State-Observation Sampling

Laurent E. Calvet and Veronika Czellar*

First version: February 2010

This version: October 2011

Abstract

In this paper, we introduce a new sequential Monte Carlo method, the state-observation sampling (SOS) filter. SOS extends the particle filtering methodology to general state-space models where the density of the observation conditional on the state is unavailable. In the mutation stage of SOS, a set of state-observation pairs are sampled from past state particles. In the importance sampling step, each state particle is weighted by the kernel distance between its corresponding simulated observation and the actual data point.

*Calvet: Department of Finance, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France; calvet@hec.fr. Czellar: Department of Economics and Decision Sciences, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France, czellarv@hec.fr. We received helpful comments from Per Mykland, Nicholas Polson, Elvezio Ronchetti, Andrew Siegel and seminar participants at CORE, the University of Chicago, the University of Evry, the Second HEC Finance and Statistics Conference, CEF 2010 in London, the 2010 Toulouse School of Economics Financial Econometrics Conference, ISBIS 2010 in Portoroz, and the 2010 Econometric Society World Congress in Shanghai. We gratefully acknowledge the computer support of EUROFIDAI and the financial support of the American Statistical Association, the Europlace Institute of Finance, and the National Institute of Statistical Sciences.

We establish that the convergence rate of SOS coincides with the convergence rate of a kernel density estimator on the observation space. SOS overcomes the curse of dimensionality with respect to the size of state space. We develop a plug-in rule for the selection of the bandwidth. The good finite-sample performance of SOS is demonstrated on an asset pricing model with investor learning.

Keywords: State-space model; Particle filter; Kernels; Bandwidth; Likelihood estimation.

1 Introduction

Since their introduction by Gordon, Salmond, and Smith (1993), particle filters have considerably expanded the range of applications of hidden Markov models and now pervade fields as diverse as biomedical research (Acton and Ray, 2006; Liu *et al.*. 2011), biometrics (Tistarelli, 2009), ecology (Newman *et al*, 2009), finance (Kim, Shephard and Chib, 1998; Johannes and Polson, 2009) and macroeconomics (Fernandez-Villaverde et al., 2011; Hansen, Polson and Sargent, 2011). Particle filters help track phenomena as diverse as virus dynamics in epidemiology, tumor cell kinetics in cancer research, and market behavior in economics and finance. These methods provide estimates of the distribution of a hidden Markov state x_t conditional on a time series of observations $Y_t = (y_1, ..., y_t)$, $y_t \in \mathbb{R}^{n_Y}$, by way of a set of "particles" $(x_t^{(1)}, ..., x_t^{(N)})$. In the original sampling and importance resampling algorithm of Gordon *et al.* (1993), the construction of the date-t filter from the date-$(t-1)$

particles proceeds in two steps. In the mutation phase, a new set of particles is obtained by drawing a hidden state $\tilde{x}_t^{(n)}$ from each date-$(t-1)$ particle $x_{t-1}^{(n)}$ under the transition kernel of the Markov state. Given a new observation y_t, the particles are then resampled using weights that are proportional to the observation density $f_Y(y_t|\tilde{x}_t^{(n)}, Y_{t-1})$. Important refinements of the algorithm include sampling from an auxiliary model in the mutation phase (Pitt and Shephard, 1999), or implementing variance-reduction techniques such as stratified (Kitagawa 1996) and residual (Liu and Chen 1998) resampling.[1]

A common feature of existing filters is the requirement that the observation density $f_Y(y_t|x_t, Y_{t-1})$ be available analytically up to a normalizing constant. This condition need not hold in complex environments such as state-space models of animal populations, HIV dynamics in the presence of mixed effects, or dynamic equilibria with hidden agents beliefs or latent macroeconomic fundamentals. To overcome the unavailability of the observation density nonparametric (Rossi and Vila, 2006, 2009) and Approximate Bayesian Computation ("ABC"; Jasra *et al.*, 2011) methods, both based on simulations of pseudo-observations, have been considered. However, the former approach is only applicable when the state x_t is a continuous random variable evolving in a Euclidean space $\mathbb{R}^{n_X}$. It is numerically challenging because N conditional densities, and therefore $2N^2$ kernels, must be evaluated every period. The method is also prone to the curse of dimensionality since the rate of convergence decreases both with the dimension of the state space n_X and the dimension

[1]Further advances of particle filtering in the statistics literature include Andrieu, Doucet, and Holenstein (2010), Del Moral (2004), Fearnhead and Clifford (2003), Gilks and Berzuini (2001), Godsill, Doucet, and West (2004), Kitagawa (1998), Liu and Chen (1995) and Storvik (2002).

of the observation space n_Y. In the ABC approach, the pseudo-observations need to be sorted every period, which requires $N \log_2 N$ steps, and convergence is an open question.

The present paper develops a new method, the State-Observation Sampling (SOS) filter, which consists of simulating a state and a pseudo-observation $(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})$ from each date-$(t-1)$ particle. In the resampling stage, we assign to each particle $\tilde{x}_t^{(n)}$ an importance weight determined by the proximity between the pseudo-observation $\tilde{y}_t^{(n)}$ and the actual observation y_t. We quantify proximity by a kernel of the type considered in nonparametric statistics:

$$p_t^{(n)} \propto \frac{1}{h_t^{n_Y}} K\left(\frac{y_t - \tilde{y}_t^{(n)}}{h_t}\right),$$

where h_t is a bandwidth and K is a probability density function supported on the real line. SOS requires the calculation of only N kernels each period and makes no assumptions on the characteristics of the state space, which may or may not be Euclidean. We demonstrate that as the number of particles N goes to infinity, the filter converges to the target distribution under a wide range of conditions on the bandwidth h_t. The root mean squared error of moments computed using the filter decays at the rate $N^{-2/(n_Y+4)}$, that is at the same rate as the kernel density estimator of a random vector on $\mathbb{R}^{n_Y}$ (e.g. Fan and Yao, 2003; Hart, 1997). The asymptotic rate of convergence is thus invariant to the size of the state space, indicating that SOS overcomes a form of the curse of dimensionality. We also prove that the SOS filter provides consistent estimates of the likelihood function and propose a plug-in rule for the choice of bandwidth.

We demonstrate the good performance of the SOS filter on the dynamic asset pricing model with agent learning of Calvet and Fisher ("CF" 2007), where the Markov chain driving fundamentals (M_t) and agent belief (Π_t) about the Markov chain both drive the observation dynamics. The advantages of the CF 2007 model is that the state variable is mixed (discrete M_t and continuous Π_t), which allows us to investigate the accuracy of SOS in the most general case. Moreover, in the special case when the agent is fully informed about the state of fundamentals, the state reduces to M_t and the likelihood function is available in closed form. This feature allows us to investigate the convergence of the SOS estimated likelihood to the true likelihood under different state-space dimensions and to illustrate the defeat of the curse of dimensionality with respect to the size of the state space (see Figure 4).

The paper is organized as follows. Section 2 defines the SOS filter for general state-space models. Section 3 applies these methods to a dynamic asset pricing model with agent learning; we verify the accuracy of our inference methodology by Monte Carlo simulations. Section 4 concludes.

2 The State-Observation Sampling (SOS) Filter

2.1 Definition in a Euclidean State Space

We consider a Markov process x_t defined on a measurable space $(\mathcal{X}, \mathcal{F}_X)$. Time is discrete and indexed by $t = 0, 1, ..., \infty$. For expositional simplicity, we assume in this subsection that $\mathcal{X} = \mathbb{R}^{n_X}$.

