcal{L}_{t,T}\mathbf{1}(x)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \geq \dfrac{\sigma_-}{\sigma_+}$ *and* $\dfrac{\mathcal{L}_{t-1,T}\mathbf{1}(x)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \geq c_- \dfrac{\sigma_-}{\sigma_+}$.

Proof. The proof of (i) is given by [8, Lemma 3].

Proof of (ii). Let $\Pi_{s-r:s,T}$ be the operator which associates to any bounded and measurable function h on $\mathbb{X}^{r+1}$ the function $\Pi_{s-r:s,T} h$ given, for any $(x_0, \ldots, x_T) \in \mathbb{X}^{T+1}$, by

$$\Pi_{s-r:s,T} h(x_{0:T}) \stackrel{\text{def}}{=} h(x_{s-r:s}) \ .$$

Then, we may write $S_{T,r} = \sum_{s=r}^{T} \Pi_{s-r:s,T} h_s$ and $G_{t,T}^N S_{T,r} = \sum_{s=r}^{T} G_{t,T}^N \Pi_{s-r:s,T} h_s$. By (3), we have

$$\frac{G_{t,T}^N \Pi_{s-r:s,T} h_s(x_t)}{\mathcal{L}_{t,T}^N \mathbf{1}(x_t)} = \frac{\mathcal{L}_{t,T}^N \Pi_{s-r:s,T} h_s(x_t)}{\mathcal{L}_{t,T}^N \mathbf{1}(x_t)} - \frac{\phi_{t-1}^N [\mathcal{L}_{t-1,T}^N \Pi_{s-r:s,T} h_s]}{\phi_{t-1}^N [\mathcal{L}_{t-1,T}^N \mathbf{1}]} ,$$

and, following the same lines as in [8, Lemma 10],

$$|G_{t,T}^N \Pi_{s-r:s,T} h_s|_\infty \leq \rho^{s-r-t} \mathrm{osc}(h_s) |\mathcal{L}_{t,T} \mathbf{1}|_\infty \quad \text{if} \quad t \leq s - r ,$$

$$|G_{t,T}^N \Pi_{s-r:s,T} h_s|_\infty \leq \rho^{t-s} \mathrm{osc}(h_s) |\mathcal{L}_{t,T} \mathbf{1}|_\infty \quad \text{if} \quad t > s ,$$

where ρ is defined in A1(i). Furthermore, for any $s - r < t \leq s$,

$$|G_{t,T}^N \Pi_{s-r:s,T} h_s|_\infty \leq \mathrm{osc}(h_s) |\mathcal{L}_{t,T} \mathbf{1}|_\infty ,$$

which shows (ii).

Proof of (iii). From the definition (10), for all $x \in \mathbb{X}$ and all $t \in \{1, \ldots, T\}$,

$$\mathcal{L}_{t,T} \mathbf{1}(x) = \int m(x, x_{t+1}) g_{t+1}(x_{t+1}) \prod_{u=t+2}^{T} M(x_{u-1}, \mathrm{d}x_u) g_u(x_u) \lambda(\mathrm{d}x_{t+1}) ,$$

hence, by assumption A1,

$$|\mathcal{L}_{t,T} \mathbf{1}|_\infty \leq \sigma_+ \int g_{t+1}(x_{t+1}) \mathcal{L}_{t+1,T} \mathbf{1}(x_{t+1}) \lambda(\mathrm{d}x_{t+1})$$
$$\mathcal{L}_{t,T} \mathbf{1}(x) \geq \sigma_- \int g_{t+1}(x_{t+1}) \mathcal{L}_{t+1,T} \mathbf{1}(x_{t+1}) \lambda(\mathrm{d}x_{t+1}) ,$$

which concludes the proof of the first statement. By construction, for any $x \in \mathbb{X}$ and any $t \in \{1, \ldots, T\}$,

$$\mathcal{L}_{t-1,T} \mathbf{1}(x) = \int M(x, \mathrm{d}x') g_t(x') \mathcal{L}_{t,T} \mathbf{1}(x') ,$$

and then, by assumption A1,

$$\frac{\mathcal{L}_{t-1,T} \mathbf{1}(x)}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} = \int M(x, \mathrm{d}x') g_t(x') \frac{\mathcal{L}_{t,T} \mathbf{1}(x')}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \geq c_- \frac{\sigma_-}{\sigma_+} .$$

□

Proposition 1. *Assume A1–3. For all $q \geq 2$, there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t \geq 1} |\vartheta_t|_\infty$ and $\sup_{t \geq 0} |\omega_t|_\infty$) such that for any $T < \infty$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\left\| \sum_{t=0}^{T} D_{t,T}^N (S_{T,r}) \right\|_q \leq \frac{C}{\sqrt{N}} \sqrt{1+r} \left(\sqrt{1+r} \wedge \sqrt{T-r+1} \right) \left(\sum_{s=r}^{T} \mathrm{osc}(h_s)^2 \right)^{1/2} , \tag{19}$$

where $D_{t,T}^N$ is defined in (14).

