Accurate Methods for Approximate Bayesian Computation Filtering

Laurent Calvet and Veronika Czellar*

February 2012

Abstract

The Approximate Bayesian Computation (ABC) filter extends the particle filtering methodology to general state-space models in which the density of the observation conditional on the state is intractable. We provide an exact upper bound for the mean squared error of the ABC filter and show that under appropriate bandwidth and kernel specifications, ABC converges to the target distribution as the number of particles goes to infinity. The optimal convergence rate decreases with the dimension of the observation space but is invariant to the complexity of the state space. We also show that the usual adaptive bandwidth used in the ABC literature leads to an inconsistent filter. We develop a plug-in rule for the bandwidth and demonstrate the good accuracy of the resulting filter on a multifractal asset pricing model with investor learning. Despite the recent criticism of ABC methods, we find that under

*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 CEF, CORE, the Econometric Society World Congress, ISBIS, the Toulouse School of Economics, the University of Chicago, the University of Evry, the University of Geneva, and the Second HEC Finance and Statistics Conference. 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.

appropriate kernel and bandwidth specifications, ABC filtering is a very powerful tool for model selection when the likelihood function is unavailable in closed-form.

Keywords: Bandwidth; Kernel density estimation; Likelihood estimation; Model selection; Particle filter; State-space model.

1 Introduction

Since their introduction by Gordon et al. (1993), particle filters have considerably expanded the range of applications of hidden Markov models and now pervade fields as diverse as biomedical research (Acton & Ray, 2006; Liu et al., 2011), biometrics (Tistarelli, 2009), ecology (Newman et al., 2009), finance (Kim et al., 1998; Johannes & Polson, 2009) and macroeconomics (Fernández-Villaverde et al., 2011; Hansen et al., 2011). These methods provide estimates of the distribution of a hidden Markov state x_t conditional on a time series of observations $y_{1:t} = (y_1, \dots, y_t)$, $y_t \in \mathbb{R}^{n_Y}$, by way of a set of particles $(x_t^{(1)}, \dots, x_t^{(N)})$. 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. In the selection phase, a new data point y_t is observed and the particles are resampled using weights that are proportional to the *observation density*

$$f_Y(y_t \mid \tilde{x}_t^{(n)}, y_{1:t-1}). \tag{1.1}$$

Under a wide range of conditions, the sample mean $N^{-1} \sum_{n=1}^{N} \Phi(x_t^{(n)})$ converges to $E\{\Phi(x_t) \mid y_{1:t}\}$ as $N \to \infty$ for any bounded measurable function Φ (Chopin, 2004; Crisan & Doucet, 2002).

A common feature of particle filters is the requirement that the observation density (1.1) be available analytically up to a normalizing constant. This condition does not hold in complex environments such as, for instance, state-space models of animal populations, HIV dynamics in the presence of mixed effects, or dynamic equilibria with agent learning. A possible solution, proposed by Rossi & Vila (2006, 2009), is to estimate each conditional density (1.1) nonparametrically by sampling pseudo-obervations from $f_Y(\cdot \mid \tilde{x}_t^{(n)}, y_{1:t-1})$. The resulting convolution filter is numerically challenging because N conditional densities must be evaluated every period. It 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 of the observation space n_Y.

An alternative, proposed by Jasra et al. (2011), is to use a filter based on Approximate Bayesian Computation. It consists of sampling a state-pseudo observation couple $(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})$ from each date-$(t-1)$ particle $x_{t-1}^{(n)}$ in the mutation phase. In the selection phase, new particles are obtained by resampling from $(\tilde{x}_t^{(1)}, \ldots, \tilde{x}_t^{(N)})$ with probability:

$$p_t^{(n)} \propto \mathbb{1}_{\|\tilde{y}_t^{(n)} - y_t\| < \varepsilon}\,, \tag{1.2}$$

where ε is a tolerance level, $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^{n_Y}$ and $\mathbb{1}_A = 1$ if A is true and 0 otherwise. Jasra et al. (2011) show that: (i) for a fixed ε, the filter converges to a biased distribution as the number of particles N goes to infinity, and (ii) the asymptotic bias itself goes to zero as the tolerance ε goes to 0.