Let $y_t \in \mathbb{R}^{n_Y}$ denote the observation at date t and $Y_{t-1} = (y_1, ..., y_{t-1})$, the vector

of observations up to date $t-1$. The building block of our model is the conditional density of (x_t, y_t) given (x_{t-1}, Y_{t-1}),

$$f_{X,Y}(x_t, y_t | x_{t-1}, Y_{t-1}). \tag{2.1}$$

Let f_{X_0} denote a prior over the state space. The inference problem consists of estimating the density of the latent state x_t conditional on the set of current and past observations: $f_X(x_t|Y_t)$ at all $t \geq 1$.

A large literature proposes estimation by way of a particle filter. The sampling importance resampling method of Gordon *et al.* (1993) is based on Bayes'rule:

$$f_X(x_t|Y_t) = \frac{f_Y(y_t|x_t, Y_{t-1})\, f_X(x_t|Y_{t-1})}{f_Y(y_t|Y_{t-1})}\,.$$

The construction begins by drawing N independent states $x_0^{(1)}, ..., x_0^{(N)}$ from f_{X_0}. Given the date$-(t-1)$ filter $(x_{t-1}^{(1)}, \ldots, x_{t-1}^{(N)})$, the construction of the date$-t$ filter proceeds in two steps. First, we sample $\tilde{x}_t^{(n)}$ from $x_{t-1}^{(n)}$ using the transition kernel of the Markov process. Second, in the resampling step, we sample N particles $(x_t^{(1)}, \ldots, x_t^{(N)})$ from $(\tilde{x}_t^{(1)}, \ldots, \tilde{x}_t^{(N)})$ with normalized importance weights

$$p_t^{(n)} = \frac{f_Y(y_t|\tilde{x}_t^{(n)}, Y_{t-1})}{\sum_{n'=1}^{N} f_Y(y_t|\tilde{x}_t^{(n')}, Y_{t-1})}. \tag{2.2}$$

Under a wide range of conditions, the sample mean $N^{-1}\sum_{n=1}^{N}\Phi(x_t^{(n)})$ converges to $\mathbb{E}[\Phi(x_t)|Y_t]$ for any bounded measurable function Φ (e.g. Crisan and Doucet, 2002).

The sampling and importance resampling algorithm, and its various refinements,

assume that the observation density $f_Y(y_t|x_t, Y_{t-1})$ is readily available up to a normalizing constant, which is a restrictive assumption in many applications. We propose a solution to this difficulty when it is possible to simulate from (2.1). Our filter makes no assumption on the tractability of $f_{X,Y}(\cdot|x_{t-1}, Y_{t-1})$, and in fact does not even require that the transition kernel of the Markov state x_t be available explicitly. The principle of our new filter is to simulate from each $x_{t-1}^{(n)}$ a state-observation pair $(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})$, and then select particles $\tilde{x}_t^{(n)}$ associated with pseudo-observations $\tilde{y}_t^{(n)}$ that are close to the actual data point y_t. The definition of the importance weights is based on Bayes' rule applied to the joint distribution of $\tilde{y}_t^{(n)}, \tilde{x}_t^{(n)}, x_{t-1}^{(n)}$ conditional on Y_t:

$$\tilde{y}_t^{(n)}, \tilde{x}_t^{(n)}, x_{t-1}^{(n)}|Y_t \sim \frac{\delta(y_t - \tilde{y}_t^{(n)})\, f_{X,Y}(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)}|x_{t-1}^{(n)}, Y_{t-1})\, f_X(x_{t-1}^{(n)}|Y_{t-1})}{f_Y(y_t|Y_{t-1})}, \tag{2.3}$$

where δ denotes the Dirac distribution on $\mathbb{R}^{n_Y}$. Since the Dirac distribution produces degenerate weights, we consider a strictly positive kernel K that satisfies the usual properties.

Assumption 1 (Kernel). *The function $K : \mathbb{R}^{n_Y} \to \mathbb{R}$ satisfies:*

(i) $K(u) > 0$ *for all* $u \in \mathbb{R}^{n_Y}$*;*

(ii) $\int K(u)du = 1$*;*

(iii) $\int uK(u)du = 0$*;*

(iv) $A(K) = \int \|u\|^2 K(u)du < \infty$*;*

(v) $B(K) = \int [K(u)]^2 du < \infty$.

For any $y \in \mathbb{R}^{n_Y}$, let

$$K_{h_t}(y) = \frac{1}{h_t^{n_Y}} K\left(\frac{y}{h_t}\right)$$

denote the corresponding kernel with bandwidth h_t at date t. The kernel K_{h_t} converges to the Dirac distribution as h_t goes to zero, which we use to approximate (2.3). This suggests the following algorithm.

SOS filter

Step 1 (State-observation sampling): For every $n = 1, \ldots, N$, we simulate a state-observation pair $(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})$ from $f_{X,Y}(\cdot | x_{t-1}^{(n)}, Y_{t-1})$.

Step 2 (Importance weights): We observe the new data point y_t and compute

$$p_t^{(n)} = \frac{K_{h_t}\left(y_t - \tilde{y}_t^{(n)}\right)}{\sum_{n'=1}^{N} K_{h_t}\left(y_t - \tilde{y}_t^{(n')}\right)}, \quad n = 1, \ldots, N.$$

Step 3 (Multinomial resampling): For every $n = 1, \ldots, N$, we draw $x_t^{(n)}$ from $\tilde{x}_t^{(1)}, \ldots, \tilde{x}_t^{(N)}$ with importance weights $p_t^{(1)}, \ldots, p_t^{(N)}$.

The state-observation pairs $\{(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})\}_{n=1,\ldots,N}$ constructed in step 1 provide a discrete approximation to the conditional distribution of (x_t, y_t) given the data Y_{t-1}. In step 2, we construct a measure of the proximity between the pseudo and the actual data points, and in Step 3 we select particles for which this measure is large. The variance of multinomial resampling in step 3 can be reduced and computational speed can be improved by alternatives such as residual (Liu and Chen, 1998) or stratified (Kitagawa, 1996) resampling. In section 3, we obtain good results with a

combined residual-stratified approach.[2] The convergence proof below applies equally well to these alternatives.

Step 2 of SOS is based on the calculation of N kernels, analogous to the evaluation of N observation densities in the traditional sampling and importance resampling filter. The overall numerical complexity SOS is therefore the same as the $O(N)$ complexity of the traditional filter (in cases when the observation density is available analytically). Moreover, SOS improves on the $O(N^2)$ complexity of earlier nonparametric methods (e.g. Rossi and Vila, 2009) and the $O(N \log_2(N))$ complexity of the ABC algorithm proposed by Jasra *et al.* (2011).

2.2 Extension to Non-Euclidean State Spaces and Convergence

The SOS filter easily extends to the case of a general measurable state space $\mathcal{X}$. The building blocks of the model are the conditional probability measure of (x_t, y_t) given (x_{t-1}, Y_{t-1}):

$$g(\cdot|x_{t-1}, Y_{t-1}),$$

and a prior measure λ_0 over the state space. The SOS filter targets the probability measure of the latent state x_t conditional on the set of current and past obser-

[2]We select $\sum_{n=1}^{N} \lfloor N p_t^{(n)} \rfloor$ particles deterministically by setting $\lfloor N p_t^{(n)} \rfloor$ particles equal to $\tilde{x}_t^{(n)}$ for every $n \in \{1, \ldots, N\}$, where $\lfloor \cdot \rfloor$ denotes the floor of a real number. The remaining $N_{r,t} = N - \sum_{n=1}^{N} \lfloor N p_t^{(n)} \rfloor$ particles are selected by the stratified sampling that produces $\tilde{x}_t^{(n)}$ with probability $q_t^{(n)} = (N p_t^{(n)} - \lfloor N p_t^{(n)} \rfloor)/N_{r,t}$, $n = 1, \ldots, N$. That is, for every $k \in \{1, \ldots, N_{r,t}\}$, we draw $\tilde{U}_k$ from the uniform distribution on $(\frac{k-1}{N_{r,t}}, \frac{k}{N_{r,t}}]$, and select the particle $\tilde{x}_t^{(n)}$ such that $\tilde{U}_k \in (\sum_{j=1}^{n-1} q_t^{(j)}, \sum_{j=1}^{n} q_t^{(j)}]$.

vations, $\lambda(\cdot|Y_t)$. It is defined by sampling $(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})$ from the conditional measure $g(\cdot|x_{t-1}^{(n)}, Y_{t-1})$ in step 1, and then implementing steps 2 and 3 as in Section 2.