Proof. Since $\left\{D^N_{t,T}(S_{T,r})\right\}_{0\le t\le T}$ is a is a forward martingale difference and $q \ge 2$, Burkholder's inequality (see [13, Theorem 2.10, page 23]) states the existence of a constant C depending only on q such that:

$$\mathbb{E}\left[\left|\sum_{t=0}^{T} D^N_{t,T}(S_{T,r})\right|^q\right] \le C\mathbb{E}\left[\left|\sum_{t=0}^{T} D^N_{t,T}(S_{T,r})^2\right|^{\frac{q}{2}}\right] .$$

Moreover, by application of the last statement of Lemma 1(iii),

$$\frac{\phi^N_{t-1}[\vartheta_t]}{\phi^N_{t-1}\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]} \le \frac{\sigma_+ \sup_{t\ge 0}|\vartheta_t|_\infty}{\sigma_- c_-} ,$$

and thus,

$$\mathbb{E}\left[\left|\sum_{t=0}^{T} D^N_{t,T}(S_{T,r})^2\right|^{\frac{q}{2}}\right] \le \left(\frac{\sigma_+ \sup_{t\ge 0}|\vartheta_t|_\infty}{\sigma_- c_-}\right)^q \mathbb{E}\left[\left|\sum_{t=0}^{T}\left(N^{-1}\sum_{\ell=1}^{N} a^{N,\ell}_{t,T}\right)^2\right|^{\frac{q}{2}}\right] ,$$

where $a^{N,\ell}_{t,T} \stackrel{\text{def}}{=} \omega^{N,\ell}_t \frac{G^N_{t,T} S_{T,r}(\xi^{N,\ell}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}$. By the Minkowski inequality,

$$\left\|\sum_{t=0}^{T} D^N_{t,T}(S_{T,r})\right\|_q \le C\left\{\sum_{t=0}^{T}\left(\mathbb{E}\left[\left|N^{-1}\sum_{\ell=1}^{N} a^{N,\ell}_{t,T}\right|^q\right]\right)^{2/q}\right\}^{1/2} . \tag{20}$$

Since for any $t \ge 0$ the random variables $\left\{a^{N,\ell}_{t,T}\right\}_{\ell=1}^{N}$ are conditionally independent and centered conditionally to $\mathcal{F}^N_{t-1}$, using again the Burkholder and the Jensen inequalities we obtain

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{\ell=1}^{N} a^{N,\ell}_{t,T}\right|^q \middle| \mathcal{F}^N_{t-1}\right] &\le CN^{q/2-1}\sum_{\ell=1}^{N}\mathbb{E}\left[\left|a^{N,\ell}_{t,T}\right|^q \middle| \mathcal{F}^N_{t-1}\right] \\ &\le C\left[\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right]^q N^{q/2} ,\end{aligned} \tag{21}$$

where the last inequality comes from (18). Finally, by (20) and (21) we get

$$\left\|\sum_{t=0}^{T} D^N_{t,T}(S_{T,r})\right\|_q \le CN^{-1/2}\left\{\sum_{t=0}^{T}\left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right)^2\right\}^{1/2} .$$

By the Holder inequality, we have

$$\sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \mathrm{osc}(h_s)$$
$$\leq \left(\sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \right)^{1/2} \times \left(\sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \mathrm{osc}(h_s)^2 \right)^{1/2}$$
$$\leq C\sqrt{1+r} \left(\sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \mathrm{osc}(h_s)^2 \right)^{1/2} ,$$

which yields

$$\left\| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r}) \right\|_q \leq CN^{-1/2}(1+r) \left(\sum_{s=r}^{T} \mathrm{osc}(h_s)^2 \right)^{1/2} .$$

We obtain similarly

$$\left\| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r}) \right\|_q \leq CN^{-1/2}(1+r)^{1/2} \sum_{s=r}^{T} \mathrm{osc}(h_s) ,$$

which concludes the proof. □

Proposition 2. *Assume A1–3. For all $q \geq 2$, there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T < +\infty$, any $0 \leq t \leq T$, any integer r, and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\left\| C_{t,T}^N(S_{T,r}) \right\|_q \leq \frac{C}{N} \sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \mathrm{osc}(h_s) , \tag{22}$$

where $C_{t,T}^N$ is defined in (15).