Double convergence in ε and N, while typical of the ABC literature, is an obstacle to implementation because the tolerance level and the particle size N must be jointly selected in practice. Marin et al. (2012) thus conclude in a survey of ABC methods that: "the convergence results obtained so far are unpractical [...] Obtaining exact error bounds for positive tolerances and finite sample sizes would bring a strong improvement in both the implementation of the method and the assessment of its worth." Given the enormous potential of the ABC filters, one would like to know how to select the tolerance level ε as a function the number of particles in order to obtain proper convergence. These answers are essential for reliable automated applications of ABC filtering.

This paper proposes answers to these questions by drawing insights from the nonparametric literature. A similar approach has been frutifully adopted for constructing summary statistics in an as yet unpublished 2011 Lancaster University technical report by Fearnhead and Prangle. We interpret the tolerance parameter ε of the ABC filter as a bandwidth and allow it to be a time-varying function $h_t(N)$ of the filter size. The nonparametric kernel density estimation literature suggests that the bandwidth should decrease with the filter size N, so that the asymptotic bias and simulation error can both vanish as N gets large, but not too quickly in order to avoid instability. This static intuition extends to the ABC filter. In order to avoid degeneracy problems, we consider a strictly positive kernel ($K(u) > 0$ for all $u \in \mathbb{R}^{n_Y}$) in the definition of the importance weights:

$$p_t^{(n)} \propto K\left\{\frac{y_t^{(n)} - y_t}{h_t(N)}\right\}.$$

We provide conditions on the bandwidth $h_t(N)$ under which the filter converges in L^2 to the target distribution as the number of particles N goes to infinity. These conditions also guarantee that the ABC filter provides consistent estimates of the likelihood function as N goes to infinity. The adaptive method commonly used in the ABC literature does not meet these sufficient conditions, and in fact leads to an inconsistent ABC filter.

We show that using the optimal decay rate of the bandwidth, the root mean squared error of moments computed using the filter decays at the optimal rate $N^{-2/(n_Y+4)}$, that is at the same rate as the kernel density estimator of a random vector on $\mathbb{R}^{n_Y}$ (Fan & Yao, 2003; Hall et al., 1995). The asymptotic rate of convergence is thus invariant to the dimension of the state space, indicating that the ABC filter overcomes a form of the curse of dimensionality. We propose a plug-in rule for the choice of bandwidth that provides good results for the estimation of the likelihood function.

We demonstrate the accuracy of the ABC filter on the multifractal asset pricing model with agent learning of Calvet & Fisher (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 Calvet & Fisher (2007) model is that the state variable is mixed and easily scalable (discrete M_t and continuous Π_t), which allows us to investigate the accuracy of ABC 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 ABC 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. We also show that using kernel and plug-in bandwidth satisfying conditions described in Sections 2 and 3, enhances ABC methods with powerful model selection properties.

The paper is organized as follows. Section 2 presents the ABC filter. In section 3, we analyze the convergence of the filter and propose a selection rule for the bandwith. Section 4 applies these methods to a multifractal asset pricing model with agent learning.

2 ABC Filtering

We consider a Markov process x_t defined on a measurable space $(\mathcal{X}, \mathcal{F}_X)$. Time is discrete and indexed by $t = 0, 1, \ldots, \infty$. For expositional simplicity, we assume for now that $\mathcal{X} = \mathbb{R}^{n_X}$. Let $y_t \in \mathbb{R}^{n_Y}$ denote the observation at date t and $y_{1:t-1} = (y_1, \ldots, y_{t-1})$, the vector of observations up to date $t-1$. The building blocks of the model are the conditional density of a period-t state-observation pair, $f_{X,Y}(x_t, y_t \mid x_{t-1}, y_{1:t-1})$, and a prior λ_0 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: $\lambda(x_t \mid y_{1:t})$ at all $t \geq 1$.

The principle of the ABC 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:

$$\left(\tilde{y}_t^{(n)}, \tilde{x}_t^{(n)}, x_{t-1}^{(n)}\right) \mid y_{1:t} \sim \frac{\delta(y_t - \tilde{y}_t^{(n)})\, f_{X,Y}(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)} \mid x_{t-1}^{(n)}, y_{1:t-1})\, f_X(x_{t-1}^{(n)} \mid y_{1:t-1})}{f_Y(y_t \mid y_{1:t-1})}, \tag{2.1}$$

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)* $\int K(u)du = 1$*; (ii)* $\int uK(u)du = 0$*; (iii)* $A(K) = \int \|u\|^2 K(u)du < \infty$*; and (iv)* $B(K) = \int [K(u)]^2 du < \infty$.