We now specify conditions under which for a fixed history $Y_T = (y_1, \ldots, y_T)$, $T \leq \infty$, the SOS filter converges in mean squared error to the target $\lambda(\cdot|Y_t)$ as the number of particles N goes to infinity.

Assumption 2 (Conditional Distributions). *The observation process satisfies the following hypotheses:*

(i) the conditional density $f_Y(\tilde{y}_t|x_{t-1}, Y_{t-1})$ exists and

$$\kappa_t = \sup\{f_Y(\tilde{y}_t|x_{t-1}, Y_{t-1}); (x_{t-1}, \tilde{y}_t) \in \mathcal{X} \times \mathbb{R}^{n_Y}\} < \infty\,;$$

(ii) the observation density $f_Y(\tilde{y}_t|x_t, Y_{t-1})$ is well-defined and there exists $\kappa_t' \in \mathbb{R}_+$ such that:

$$|f_Y(\tilde{y}_t|x_t, Y_{t-1}) - f_Y(y_t|x_t, Y_{t-1}) - \frac{\partial f_Y}{\partial y_t^T}(y_t|x_t, Y_{t-1})(\tilde{y}_t - y_t)| \leq \kappa_t' \|\tilde{y}_t - y_t\|^2$$

for all $(x_t, \tilde{y}_t) \in \mathcal{X} \times \mathbb{R}^{n_Y}$ and $t \leq T$.

Assumption 3 (Bandwidth). *The bandwidth is a function of N, $h_t = h_t(N)$, and satisfies*

(i) $\lim_{N\to\infty} h_t(N) = 0$,

(ii) $\lim_{N\to\infty} N[h_t(N)]^{n_Y} = +\infty$,

for all $t = 1, \ldots, T$.

We establish the following result in Appendix A.

Theorem 1 (Convergence of the SOS Filter). *Under assumptions 1 and 2 and for every* t *and* $N \geq 1$*, there exists* $U_t(N) \in \mathbb{R}_+$ *such that*

$$\mathbb{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t|Y_{t-1})\right]^2\right\} \leq \frac{[f_Y(y_t|Y_{t-1})]^2}{4} U_t(N), \tag{2.4}$$

where the expectation is over all the realizations of the random particle method. Furthermore, for any bounded measurable function, $\Phi : \mathcal{X} \to \mathbb{R}$,

$$MSE_t = \mathbb{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N} \Phi(x_t^{(n)}) - \mathbb{E}[\Phi(x_t)|Y_t]\right]^2\right\} \leq U_t(N)\|\Phi\|^2, \tag{2.5}$$

where $\|\Phi\| = \sup_{s\in\mathcal{X}} |\Phi(s)|$*. If assumption 3 also holds, then*

$$\lim_{N\to\infty} U_t(N) = 0\,,$$

and the filter converges in mean squared error. Furthermore, if the bandwidth sequence is of the form $h_t(N) = h_t(1)N^{-1/(n_Y+4)}$*, then* $U_t(N)$ *decays at rate* $N^{-4/(n_Y+4)}$ *and the root mean squared error* $MSE_t^{1/2}$ *at rate* $N^{-2/(n_Y+4)}$ *for all* t.

The convergence rate of SOS is the same as the convergence rate of a kernel density estimator on the observation space $\mathbb{R}^{n_Y}$. SOS therefore defeats the curse

of dimensionality with respect to the size of the state space, thus improving on the nonparametric approach proposed by Rossi and Vila (2006, 2009).

By (2.4), the kernel estimator

$$\hat{f}_t = \frac{1}{N} \sum_{n=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n)}), \tag{2.6}$$

converges to the conditional density of y_t given past observations. Consequently, we can estimate the log-likelihood function by $\sum_{t=1}^{T} \ln \hat{f}_t$. We now discuss the appropriate choice of bandwidth.

2.3 Choice of Bandwidth

We begin by observing that computing the vector $\{h_t(N)\}_{t=1}^{T}$ that minimizes the mean squared error (MSE) of the log likelihood estimator is a seemingly intractable high-dimensional problem. For this reason, we consider a much simpler problem. At each date t, we select the bandwidth that minimizes the integrated MSE in the estimation of $f_Y(y_t|Y_{t-1}) = \mathbb{E}[f_Y(y_t|x_{t-1}, Y_{t-1})|Y_{t-1}]$ under the simplifying assumption that the conditional measure $\lambda(\cdot|Y_{t-1})$ concides with the date$-(t-1)$ filter. In this setting, the state x_{t-1} takes values on the fixed finite support $\mathcal{X}_{t-1}^{(N)} = (x_{t-1}^{(1)}, \ldots, x_{t-1}^{(N)})$ with equal probabilities. Suppose that $y_t \in \mathbb{R}^{n_Y}$ is fixed, and let $f_t = N^{-1} \sum_{n=1}^{N} f_Y(y_t|x_{t-1}^{(n)}, Y_{t-1})$. The properties of the mean squared error

$$\mathbb{E}[(\hat{f}_t - f_t)^2|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}] = \mathrm{Var}(\hat{f}_t|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}) + \left[\mathbb{E}(\hat{f}_t|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}) - f_t\right]^2$$

are summarized in the following proposition established in Appendix A.

Proposition 1 *Assume that x_{t-1} takes values on the fixed finite support $\mathcal{X}_{t-1}^{(N)}$ with equal probabilities. Then,*

$$\mathbb{E}(\hat{f}_t|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}) - f_t = \frac{h_t^2}{2}\mathrm{tr}\left[\frac{\partial^2 f_t}{\partial y_t \partial y_t^T}\mathrm{Var}_K(u)\right] + \mathcal{O}(h_t^3)\,, \tag{2.7}$$

$$\mathrm{Var}(\hat{f}_t|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}) = \frac{B(K)}{Nh_t^{n_Y}} f_t + \mathcal{O}(h_t^{-n_Y+1})\,. \tag{2.8}$$

where $\frac{\partial^2 f_t}{\partial y_t \partial y_t^T} = \frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2 f_Y}{\partial y_t \partial y_t^T}(y_t|x_{t-1}^{(n)}, Y_{t-1})$, tr *is the trace operator and* $\mathrm{Var}_K(u) = \int uu^T K(u)du$.