Proof. According to (15), $C_{t,T}^N(S_{T,r})$ can be written

$$C_{t,T}^N(S_{T,r}) = U_{t,T}^N V_{t,T}^N W_{t,T}^N , \tag{23}$$

where

$$U_{t,T}^N = \frac{N^{-1} \sum_{\ell=1}^{N} \omega_t^{N,\ell} \frac{G_{t,T}^N S_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}}{N^{-1}\Omega_t^N} ,$$
$$V_{t,T}^N = N^{-1} \sum_{\ell=1}^{N} \left(\mathbb{E}\left[\omega_t^{N,1} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1} \right] - \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right) ,$$
$$W_{t,T}^N = \frac{N^{-1}\Omega_t^N}{\mathbb{E}\left[\omega_t^{N,1} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1} \right] N^{-1} \sum_{\ell=1}^{N} \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}} ,$$

and where Ω_t^N is defined by (5). Using the last statement of Lemma 1, we get the following bound:

$$\mathbb{E}\left[\omega_t^{N,1}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\middle|\mathcal{F}_{t-1}\right] = \frac{\phi_{t-1}^N\left[\mathcal{L}_{t-1,T}\mathbf{1}/|\mathcal{L}_{t,T}\mathbf{1}|_\infty\right]}{\phi_{t-1}^N[\vartheta_t]} \geq \frac{c_-\sigma_-}{|\vartheta_t|_\infty\,\sigma_+}$$

$$\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \geq \frac{\sigma_-}{\sigma_+}\,,$$

which implies

$$\left|W_{t,T}^N\right| \leq \left(\frac{\sigma_+}{\sigma_-}\right)^2\frac{|\vartheta_t|_\infty}{c_-}\,. \tag{24}$$

Then, $\left|C_{t,T}^N(S_{T,r})\right| \leq C\left|U_{t,T}^N\right|\left|V_{t,T}^N\right|$ and we can use the decomposition

$$U_{t,T}^N V_{t,T}^N = V_{t,T}^N\left[\frac{N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}}{\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]} + \frac{N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}}{\widetilde{\Omega}_t^N\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]}\left(\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]-\widetilde{\Omega}_t^N\right)\right]\,,$$

where $a_{t,T}^{N,\ell} \stackrel{\text{def}}{=} \omega_t^{N,\ell}\frac{G_{t,T}^N S_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}$ and $\widetilde{\Omega}_t^N \stackrel{\text{def}}{=} N^{-1}\Omega_t^N$. By (13), $\mathbb{E}\left[\omega_t^{N,1}\middle|\mathcal{F}_{t-1}^N\right] = \frac{\phi_{t-1}^N[Mg_t]}{\phi_{t-1}^N[\vartheta_t]}$ and then, by A1(ii), A3 and (18),

$$\frac{1}{\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]} \leq \frac{|\vartheta_t|_\infty}{c_-} \quad\text{and}\quad \frac{N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}}{\widetilde{\Omega}_t^N\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]} \leq C\frac{|\vartheta_t|_\infty}{c_-}\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\,.$$

Therefore, $\left|C_{t,T}^N(S_{T,r})\right| \leq C\left(C_{t,T}^{1,N} + \sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)C_{t,T}^{2,N}\right)$ where

$$C_{t,T}^{1,N} \stackrel{\text{def}}{=} V_{t,T}^N\cdot N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell} \quad\text{and}\quad C_{t,T}^{2,N} \stackrel{\text{def}}{=} V_{t,T}^N\left|\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]-\widetilde{\Omega}_t^N\right|\,.$$

The random variables $\left\{\omega_t^{N,\ell}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right\}_{\ell=1}^N$ being bounded and conditionally independent given $\mathcal{F}_{t-1}^N$, following the same steps as in the proof of Proposition 1, there exists a constant C (depending only on q, σ_-, σ_+, c_- and $\sup_{t\geq 0}|\omega_t|_\infty$) such that $\left\|V_{t,T}^N\right\|_{2q} \leq CN^{-1/2}$. Similarly

$$\left\|N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}\right\|_{2q} \leq C\frac{\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)}{N^{1/2}}\,,$$

and

$$\left\|\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]-\widetilde{\Omega}_t^N\right\|_{2q} \leq \frac{C}{N^{1/2}}\,.$$