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 suggests the following algorithm.

ABC filter

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

Step 2 (Importance weights): Given the new data point y_t, compute

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

Step 3 (Multinomial resampling): For every $n = 1, \ldots, N$, 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 variance of multinomial resampling in step 3 can be reduced and computational speed can be improved by using a combination of residual (Liu & Chen, 1998) and stratified (Kitagawa, 1996) resampling that we use in Section 4. That is, we set $\lfloor Np_t^{(n)} \rfloor$ particles equal to $\tilde{x}_t^{(n)}$ for every n where $\lfloor \cdot \rfloor$ denotes the floor of a real number. The remaining $N_{r,t} = N - \sum_{n=1}^{N} \lfloor Np_t^{(n)} \rfloor$ particles are selected by stratified sampling, as we now explain. Let $q_t^{(n)} = (Np_t^{(n)} - \lfloor Np_t^{(n)} \rfloor)/N_{r,t}$ for all n. For every $k = 1, \ldots, N_{r,t}$, we draw $\tilde{U}_k$ from the uniform distribution on $((k-1)/N_{r,t}, k/N_{r,t}]$ and select the particle $\tilde{x}_t^{(n)}$ such that $\tilde{U}_k \in (\sum_{j=1}^{n-1} q_t^{(j)}, \sum_{j=1}^{n} q_t^{(j)}]$. The convergence proof below applies equally well to this alternative.

The ABC filter applies just as well to dynamic systems with a general measurable state space $\mathcal{X}$. The building blocks of the model are the conditional probability measure associated with the period-t state-observation pair, $g(\cdot \mid x_{t-1}, y_{1:t-1})$ and a prior measure λ_0 over the state space. The ABC filter is defined by sampling $(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})$ from the conditional measure $g(\cdot \mid x_{t-1}^{(n)}, y_{1:t-1})$ in step 1 and by implementing steps 2 and 3 as above.

3 Mean Square Convergence and Bandwidth Selection

3.1 Convergence

We now specify conditions under which for a fixed history $y_{1:T} = (y_1, \ldots, y_T)$, $T \leq \infty$, the ABC filter converges in mean squared error to the target conditional measure $\lambda(\cdot \mid y_{1:t})$ as the number of particles N goes to infinity.

Assumption 2 (Conditional Distributions) *The conditional density $f_Y(\cdot \mid x_{t-1}, y_{1:t-1})$ exists and satisfies $\kappa_t = \sup\{f_Y(\tilde{y}_t \mid x_{t-1}, y_{1:t-1}) \mid (x_{t-1}, \tilde{y}_t) \in \mathcal{X} \times \mathbb{R}^{n_Y}\} < \infty$. Moreover, the observation density $f_Y(\cdot \mid x_t, y_{1:t-1})$ is well-defined and there exists $\kappa_t' \in \mathbb{R}_+$ such that:*

$$\left| f_Y(\tilde{y}_t \mid x_t, y_{1:t-1}) - f_Y(y_t \mid x_t, y_{1:t-1}) - \frac{\partial f_Y}{\partial y_t^{\mathrm{T}}}(y_t \mid x_t, y_{1:t-1})(\tilde{y}_t - y_t) \right| \leq \kappa_t' \|\tilde{y}_t - y_t\|^2$$

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) $\lim_{N\to\infty} h_t(N) = 0$, and (ii) $\lim_{N\to\infty} N\{h_t(N)\}^{n_Y} = +\infty$, for all $t = 1, \ldots, T$.*

We establish the following result in the Appendix.

Theorem 1 (Mean Square Convergence of the ABC Filter) *Under assumptions 1 and 2 and for every t and $N \geq 1$*

$$E\Big[\Big\{\frac{1}{N}\sum_{n=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t \mid y_{1:t-1})\Big\}^2\Big] \leq \frac{f_Y(y_t \mid y_{1:t-1})^2}{4} U_t(N)\,, \qquad (3.1)$$

with $U_t(N) = 4f_Y(y_t \mid y_{1:t-1})^{-2}\{2\kappa_t'^2 A(K)^2 h_t^4 + B(K)\kappa_t/(Nh_t^{n_Y}) + 2U_{t-1}(N)\kappa_t^2\} + N^{-1}$ and $U_0(N) = 1/N$, and the expectation is over all the realizations of the random particle method. Furthermore, for any bounded measurable function, $\Phi : \mathcal{X} \to \mathbb{R}$,

$$E\Big(\Big[\frac{1}{N}\sum_{n=1}^{N} \Phi(x_t^{(n)}) - E\{\Phi(x_t) \mid Y_t\}\Big]^2\Big) \leq U_t(N)\|\Phi\|_\infty^2, \qquad (3.2)$$

where $\|\Phi\|_\infty = \sup_{x\in\mathcal{X}} |\Phi(x)|$. If assumption 3 also holds, then $\lim_{N\to\infty} U_t(N) = 0$. 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 $N^{-4/(n_Y+4)}$.