The mean squared error $\mathbb{E}[(\hat{f}_t - f_t)^2|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}]$ can therefore be approximated by

$$\frac{B(K)}{Nh_t^{n_Y}} f_t + \frac{h_t^4}{4}\left\{\mathrm{tr}\Big[\frac{\partial^2 f_t}{\partial y_t \partial y_t^T}\mathrm{Var}_K(u)\Big]\right\}^2 .$$

We choose the bandwidth that minimizes the integrated mean squared error which leads to the rule:

$$h_t = \left[\frac{B(K)\, n_Y}{N\int\left\{\mathrm{tr}\big[\frac{\partial^2 f_t}{\partial y_t \partial y_t^T}\mathrm{Var}_K(u)\big]\right\}^2 dy_t}\right]^{1/(n_Y+4)} . \tag{2.9}$$

The formula reduces to the standard kernel density estimation (KDE) plug-in rule if f_t is Gaussian and the kernel satisfies $\mathrm{Var}_K(u) = A(K)n_Y^{-1}I_{n_Y}$, where I_{n_Y} is the n_Y-dimensional identity matrix. A separate observation is that if $n_Y = 1$, the rule (2.9) simplifies to:

$$h_t = \left[\frac{B(K)}{NA(K)^2\int(f_t'')^2 dy_t}\right]^{1/5} .$$

2.3.1 Example: the Quasi-Cauchy Kernel

When $n_Y = 1$ we can use the infinite support quasi-Cauchy kernel:

$$K(u) = \frac{1}{(1 + Cu^2)^2},$$

where $C = (\pi/2)^2$ is a normalizing constant. We know that $A(K) = 4/\pi^2$ and $B(K) = 5/8$. Moreover, if f_t is a Gaussian distribution $\mathcal{N}(\mu_t, \sigma_t^2)$, then $\int (f_t'')^2 dy_t = 3/(8\sqrt{\pi}\sigma_t^5)$. The quasi-Cauchy plug-in bandwidth is therefore

$$h_t = \sigma_t \left(\frac{5\pi^{9/2}}{48\,N} \right)^{1/5},$$

with σ_t estimated by the sample standard deviation of the simulated observations $\{\tilde{y}_t^{(n)}\}_{n=1}^N$. We use this rule in all the calculations reported in the following section.

3 Inference in an Asset Pricing Model with Investor Learning

We now apply our methodology to a dynamic incomplete-information equilibrium model of equity returns.

3.1 Specification

We adopt the Lucas tree economy with regime-switching fundamentals of CF (2007, 2008). We consider three levels of information corresponding to nature, an agent and

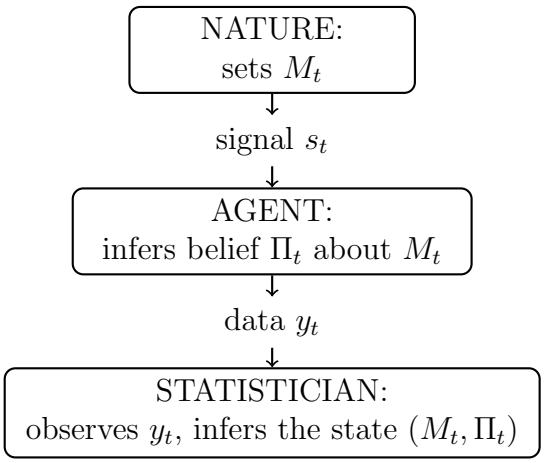

Figure 1: Information structure.

the statistician, as illustrated in Figure 1. At the beginning of every period t, nature selects a discrete first-order Markov vector $M_t \in \mathbb{R}^{\bar{k}}$. Each component of the vector can take either a high value $m_0 \in [1, 2)$ or a low value $2 - m_0$. The agent observes a signal s_t, whose distribution is contingent on the vector M_t, and uses Bayes' rule to infer the conditional probability distribution of M_t given $(s_1, \ldots, s_t)$, to which we will refer as the "belief" Π_t. The agent also computes the stock return y_t as a function of her beliefs and signals. The statistician observes y_t and aims to track the hidden state $x_t = (M_t, \Pi_t)$ of the learning economy.

3.1.1 State Model

Nature's M_t. The components of $M_t = (M_{k,t})_{1 \leq k \leq \bar{k}}$ are mutually independent across k. Conditional on $M_{k,t}$, the next-period multiplier $M_{k,t+1}$ is either: (a) unchanged: $M_{k,t+1} = M_{k,t}$ with probability $1 - \gamma_k$; or, alternatively, (b) drawn from a Bernoulli taking values m_0 or $2 - m_0$ with equal probability. The Markov vector M_t takes

$d = 2^{\overline{k}}$ possible values $m^1, \ldots, m^d$, and we denote by $a_{ij} = \mathbb{P}(M_t = m^j | M_{t-1} = m^i)$ its transition probabilities.

Signal. The components of the signal $s_t \in \mathbb{R}^{\overline{k}+2}$ consist of dividend growth $s_{1,t}$, consumption growth $s_{2,t}$, and signals $s_{3,t}, \ldots, s_{\overline{k}+2,t}$ about the state of nature. They are specified by:

$$s_{1,t} = g_D - \frac{\sigma_D^2(M_t)}{2} + \sigma_D(M_t)\varepsilon_{1,t}, \tag{3.1}$$

$$s_{2,t} = g_C + \sigma_C \varepsilon_{2,t}, \tag{3.2}$$

$$s_{i+2,t} = M_{i,t} + \sigma_\delta \varepsilon_{i,t}, \quad i = 1, \ldots, \overline{k}\,. \tag{3.3}$$

where g_D, g_C, σ_C, and σ_δ are fixed scalars. The stochastic volatility of dividends is $\sigma_D(M_t) = \overline{\sigma}_D \left(\prod_{k=1}^{\overline{k}} M_{k,t}\right)^{1/2}$, where $\overline{\sigma}_D \in \mathbb{R}_+$. The innovations $\varepsilon_{1,t}, \ldots, \varepsilon_{\overline{k},t}$ are jointly normal and have zero means and unit variances. The correlation $\rho_{1,2}$ between $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ is strictly positive, and all the other correlations are zero.

Agent belief. The agent recursively applies Bayes' rule to compute the beliefs:

$$\Pi_t^j \propto f(s_t | M_t = m^j) \sum_{i=1}^{d} a_{ij} \Pi_{t-1}^i\,,$$

for all $j \in \{1, \ldots, d\}$ and $t \geq 1$.

State of the economy. At a given date t, the state of the economy is the mixed variable:

$$x_t = (M_t, \Pi_t) \in \mathcal{X} = \{m^1, \ldots, m^d\} \times \Delta_+^{d-1}, \tag{3.4}$$

Table 1: Dimension of the state space

	Incomplete Information ($\sigma_\delta > 0$)	Full Information ($\sigma_\delta = 0$)
State space $\mathcal{X}$	$\{m^1, \ldots, m^d\} \times \Delta_+^{d-1}$	$\{m^1, \ldots, m^d\}$
Dimension of $\mathcal{X}$	$d-1$	0
Likelihood	Unavailable	Available

where M_t is the Markov vector selected by nature, Π_t is the agent's belief, and $\Delta_+^{d-1} = \{\Pi \in \mathbb{R}_+^d | \sum_{i=1}^d \Pi_i = 1\}$ denotes the $(d-1)$–dimensional unit simplex. The transition kernel of the Markov state x_t is generally unavailable in closed-form.

In the special case where $\sigma_\delta = 0$, the agent is fully informed (FI) about the state of nature (as can be seen from (3.3)) and the belief vector reduces to $\Pi_t = \mathbb{1}_{M_t}$, where $\mathbb{1}_{M_t}$ denotes the vector whose j^{th} component is equal to 1 if $M_t = m^j$ and 0 otherwise. The log-likelihood under full information, denoted $\mathcal{L}_{FI}$, is available analytically (see CF 2007, 2008). Under incomplete information (II), when $\sigma_\delta > 0$, the log-likelihood, denoted $\mathcal{L}_{II}$ is unavailable. The topological dimensions of the state spaces under full and incomplete information are summarized in Table 1.