The Cauchy-Schwarz inequality concludes the proof of (22). □

The proof of Theorem 1 is now concluded for the FFBS estimator $\phi^N_{0:T|T}[S_{T,r}]$ and we can proceed to the proof for the FFBSi estimator. We preface the proof of Theorem 1 for the FFBSi estimator $\widetilde{\phi}^N_{0:T|T}$ by the following Lemma. We first define the backward filtration $\left\{\mathcal{G}^N_{t,T}\right\}_{t=0}^{T+1}$ by

$$\begin{cases} \mathcal{G}^N_{T+1,T} \stackrel{\text{def}}{=} \mathcal{F}^N_T \ , \\ \mathcal{G}^N_{t,T} \stackrel{\text{def}}{=} \mathcal{F}^N_T \vee \sigma\left\{J^\ell_u, 1 \leq \ell \leq N, t \leq u \leq T\right\}, \quad \forall\, t \in \{0, \ldots, T\} \ . \end{cases}$$

Lemma 2. *Assume A1–3. Let $\ell \in \{1, \ldots, N\}$ and $T < +\infty$. For any bounded measurable function h on $\mathbb{X}^{r+1}$ we have,*

(i) for all u, t such that $r \leq t \leq u \leq T$,

$$\left|\mathbb{E}\left[h\left(\xi^{N,J^\ell_{t-r:t}}_{t-r:t}\right)\middle|\mathcal{G}^N_{u,T}\right] - \mathbb{E}\left[h\left(\xi^{N,J^\ell_{t-r:t}}_{t-r:t}\right)\middle|\mathcal{G}^N_{u+1,T}\right]\right| \leq \rho^{u-t}\mathrm{osc}(h) \ ,$$

where ρ is defined in A1(i).

(ii) for all u, t such that $t - r \leq u \leq t - 1 \leq T$,

$$\left|\mathbb{E}\left[h\left(\xi^{N,J^\ell_{t-r:t}}_{t-r:t}\right)\middle|\mathcal{G}^N_{u,T}\right] - \mathbb{E}\left[h\left(\xi^{N,J^\ell_{t-r:t}}_{t-r:t}\right)\middle|\mathcal{G}^N_{u+1,T}\right]\right| \leq \mathrm{osc}(h) \ .$$

Proof. According to Section 2.2, for all $\ell \in \{1, \ldots, N\}$, $\{J^{N,\ell}_u\}_{u=0}^T$ is an inhomogeneous Markov chain evolving backward in time with backward kernel $\{\Lambda^N_u\}_{u=0}^{T-1}$. For any $r \leq t \leq u \leq T$, we have

$$\begin{aligned} &\mathbb{E}\left[h\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right)\middle|\mathcal{G}^N_{u,T}\right] - \mathbb{E}\left[h\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right)\middle|\mathcal{G}^N_{u+1,T}\right] \\ &\qquad = \sum_{j_{t:u}}\left[\delta_{J^{N,\ell}_u}(j_u) - \left(\Lambda_u(J^{N,\ell}_{u+1}, j_u)\mathbf{1}_{u<T} + \frac{\omega^{N,j_u}_T}{\Omega_u}\mathbf{1}_{u=T}\right)\right] \\ &\qquad\qquad \times \prod_{\ell=u}^{t+1}\Lambda^N_{\ell-1}(j_\ell, j_{\ell-1}) \sum_{j_{t-r:t-1}} \prod_{\ell=t}^{t-r+1}\Lambda^N_{\ell-1}(j_\ell, j_{\ell-1}) h\left(\xi^{N,j_{t-r:t}}_{t-r:t}\right) \ . \end{aligned}$$

The RHS of this equation is the difference between two expectations started with two different initial distributions. Under A1(i), the backward kernel satisfies the uniform Doeblin condition,

$$\forall (i,j) \in \{1, \ldots, N\}^2 \quad \Lambda^N_s(i,j) \geq \frac{\sigma_-}{\sigma_+}\frac{\omega^i_s}{\Omega^N_s} \ ,$$

and the proof is completed by the exponential forgetting of the backward kernel (see [2, 7]). The proof of (ii) follows exactly the same lines. □

To compute an upper-bound for the L_q-mean error of the FFBSi algorithm, we may define the difference between the FFBS and the FFBSi estimators:

$$\delta_T^N\left[S_{T,r}\right] = \widetilde{\phi}_{0:T|T}^N\left[S_{T,r}\right] - \phi_{0:T|T}^N\left[S_{T,r}\right] \ . \tag{25}$$

Proof of Theorem 1 for the FFBSi estimator. The difference between the FFBS and the FFBSi estimators, δ_T^N, defined in (25), can be written

$$\begin{aligned}
\delta_T^N\left[S_{T,r}\right] &= \frac{1}{N}\sum_{\ell=1}^{N}\sum_{t=r}^{T} h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right) - \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,1}}\right)\middle|\mathcal{F}_T^N\right] \\
&= \frac{1}{N}\sum_{\ell=1}^{N}\sum_{t=r}^{T}\sum_{u=t-r}^{T} \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u,T}^N\right] - \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u+1,T}^N\right] \\
&= \frac{1}{N}\sum_{\ell=1}^{N}\sum_{u=0}^{T} \zeta_u^{N,\ell} \ ,
\end{aligned}$$

where

$$\zeta_u^{N,\ell} \stackrel{\text{def}}{=} \sum_{t=r}^{(u+r)\wedge T} \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u,T}^N\right] - \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u+1,T}^N\right] \ .$$