3.1.2 Observation Model

The statistician observes the log excess return computed by the agent according to her beliefs:

$$y_t = \ln \left[\frac{1 + Q(\Pi_t)}{Q(\Pi_{t-1})} \right] + s_{1,t} - r_f \,, \tag{3.5}$$

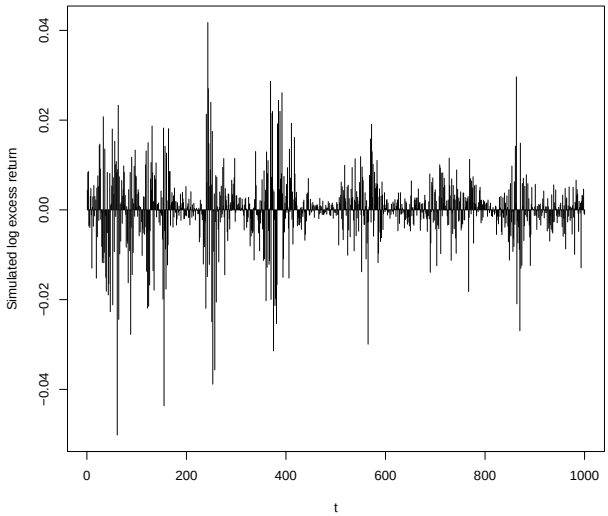

Figure 2: Simulated observations.

where r_f is the log interest rate and $Q(\Pi_t)$ is the stock's price-dividend ratio defined by $Q(\Pi_t) = \sum_{j=1}^{d} Q(m^j)\Pi_t^j$. The definition of the linear coefficients $Q(m^j)$ is provided in Appendix B. The observation density $f_Y(y_t|M_t, \Pi_t, Y_{t-1})$ is in general unavailable analytically.

3.2 Accuracy of the SOS filter

We now present the results of Monte Carlo simulations of the SOS particle filter.

Comparison with a Known Likelihood. We generate a simulated sample of excess returns of size $T = 1,000$ periods from the learning model in Section 3.1, with $\overline{k} = 3$ volatility components and signal noise parameter $\sigma_\delta = 0$. The values of the other parameters are given in Appendix B. Simulated data are illustrated in Figure 2. The

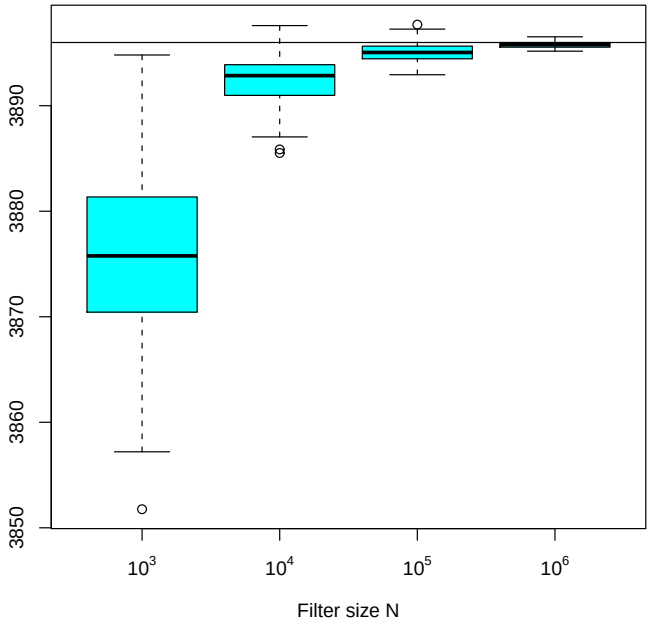

Figure 3: Boxplots of SOS likelihood estimates. The figure illustrates estimates of the full-information log-likelihood function $\mathcal{L}_{FI}$ for various filter sizes. All the filters are applied to the dataset reported in Figure 2. For each filter size $N = 10^3$, 10^4, 10^5, 10^6, one hundred SOS estimates of $\mathcal{L}_{FI}$ are computed using the quasi-Cauchy kernel. The horizontal line illustrates the true likelihood $\mathcal{L}_{FI}$.

state space dimension is zero (see Table 1) and the likelihood function is available in this case (see CF 2007). We apply to the simulated excess return series the SOS filter with the quasi-Cauchy kernel and bandwidth derived in section 2. Figure 3 illustrates boxplots of 100 estimates of the log-likelihood function for various values of the filter size $N = 10^3$, 10^4, 10^5, 10^6. The estimated log-likelihood increases with the filter size, as one expects from Jensen's inequality. SOS provides very accurate estimates of the likelihood function for $N \geq 10^6$.

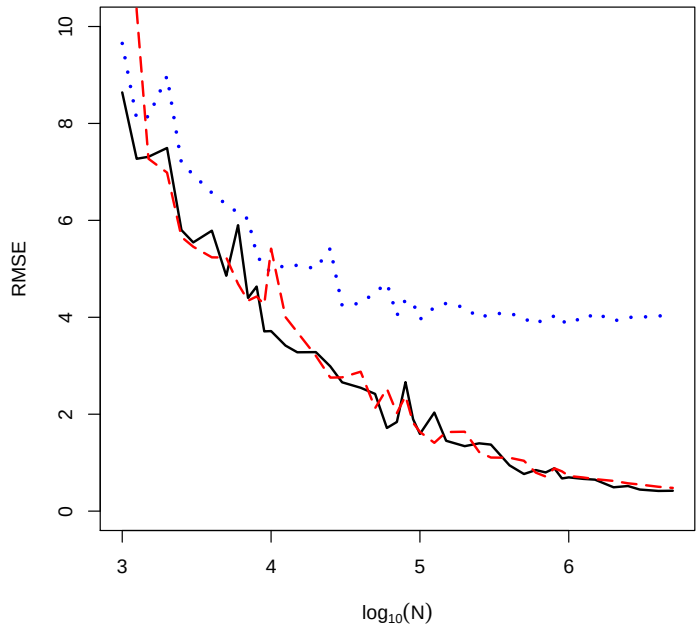

Figure 4: Plot of the RMSE as a function of the filter size N. We consider SOS with a correctly specified model ($\sigma_\delta = 0$, continuous line) and misspecified models with a higher state space dimension ($\sigma_\delta = 0.1$, dashed line, and 0.5, dotted line). Convergence of SOS is nearly identical for the correct ($\sigma_\delta = 0$) and the misspecified high-dimensional ($\sigma_\delta = 0.1$) models, even though the topological dimension of the state space is 0 for the former and 7 for the latter. The figure confirms the result of Theorem 1 that the convergence rate of SOS is independent of the dimension of the state space.

Defeat of the curse of dimensionality. We now increase the state space dimension and apply the SOS filter to data in Figure 2 using two misspecified models, that is with $\sigma_\delta = 0.1$ and 0.5, where the state space dimension is $n_X = 7$. We compare the convergence of the SOS estimation with increased state space dimension to the SOS estimation using the correctly specified model where $n_X = 0$. To investigate convergence in mean squared error of Theorem 1, we use the following measure:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{n=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t|Y_{t-1})\right]^2}, \tag{3.6}$$

where $f_Y(y_t|Y_{t-1})$ is the analytically available score function of the FI model. Figure 4 plots the RMSE as a function of the filter size N for SOS with correctly specified learning model and misspecified learning models with increased state space dimension. Convergence of SOS is nearly identical for the FI model and for the II model with $\sigma_\delta = 0.1$, even though II has a much larger state space. These findings confirm the result of Theorem 1 that the convergence rate of SOS is independent of the dimension of the state space.