For all $\ell \in \{1,\ldots,N\}$ and all $u \in \{0,\ldots,T\}$, the random variable $\zeta_u^{N,\ell}$ is $\mathcal{G}_{u,T}^N$-measurable and $\mathbb{E}\left[\zeta_u^{N,\ell}|\mathcal{G}_{u+1,T}^N\right] = 0$ so that $\zeta_u^{N,\ell}$ can be seen as the increment of a backward martingale. Hence, since $q \geq 2$, using the Burkholder inequality (see [13, Theorem 2.10, page 23]), there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that:

$$\left\|\delta_T^N\left[S_{T,r}\right]\right\|_q \leq C\left\{\sum_{u=0}^{T}\mathbb{E}\left[\left|N^{-1}\sum_{\ell=1}^{N}\zeta_u^{N,\ell}\right|^q\right]^{2/q}\right\}^{1/2} \ . \tag{26}$$

Then, since the random variables $\{\zeta_u^{N,\ell}\}_{\ell=1}^N$ are conditionally independent and centered conditionally to $\mathcal{G}_{u+1,T}^N$, using the Burkholder inequality once again implies:

$$\mathbb{E}\left[\left|\sum_{\ell=1}^{N}\zeta_u^{N,\ell}\right|^q\middle|\mathcal{G}_{u+1,T}^N\right] \leq CN^{q/2-1}\sum_{\ell=1}^{N}\mathbb{E}\left[\left|\zeta_u^{N,\ell}\right|^q\middle|\mathcal{G}_{u+1,T}^N\right] \ . \tag{27}$$

Furthermore, according to Lemma 2(i),

$$\begin{aligned}\left|\zeta_u^{N,\ell}\right| &\leq \sum_{t=r}^{u}\left|\mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u,T}^N\right]-\mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u+1,T}^N\right]\right| \\ &+\sum_{t=u+1}^{(u+r)\wedge T}\left|\mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u,T}^N\right]-\mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u+1,T}^N\right]\right| \\ &\leq \sum_{t=r}^{u}\rho^{u-t}\mathrm{osc}(h_t)+\sum_{t=u+1}^{(u+r)\wedge T}\mathrm{osc}(h_t)\ . \quad (28)\end{aligned}$$

Putting (26), (27) and (28) together leads to

$$\left\|\delta_T^N\left[S_{T,r}\right]\right\|_q \leq \frac{C}{\sqrt{N}}\left\{\sum_{u=0}^{T}\left(\sum_{t=r}^{(u+r)\wedge T}\rho^{(u-t)\vee 0}\mathrm{osc}(h_t)\right)^2\right\}^{1/2}\ .$$

Using the Holder inequality as in the proof of Proposition 1 yields

$$\left\|\delta_T^N\left[S_{T,r}\right]\right\|_q \leq \frac{C}{\sqrt{N}}\sqrt{1+r}\left(\sqrt{1+r}\wedge\sqrt{T-r+1}\right)\left(\sum_{s=r}^{T}\mathrm{osc}(h_s)^2\right)^{1/2}\ ,$$

and the proof of Theorem 1 for the FFBSi estimator is derived from the triangle inequality:

$$\left\|\phi_{0:T|T}\left(S_{T,r}\right)-\widetilde{\phi}_{0:T|T}^N\left(S_{T,r}\right)\right\|_q \leq \left\|\Delta_T^N[S_{T,r}]\right\|_q+\left\|\delta_T^N\left[S_{T,r}\right]\right\|_q\ ,$$

where $\Delta_T^N[S_{T,r}]$ is defined by (9) and $\delta_T^N\left[S_{T,r}\right]$ is defined by (25). □

6 Proof of Theorem 2

We preface the proof of the Theorem by showing that the martingale term of the error $\Delta_T^N[S_{T,r}]$ (which is defined by (9)) satisfies an exponential deviation inequality in the following Proposition.

Proposition 3. *Assume A1–3. There exists a constant C (depending only on σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T<\infty$, any $N\geq 1$, any $\varepsilon>0$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\mathbb{P}\left\{\left|\sum_{t=0}^{T}D_{t,T}^N(S_{T,r})\right|>\varepsilon\right\}\leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^{T}\mathrm{osc}(h_s)^2}\right)\ , \quad (29)$$

where $D_{t,T}^N$ is defined in (14) *and $\Theta_{r,T}$ is defined by* (17).

Proof. According to the definition of $D^N_{t,T}(S_{T,r})$ given in (14), we can write

$$\sum_{t=0}^{T} D^N_{t,T}(S_{T,r}) = \sum_{k=1}^{N(T+1)} \upsilon^N_k \ ,$$

where for all $t \in \{0,\dots,T\}$ and $\ell \in \{1,\dots,N\}$, $\upsilon^N_{Nt+\ell}$ is defined by

$$\upsilon^N_{Nt+\ell} = \frac{\phi^N_{t-1}[\vartheta_t]}{\phi^N_{t-1}\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]} N^{-1}\omega^{N,\ell}_t \frac{G^N_{t,T} S_{T,r}(\xi^{N,\ell}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \ ,$$

and is bounded by (see (18))

$$\left|\upsilon^N_{Nt+\ell}\right| \leq CN^{-1}\sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s) \ .$$

Furthermore, we define the filtration $\left\{\mathcal{H}^N_k\right\}_{k=1}^{N(T+1)}$, for all $t \in \{0,\dots,T\}$ and $\ell \in \{1,\dots,N\}$, by:

$$\mathcal{H}^N_{Nt+\ell} \stackrel{\text{def}}{=} \mathcal{F}^N_{t-1} \vee \sigma\left\{\left(\omega^{N,i}_t, \xi^{N,i}_t\right), 1 \leq i \leq \ell\right\} \ ,$$

with the convention $\mathcal{F}^N_{-1} = \sigma(Y_{0:T})$. Then, according to Lemma 1, $\{\upsilon_k\}_{k=1}^{N(T+1)}$ is martingale increment for the filtration $\{\mathcal{H}^N_k\}_{k=1}^{N(T+1)}$ and the Azuma-Hoeffding inequality completes the proof. □

Proposition 4. *Assume A1–3. There exists a constant C (depending only on σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T < \infty$, any $N \geq 1$, any $\varepsilon > 0$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\mathbb{P}\left\{\left|\sum_{t=0}^{T} C^N_{t,T}(S_{T,r})\right| > \varepsilon\right\} \leq 8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right) \ . \tag{30}$$

where $C^N_{t,T}(F)$ is defined in (15).

Proof. In order to apply Lemma 4 in the appendix, we first need to find an exponential deviation inequality for $C^N_{t,T}(S_{T,r})$ which is done by using the decomposition $C^N_{t,T}(S_{T,r}) = U^N_{t,T}V^N_{t,T}W^N_{t,T}$ given in (23). First, the ratio $U^N_{t,T}$ is dealt with through Lemma 3 in the appendix by defining

$$\begin{cases} a_N \stackrel{\text{def}}{=} N^{-1}\sum_{\ell=1}^N \omega^{N,\ell}_t G^N_{t,T} S_{T,r}(\xi^{N,\ell}_t)/|\mathcal{L}_{t,T}\mathbf{1}|_\infty \ , \\ b_N \stackrel{\text{def}}{=} N^{-1}\sum_{\ell=1}^N \omega^{N,\ell}_t \ , \\ b \ \stackrel{\text{def}}{=} \mathbb{E}[\omega^1_t|\mathcal{F}^N_{t-1}] = \phi^N_{t-1}\left[Mg_t\right]/\phi^N_{t-1}[\vartheta_t] \ , \\ \beta \ \stackrel{\text{def}}{=} c_-/|\vartheta_t|_\infty \ . \end{cases}$$

Assumption A1(ii) and A3 shows that $b \geq \beta$ and (18) shows that $|a_N/b_N| \leq C(1+r)\max_{r\leq t\leq T}\{\mathrm{osc}(h_t)\}$. Therefore, Condition (I) of Lemma 3 is satisfied. The bounds $0 < \omega_t^l \leq |\omega_t|_\infty$ and the Hoeffding inequality lead to

$$\mathbb{P}[|b_N - b| \geq \varepsilon] = \mathbb{E}\left[\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^{N}\left(\omega_t^{N,\ell} - \mathbb{E}[\omega_t^{N,1}|\mathcal{F}_{t-1}^N]\right)\right| \geq \varepsilon \middle| \mathcal{F}_{t-1}^N\right]\right] \leq 2\exp\left(-\frac{2N\varepsilon^2}{|\omega_t|_\infty^2}\right) ,$$

establishing Condition (ii) in Lemma 3. Finally, Lemma 1(i) and the Hoeffding inequality imply that

$$\mathbb{P}\left[|a_N| \geq \varepsilon\right] = \mathbb{E}\left[\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^{N}\omega_t^{N,\ell}G_{t,T}^N S_{T,r}(\xi_t^{N,\ell})/|\mathcal{L}_{t,T}\mathbf{1}|_\infty\right| \geq \varepsilon \middle| \mathcal{F}_{t-1}^N\right]\right] \leq 2\exp\left(-\frac{N\varepsilon^2}{2|\omega_t|_\infty^2\left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right)^2}\right) .$$