4 Conclusion

In this paper, we have developed a powerful algorithm, the State-Observation Sampling filter, for general state-space models in which state-observation pairs can be conveniently simulated. Our method makes no assumption on the availability of the observation density and therefore expands the scope of sequential Monte Carlo

methods. The rate of convergence does not depend on the size of the state space, which shows that our filter defeats a form of the curse of dimensionality. Among many possible applications, SOS is useful to estimate the likelihood function, conduct likelihood-based specification tests, and generate forecasts.

The new filter naturally applies to nonlinear economies with agent learning of the type often considered in financial economics. In this context, SOS permits to track in real time both fundamentals and agent beliefs about fundamentals.

The paper opens multiple directions for future work. For instance, SOS can be used to price complex instruments, such as derivatives contracts, which crucially depend on the distribution of the hidden state. We can expand the role of learning in the analysis, for instance by letting the agent learn the parameter of the economy over time, or by conducting the joint online estimation of the structural parameter θ and the state of the economy x_t, as in Polson, Stroud, and Mueller (2008) and Storvik (2002). Applications to other fields, such as epidemiology or ecology, are also envisioned and will be the object of further research.

A Convergence of the SOS Filter (Section 2)

A.1 A Preliminary Result

In this appendix, we show the convergence of the SOS particle filter defined in section 2 as the number of particles N goes to infinity. Since the path Y_T is fixed, our focus is on simulation noise, and expectations in this section are over all the realizations of the random particle method. We begin by establishing the following result for a

given $N \geq 1$ and $t \geq 1$.

Lemma A1. *Assume that there exists $U_{t-1}(N)$ such that for every bounded measurable function $\Phi : \mathcal{X} \to \mathbb{R}$,*

$$\mathbb{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}\Phi(x_{t-1}^{(n)}) - \mathbb{E}[\Phi(x_{t-1})|Y_{t-1}]\right]^2\right\} \leq U_{t-1}(N)\|\Phi\|^2. \tag{A.1}$$

Let $U_t^*(N) = 2\kappa_t'^2 A(K)^2 h_t^4 + B(K)\kappa_t/(Nh_t^{n_Y}) + 2U_{t-1}(N)\kappa_t^2$. Then, the inequality

$$\mathbb{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}\Phi(\tilde{x}_t^{(n)})K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t|Y_{t-1})\mathbb{E}\left[\Phi(x_t)|\, Y_t\right]\right]^2\right\} \leq U_t^*(N)\|\Phi\|^2$$

holds for every bounded measurable function Φ.

Proof of Lemma A1. We consider the function

$$a_{t-1}(x_{t-1}) = \int \Phi(\tilde{x}_t)K_{h_t}(y_t - \tilde{y}_t)g(d\tilde{x}_t, d\tilde{y}_t|x_{t-1}, Y_{t-1}).$$

We note that

$$\begin{aligned}|a_{t-1}(x_{t-1})| &\leq \|\Phi\| \int K_{h_t}(y_t - \tilde{y}_t)g(d\tilde{x}_t, d\tilde{y}_t|x_{t-1}, Y_{t-1}) \\ &= \|\Phi\| \int K_{h_t}(y_t - \tilde{y}_t)f_Y(\tilde{y}_t|x_{t-1}, Y_{t-1})d\tilde{y}_t.\end{aligned}$$

The function a_{t-1} is therefore bounded above by $\kappa_t\, \|\Phi\|$.

The difference $Z = N^{-1}\sum_{n=1}^{N}\Phi(\tilde{x}_t^{(n)})K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t|Y_{t-1})\mathbb{E}\left[\Phi(x_t)|\, Y_t\right]$ is

the sum of the following three terms:

$$
\begin{aligned}
Z_1 &= \frac{1}{N}\sum_{n=1}^{N}\left[\Phi(\tilde{x}_t^{(n)})K_{h_t}(y_t-\tilde{y}_t^{(n)}) - a_{t-1}(x_{t-1}^{(n)})\right],\\
Z_2 &= \frac{1}{N}\sum_{n=1}^{N} a_{t-1}(x_{t-1}^{(n)}) - \int a_{t-1}(x_{t-1})\lambda(dx_{t-1}|Y_{t-1}),\\
Z_3 &= \int a_{t-1}(x_{t-1})\lambda(dx_{t-1}|Y_{t-1}) - f_Y(y_t|Y_{t-1})\mathbb{E}\left[\Phi(x_t)|\,Y_t\right].
\end{aligned}
$$

Let $\mathcal{X}_{t-1}^{(N)} = (x_{t-1}^{(1)},\ldots,x_{t-1}^{(N)})$ denote the vector of period$-(t-1)$ particles. Conditional on $\mathcal{X}_{t-1}^{(N)}$, Z_1 has a zero mean, while Z_2 and Z_3 are deterministic. Hence:

$$
\mathbb{E}(Z^2) = \mathbb{E}(Z_1^2) + \mathbb{E}[(Z_2+Z_3)^2] \leq \mathbb{E}(Z_1^2) + 2\mathbb{E}(Z_2^2) + 2\mathbb{E}(Z_3^2).
$$

Conditional on $\mathcal{X}_{t-1}^{(N)}$, the state-observation pairs $\{(\tilde{x}_t^{(n)},\tilde{y}_t^{(n)})\}_{n=1}^N$ are independent, and each $(\tilde{x}_t^{(n)},\tilde{y}_t^{(n)})$ is drawn from $g(\cdot|x_{t-1}^{(n)},Y_{t-1})$; the addends of $\Phi(\tilde{x}_t^{(n)})K_{h_t}(y_t-\tilde{y}_t^{(n)}) - a_{t-1}(x_{t-1}^{(n)})$ are thus independent and have mean zero. We infer that the conditional expectation of Z_1^2 is bounded above by:

$$
\frac{1}{N^2}\sum_{n=1}^{N}\int \Phi(\tilde{x}_t)^2 K_{h_t}(y_t-\tilde{y}_t)^2 g(d\tilde{x}_t,d\tilde{y}_t|x_{t-1}^{(n)},Y_{t-1}) \leq \frac{\kappa_t\|\Phi\|^2}{N}\int K_{h_t}(y_t-\tilde{y}_t)^2 d\tilde{y}_t.
$$

We apply the change of variable $u=(y_t-\tilde{y}_t)/h_t$:

$$
\int K_{h_t}(y_t-\tilde{y}_t)^2 d\tilde{y}_t = \frac{B(K)}{h_t^{n_Y}},
$$

and infer that $\mathbb{E}(Z_1^2) \leq \|\Phi\|^2 B(K)\kappa_t/(Nh_t^{n_Y})$.

Since the function $a_{t-1}(x_{t-1})$ is bounded above by $\kappa_t \|\Phi\|$, we infer from (A.1) that: $\mathbb{E}(Z_2^2) \leq U_{t-1}(N)\kappa_t^2 \|\Phi\|^2$.

Finally, we observe that $f_Y(y_t|Y_{t-1})\mathbb{E}\left[\Phi(x_t)|\, Y_t\right] = \int \Phi(x_t) f_Y(y_t|x_t, Y_{t-1})\lambda(dx_t|Y_{t-1})$, and therefore

$$\begin{aligned} Z_3 &= \int \Phi(x_t) \left\{ \int K_{h_t}(y_t - \tilde{y}_t)[f_Y(\tilde{y}_t|x_t, Y_{t-1}) - f_Y(y_t|x_t, Y_{t-1})]d\tilde{y}_t \right\} \lambda(dx_t|Y_{t-1}) \\ &= \int \Phi(x_t) \left\{ \int K(u)[f_Y(y_t - h_t u|x_t, Y_{t-1}) - f_Y(y_t|x_t, Y_{t-1})]du \right\} \lambda(dx_t|Y_{t-1}). \end{aligned}$$

Note that $\left|\int K(u)[f_Y(y_t - h_t u|x_t, Y_{t-1}) - f_Y(y_t|x_t, Y_{t-1})]du\right| \leq \kappa_t' A(K) h_t^2$. Hence $|Z_3| \leq \kappa_t' A(K) h_t^2 \|\Phi\|$ and therefore $\mathbb{E}(Z_3^2) \leq \kappa_t'^2 A(K)^2 h_t^4 \|\Phi\|^2$. We conclude that the lemma holds. *Q.E.D.*

A.2 Proof of Theorem 1

The proof of (2.5) proceeds by induction. When $t = 0$, the particles are drawn from the prior λ_0, and the conditional expectation is computed under the same prior. Hence the property (2.5) holds with $U_0(N) = 1/N$.