Lemma 3 therefore yields

$$\mathbb{P}\left\{\left|U_{t,T}^N\right| \geq \varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right)^2}\right) .$$

Then $V_{t,T}^N$ is dealt with by using again the Hoeffding inequality and the bounds $0 < b_{t,T}^{N,\ell} \leq |\omega_t|_\infty$, where $b_{t,T}^{N,\ell} \stackrel{\mathrm{def}}{=} \omega_t^{N,\ell}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}$:

$$\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^{N}b_{t,T}^{N,\ell} - \mathbb{E}\left[b_{t,T}^{N,1}\middle|\mathcal{F}_{t-1}\right]\right| \geq \varepsilon\right]$$
$$= \mathbb{E}\left[\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^{N}\left(b_{t,T}^{N,\ell} - \mathbb{E}\left[b_{t,T}^{N,\ell}\middle|\mathcal{F}_{t-1}^N\right]\right)\right| \geq \varepsilon \middle| \mathcal{F}_{t-1}^N\right]\right] \leq 2\exp\left(-CN\varepsilon^2\right) .$$

Finally, $W_{t,T}^N$ has been shown in (24) to be bounded by a constant depending only on σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$: $\left|W_{t,T}^N\right| \leq C$ so that

$$\mathbb{P}\left\{\left|C_{t,T}^N(S_{T,r})\right| > \varepsilon\right\} \leq \mathbb{P}\left\{\left|U_{t,T}^N V_{t,T}^N\right| > \varepsilon/C\right\} \leq \mathbb{P}\left\{\left|U_{t,T}^N\right| > \varepsilon_u\right\} + \mathbb{P}\left\{\left|V_{t,T}^N\right| > \varepsilon_v\right\} ,$$

where

$$\varepsilon_u \stackrel{\mathrm{def}}{=} \sqrt{\varepsilon\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)/C} \quad \text{and} \quad \varepsilon_u \stackrel{\mathrm{def}}{=} \sqrt{\frac{\varepsilon}{C\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)}} .$$

Therefore,

$$\mathbb{P}\left\{\left|C_{t,T}^N(S_{T,r})\right| > \varepsilon\right\} \leq 4\exp\left(-\frac{CN\varepsilon}{\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)}\right) .$$

The proof of (30) is finally completed by applying Lemma 4 with

$$X_t = C_{t,T}^N(S_{T,r}) , \quad A = 4 , \quad B_t = \frac{CN}{\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)} , \quad \gamma = 1/2 .$$

□

Proof of Theorem 2 for the FFBS estimator. The result is obtained by writing

$$\mathbb{P}\left\{\left|\Delta_T^N[S_{T,r}]\right| > \varepsilon\right\} \leq \mathbb{P}\left\{\left|\sum_{t=0}^T C_{t,T}^N(S_{T,r})\right| > \varepsilon/2\right\} + \mathbb{P}\left\{\left|\sum_{t=0}^T D_{t,T}^N(S_{T,r})\right| > \varepsilon/2\right\} ,$$

and using (29) and (30). □

Proof of Theorem 2 for the FFBSi estimator. We recall the decomposition used in the proof of Theorem 1 for the FFBSi estimator:

$$\delta_T^N[S_{T,r}] = \frac{1}{N}\sum_{\ell=1}^N \sum_{u=0}^T \zeta_u^{N,\ell} ,$$

where $\delta_T^N[S_{T,r}]$ is defined by (25). Since $\left\{\zeta_u^{N,\ell}\right\}_{\ell=1}^N$ are $\mathcal{G}_{u,T}^N$ measurable and centered conditionally to $\mathcal{G}_{u+1,T}^N$ using the same steps as in the proof of Proposition 3, we get

$$\mathbb{P}\left\{\left|\delta_T^N[S_{T,r}]\right| > \varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right) ,$$

where $\Theta_{r,T}$ is defined by (17). The proof is finally completed by writing

$$\phi_{0:T|T}[S_{T,r}] - \widetilde{\phi}_{0:T|T}^N[S_{T,r}] = \Delta_T^N[S_{T,r}] + \delta_T^N[S_{T,r}] ,$$

and by using Theorem 2 for the FFBS estimator. □

A Technical results

Lemma 3. *Assume that a_N, b_N, and b are random variables defined on the same probability space such that there exist positive constants β, B, C, and M satisfying*

(i) $|a_N/b_N| \leq M$, $\mathbb{P}$*-a.s. and* $b \geq \beta$, $\mathbb{P}$*-a.s.,*

(ii) For all $\epsilon > 0$ *and all* $N \geq 1$, $\mathbb{P}[|b_N - b| > \epsilon] \leq B\mathrm{e}^{-CN\epsilon^2}$,

(iii) For all $\epsilon > 0$ and all $N \geq 1$, $\mathbb{P}\left[|a_N| > \epsilon\right] \leq B\mathrm{e}^{-CN(\epsilon/M)^2}$.

Then,

$$\mathbb{P}\left\{\left|\frac{a_N}{b_N}\right| > \epsilon\right\} \leq B\exp\left(-CN\left(\frac{\epsilon\beta}{2M}\right)^2\right).$$

Proof. See [8, Lemma 4]. □

Lemma 4. *For $T \geq 0$, let $\{X_t\}_{t=0}^T$ be $(T+1)$ random variables. Assume that there exists a constants $A \geq 1$ and for all $0 \leq t \leq T$, there exists a constant $B_t > 0$ such that and all $\varepsilon > 0$*

$$\mathbb{P}\{|X_t| > \varepsilon\} \leq Ae^{-B_t\varepsilon}\ .$$

Then, for all $0 < \gamma < 1$ and all $\varepsilon > 0$, we have

$$\mathbb{P}\left\{\left|\sum_{t=0}^T X_t\right| > \varepsilon\right\} \leq \frac{A}{1-\gamma}e^{-\gamma B\varepsilon/(T+1)}\ ,$$

where

$$B \stackrel{\text{def}}{=} \left(\frac{1}{T+1}\sum_{t=0}^T B_t^{-1}\right)^{-1}\ .$$

Proof. By the Bienayme-Tchebychev inequality, we have

$$\begin{aligned}\mathbb{P}\left\{\left|\sum_{t=0}^T X_t\right| > \varepsilon\right\} &= \mathbb{P}\left\{\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right] > e^{\gamma B\varepsilon/(T+1)}\right\} \\ &\leq e^{-\gamma B\varepsilon/(T+1)}\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right]\right]\ . \quad (31)\end{aligned}$$

It remains to bound the expectation in the RHS of (31) by $A(1-\gamma)^{-1}$. First, by the Minkowski inequality,

$$\begin{aligned}\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right]\right] &= \sum_{q=0}^\infty \frac{\gamma^q B^q}{q!(T+1)^q}\mathbb{E}\left[\left|\sum_{t=0}^T X_t\right|^q\right] \\ &\leq 1 + \sum_{q=1}^\infty \frac{\gamma^q B^q}{q!(T+1)^q}\left(\sum_{t=0}^T \|X_t\|_q\right)^q\ .\end{aligned}$$

Moreover, for $q \geq 1$, $\mathbb{E}\left[|X_t|^q\right]$ can be bounded by

$$\mathbb{E}\left[|X_t|^q\right] = \int_0^\infty \mathbb{P}\{|X_t| > \varepsilon^{1/q}\}\mathrm{d}\varepsilon \leq A\int_0^\infty e^{-B_t\varepsilon^{1/q}}\mathrm{d}\varepsilon = \frac{Aq!}{B_t^q}\ ,$$

Finally,

$$\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right]\right] \leq A\sum_{q=0}^\infty \gamma^q = \frac{A}{(1-\gamma)}\ .$$

□

References

[1] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state-space models. *Annals Institute Statistical Mathematics*, 62(1):61–89, 2010.

[2] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

[3] J. Davidson. *Stochastic limit theory*. Oxford university press, 1997.

[4] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[5] P. Del Moral, A. Doucet, and S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975, 2010.

[6] P. Del Moral, A. Doucet, and S.S. Singh. Forward smoothing using sequential monte carlo. Technical report, arXiv:1012.5390v1, 2010.

[7] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré*, 37:155–194, 2001.

[8] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.*, 21(6):2109–2145, 2011.

[9] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10:197–208, 2000.

[10] A. Doucet, G. Poyiadjis, and S.S. Singh. Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, 98(1):65–80, 2011.

[11] J. Durbin and S. J. Koopman. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. B*, 62:3–29, 2000.

[12] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for nonlinear time series. *J. Am. Statist. Assoc.*, 50:438–449, 2004.

[13] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, London, 1980.

[14] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42:281–300, 1987.

[15] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.*, 7:175–193, 1998.

[16] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1:1–25, 1996.

[17] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94(446):590–599, 1999.

[18] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 1989.