We now assume that the property (2.5) holds at date $t-1$. The estimation error $X = N^{-1} \sum_{n=1}^N \Phi(x_t^{(n)}) - \mathbb{E}[\Phi(x_t)|Y_t]$ is the sum of:

$$\begin{aligned} X_1 &= \frac{1}{N} \sum_{n=1}^N \Phi(x_t^{(n)}) - \sum_{n=1}^N p_t^{(n)} \Phi(\tilde{x}_t^{(n)}). \\ X_2 &= \left[\sum_{n=1}^N p_t^{(n)} \Phi(\tilde{x}_t^{(n)}) \right] \left[\frac{f_Y(y_t|Y_{t-1}) - N^{-1} \sum_{n'=1}^N K_{h_t}(y_t - \tilde{y}_t^{(n')})}{f_Y(y_t|Y_{t-1})} \right], \\ X_3 &= \frac{1}{N f_Y(y_t|Y_{t-1})} \sum_{n=1}^N \Phi(\tilde{x}_t^{(n)}) K_{h_t}(y_t - \tilde{y}_t^{(n)}) - \mathbb{E}[\Phi(x_t)|Y_t]. \end{aligned}$$

The first term, X_1, corresponds to step 3 resampling, the second term to the normalization of the resampling weights, and the third term to the error in the estimation of Φ using the nonnormalized weights.

Conditional on $\{(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})\}_{n=1}^N$, the particles $x_t^{(n)}$ are independent and identically distributed, and X_1 has mean zero. We infer that $\mathbb{E}[X_1^2|\{\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)}\}_{n=1}^N] \leq \|\Phi\|^2/N$, and therefore $\mathbb{E}(X_1^2) \leq \|\Phi\|^2/N$. Note that when we use stratified, residual or combined stratified-residual resampling in step 3, the inequality $\mathbb{E}(X_1^2) \leq \|\Phi\|^2/N$ remains valid, and smaller upper bounds can also be derived.[3]

Conditional on $\{(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})\}_{n=1}^N$, X_2 and X_3 are deterministic variables. The mean squared error satisfies:

$$\mathbb{E}(X^2) = \mathbb{E}(X_1^2) + \mathbb{E}[(X_2 + X_3)^2] \leq \mathbb{E}(X_1^2) + 2\mathbb{E}(X_2^2) + 2\mathbb{E}(X_3^2).$$

We note that $|X_2| \leq \|\Phi\|[f_Y(y_t|Y_{t-1})]^{-1} \left| f_Y(y_t|Y_{t-1}) - \sum_{n'=1}^N K_{h_t}(y_t - \tilde{y}_t^{(n')})/N \right|$. Using the induction hypothesis at date $t-1$, we apply Lemma A1 with $\Phi \equiv 1$ and obtain that $\mathbb{E}(X_2^2)$ is bounded above by:

$$\frac{U_t^*(N)\|\Phi\|^2}{[f_Y(y_t|Y_{t-1})]^2}. \tag{A.2}$$

Lemma A1 implies that $\mathbb{E}(X_3^2)$ is also bounded above by (A.2). We conclude that

[3]See Cappé, O., Moulines, E., and T. Rydén (2005, ch. 7) for a detailed discussion of sampling variance.

$\mathbb{E}(X^2) \leq U_t(N)\|\Phi\|^2$, where $U_t(N) = 4U_t^*(N)[f_Y(y_t|Y_{t-1})]^{-2} + N^{-1}$, or equivalently

$$U_t(N) = \frac{4}{[f_Y(y_t|Y_{t-1})]^2}\left[2\kappa_t'^2 A(K)^2 h_t^4 + \frac{B(K)\kappa_t}{N h_t^{n_Y}} + 2U_{t-1}(N)\kappa_t^2\right] + \frac{1}{N}. \tag{A.3}$$

This establishes part (2.5) of the theorem. From (2.5) and Lemma A1 with $\Phi \equiv 1$, (2.4) follows.

Assume now that the bandwidth is a function of N, and that assumption 3 holds. A simple recursion implies that $\lim_{N\to\infty} U_t(N) = 0$ for all t. The mean squared error converges to zero for any bounded measurable function Φ.

We now characterize the rate of convergence. Given $U_{t-1}(N)$, we know that the coefficient $U_t(N)$ defined by (A.3) is minimal if

$$h_t = N^{-1/(n_Y+4)}\left[\frac{\kappa_t n_Y B(K)}{8\kappa_t'^2 A(K)^2}\right]^{1/(n_Y+4)}. \tag{A.4}$$

More generally, if the bandwidth sequence is of the form $h_t(N) = h_t(1)/N^{-1/(n_Y+4)}$, then $U_t(N)$ is of the form:

$$U_t(N) = u_{1,t}N^{-4/(n_Y+4)} + u_{2,t}U_{t-1}(N) + N^{-1}. \tag{A.5}$$

where $u_{1,t}$ and $u_{2,t}$ are finite nonnegative coefficients.[4] By a simple recursion, $U_t(N)$ is of order $N^{-4/(n_Y+4)}$ for all t. *Q.E.D.*

[4]We verify that $u_{1,t} = 4[f(y_t|Y_{t-1})]^{-2}\left[2\kappa_t'^2 h_t(1)^4 A(K)^2 + B(K)\kappa_t h_t(1)^{-n_Y}\right]$ and $u_{2,t} = 8\kappa_t^2[f(y_t|Y_{t-1})]^{-2}$.

A.3 Proof of Proposition 1

We first investigate the bias. Note that:

$$\begin{aligned}\mathbb{E}[K_{h_t}(y_t - \tilde{y}_t^{(n)})|x_{t-1}^{(n)}, Y_{t-1}] &= \int K_{h_t}(y_t - \tilde{y}_t) f_Y(\tilde{y}_t|x_{t-1}^{(n)}, Y_{t-1}) d\tilde{y}_t \\ &= \int K(u) f_Y(y_t - h_t u|x_{t-1}^{(n)}, Y_{t-1}) du.\end{aligned} \tag{A.6}$$

From a Taylor expansion of $f_Y(y_t - h_t u|x_{t-1}^{(n)}, Y_{t-1})$ around $f_Y(y_t|x_{t-1}^{(n)}, Y_{t-1})$ we have:

$$\mathbb{E}[K_{h_t}(y_t - \tilde{y}_t^{(n)})|x_{t-1}^{(n)}, Y_{t-1}] = f_Y(y_t|x_{t-1}^{(n)}, Y_{t-1}) + \frac{h_t^2}{2} \int K(u) u^T \frac{\partial^2 f_Y}{\partial y_t \partial y_t^T}(y_t|x_{t-1}^{(n)}, Y_{t-1}) u \, du + \mathcal{O}(h_t^3) \,.$$

By construction, $\mathbb{E}_K(u) = \int uK(u)du = 0$. We observe that

$$\int K(u) u^T \frac{\partial^2 f_Y}{\partial y_t \partial y_t^T}(y_t|x_{t-1}^{(n)}, Y_{t-1}) u \, du = \mathrm{tr}\left[\frac{\partial^2 f_Y}{\partial y_t \partial y_t^T}(y_t|x_{t-1}^{(n)}, Y_{t-1}) \mathrm{Var}_K(u)\right].$$

The bias is therefore

$$\mathbb{E}(\hat{f}_t|\mathcal{X}_{t-1}^{(N)}, Y_{t-1}) - f_t = \frac{h_t^2}{2} \mathrm{tr}\left[\frac{\partial^2 f_t}{\partial y_t \partial y_t^T} \mathrm{Var}_K(u)\right] + \mathcal{O}(h_t^3) \,.$$

Next, we note that the variance $\mathrm{Var}\Big[K_{h_t}(y_t - \tilde{y}_t^{(n)})|x_{t-1}^{(n)}, Y_{t-1}\Big]$ can be written as

$$\int \left[K_{h_t}(y_t - \tilde{y}_t)\right]^2 f_Y(\tilde{y}_t|x_{t-1}^{(n)}, Y_{t-1}) d\tilde{y}_t - \left\{\mathbb{E}\left[K_{h_t}(y_t - \tilde{y}_t^{(n)})|x_{t-1}^{(n)}, Y_{t-1}\right]\right\}^2.$$

Hence

$$\begin{aligned}\mathrm{Var}\Big[K_{h_t}(y_t-\tilde{y}_t^{(n)})|x_{t-1}^{(n)},Y_{t-1}\Big] =& \frac{1}{h_t^{n_Y}}\int [K(u)]^2 f_Y(y_t-h_t u|x_{t-1}^{(n)},Y_{t-1})du+\mathcal{O}(1)\\ =& \frac{B(K)}{h_t^{n_Y}} f_Y(y_t|x_{t-1}^{(n)},Y_{t-1})+\mathcal{O}(h_t^{-n_Y+1})\,.\end{aligned}$$

We conclude that (2.8) holds. *Q.E.D.*

B Linear coefficients and parameter values in the learning model of CF (2007)

We choose parameter values that provide empirically plausible results in the analysis of U.S. daily excess returns reported in CF (2007). We let $m_0 = 1.7$, $\gamma_{\overline{k}} = 0.06$, $b = 2$, $g_C = 0.75$ basis point (bp) (or 1.18% per year), $r_f = 0.42$ bp per day (1% per year), $g_D - r_f = 0.5$ bp per day (about 1.2% per year), $\sigma_C = 0.189\%$ (or 2.93% per year), $\overline{\sigma}_D = 0.70\%$ per day (about 11% per year), and $\rho_{1,2} = 0.6$.

The linear coefficients are given by $\big(Q(m^1),\ldots,Q(m^d)\big)^T = (I-B)^{-1}\iota-\iota$, where $B=(b_{ij})_{1\le i,j\le d}$ is the matrix with components $b_{ij} = a_{i,j}\exp\big[g_D - r_f - \alpha\,\rho_{1,2}\,\sigma_C\,\sigma_D(m^j)\big]$ and $\iota = (1,\ldots,1)^T$. The risk aversion coefficient α is chosen so that the average price-dividend ratio is $\overline{Q} = d^{-1}\sum_{i=1}^d Q(m^i) = 6000$ in daily units (25 in yearly units).

References

Acton, S.T. and N. Ray (2006). *Biomedical Image Analysis: Tracking.* Morgan and Claypool Publishers.

Andrieu, C., Doucet, A., and R. Holenstein (2010). Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society B* **72(3)**, 269–342.

Calvet, L. E., and A. J. Fisher (2007). Multifrequency News and Stock Returns. *Journal of Financial Economics* 86: 178-212.

Calvet, L. E., and A. J. Fisher (2008). *Multifractal Volatility: Theory, Forecasting and Pricing.* Elsevier – Academic Press.

Cappé, O., Moulines, E., and T. Rydén (2005). *Inference in Hidden Markov Models.* Springer.

Crisan, D., and Doucet, A. (2002). A Survey of Convergence Results on Particle Filtering Methods for Practitioners. *IEEE Transactions on Signal Processing* **50(3)** 736–746.

Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.* Springer-Verlag, New York.

Fan, J., and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, New York.

Fearnhead, P., and P. Clifford (2003). On-line Inference for Hidden Markov Models via Particle Filters. *Journal of the Royal Statistical Society B* **65**, 887899.

Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramirez, J., and M. Uribe (2011). Risk Matters: The Real Effects of Stochastic Volatility. Forthcoming, *American Economic Review.*

Gilks, W. R., and C. Berzuini (2001). Following a Moving Target – Monte Carlo Inference for Dynamic Bayesian Models. *Journal of the Royal Statistical Society B* **63**, 127–146.

Godsill, S. J., Doucet, A., and M. West (2004). Monte Carlo Smoothing for Nonlinear Time Series. *Journal of the American Statistical Association* **99**, 156–168.

Gordon, N., Salmond, D., and A. F. Smith (1993). Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE Proceedings F* **140**, 107–113.

Hansen, L., Polson, N., and T. Sargent (2011). Nonlinear Filtering and Learning Dynamics. Forthcoming, *Journal of Business and Economic Statistics.*

Hart, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer-Verlag, New York.

Jasra, A., Singh, S., Martin, J. and E. McCoy (2011). Filtering via Approximate Bayesian Computation. Forthcoming, *Statistics and Computing.*

Johannes, M., and N. Polson (2009), "Particle Filtering," in *Handbook of Financial Time Series*, ed. by T. Andersen, R. A. Davis, J.-P. Kreiss, and Th. Mikosch, 1015–1028. Springer.

Kim, S., Shephard, N., and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* **65(3)**, 361–393.

Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics* **5**, 1–25.

Kitagawa, G. (1998). Self-Organizing State Space Model. *Journal of the American Statistical Association* **93**, 1203–1215.

Liu, J.S. and R. Chen (1995). Blind Deconvolution via Sequential Imputations. *Journal of the American Statistical Association* **90**, 567–576.

Liu, J.S. and R. Chen (1998). Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association* **93**, 1032–1044.

Liu, D., Lu, T., Niu X. and H. Wu (2011). Mixed-Effects State-Space Models for Analysis of Longitudinal Dynamic Systems. *Biometrics* **67**, 476–485.

Newman, K.B., Fernández, C., Thomas, L. and S.T. Buckland (2009). Monte Carlo Inference for State-Space Models of Wild Animal Populations. *Biometrics* **65**, 572-583.

Pitt, M., and N. Shephard (1999). Filtering Via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association* **94**, 590–599.

Polson, N., Stroud, J., and P. Mueller (2008). Practical Filtering with Sequential Parameter Learning. *Journal of the Royal Statistical Society B* **70(2)**, 413–428.

Rossi, V., and J.-P. Vila (2006). Nonlinear Filtering in Discrete Time: A Particle Convolution Approach. *Annales de l'Institut de Statistique de l'Université de Paris* **50(3)**, 71–102.

Rossi, V., and J.-P. Vila (2009). Convolution Particle Filter for Parameter Estimation in General State-Space Models. *IEEE Transactions on Aerospace and Electronic Systems* **45(3)**, 1063– 1072.

Storvik, G. (2002). Particle Filters in State Space Models with the Presence of Unknown Static Parameters. *IEEE Transactions on Signal Processing* **50**, 281–289.

Tistarelli, M. (2009). *Handbook of Remote Biometrics: For Surveillance and Security.* Springer.