The convergence rate of ABC is the same as the convergence rate of a kernel density estimator on the observation space $\mathbb{R}^{n_Y}$. ABC therefore defeats the curse of dimensionality with respect to the size of the state space $\mathcal{X}$. Moreover, by (3.1), the kernel estimator

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

converges to the conditional density of y_t given past observations. Consequently, we can estimate the loglikelihood function by $\sum_{t=1}^{T} \ln \hat{f}_t$.

Remark. The adaptive bandwidth used in the ABC literature (Del Moral et al., 2011; Jasra et al., 2011) consists of setting $h_t(N)$ equal to the αth quantile of the distribution of the $\{\|y_t - \tilde{y}_t^{(n)}\|\}$. As $N \to \infty$, the adaptive bandwidth converges to a positive finite limit and assumption 3 is violated. This method is therefore inconsistent as $N \to \infty$, as is illustrated in section 4.

3.2 Bandwidth Selection

Computing the vector $\{h_t(N)\}_{t=1}^T$ that minimizes the mean squared error of the likelihood estimator is a seemingly intractable high-dimensional problem. For this reason, we consider a much simpler criterion. At each date t, we select the bandwidth that minimizes the integrated MSE in the estimation of $f_Y(y_t \mid y_{1:t-1}) = E\{f_Y(y_t \mid x_{t-1}, y_{1:t-1}) \mid y_{1:t-1}\}$ under the simplifying assumption that the conditional measure $\lambda(\cdot \mid y_{1:t-1})$ 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 $f_t = N^{-1}\sum_{n=1}^N f_Y(y_t \mid x_{t-1}^{(n)}, y_{1:t-1})$. The properties of the mean squared error

$$E\{(\hat{f}_t - f_t)^2 \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}\} = \text{var}(\hat{f}_t \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}) + \left\{E(\hat{f}_t \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}) - f_t\right\}^2$$

are summarized in the following proposition established in the Appendix.

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

$$E(\hat{f}_t \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}) - f_t = h_t^2 \text{tr}\{H_t(f) var_K(u)\}/2 + \mathcal{O}(h_t^3)\,, \tag{3.3}$$

$$var(\hat{f}_t \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}) = B(K) f_t/(N h_t^{n_Y}) + \mathcal{O}(h_t^{-n_Y+1})\,. \tag{3.4}$$

where $H_t(f) = N^{-1}\sum_{n=1}^N \partial^2 f_Y/(\partial y_t \partial y_t^{\mathrm{T}})(y_t \mid x_{t-1}^{(n)}, y_{1:t-1})$, tr *is the trace operator and $var_K(u) = \int u u^{\mathrm{T}} K(u) du$.*

The mean squared error $E\{(\hat{f}_t - f_t)^2 \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}\}$ can therefore be approximated by $B(K)f_t/(Nh_t^{n_Y}) + h_t^4[\text{tr}\{H_t(f)\text{var}_K(u)\}]^2/4$. The mean squared error integrated on y_t is minimal for

$$h_t^*(N) = \left(\frac{B(K)\, n_Y}{N \int \left[\operatorname{tr}\left\{ H_t(f) \operatorname{var}_K(u) \right\} \right]^2 dy_t} \right)^{1/(n_Y+4)} . \tag{3.5}$$

The formula reduces to the standard kernel density estimation plug-in rule if f_t is Gaussian and the kernel satisfies $\operatorname{var}_K(u) = A(K) n_Y^{-1} I_{n_Y}$, where I_{n_Y} is the n_Y-dimensional identity matrix.

Example (Quasi-Cauchy kernel). When $n_Y = 1$ we can use the quasi-Cauchy kernel, $K(u) = (1 + Cu^2)^{-2}$ with $C = (\pi/2)^2$, which does not contain exponentials and is hence computationally cheaper than the Gaussian kernel. We obtain $A(K) = 4/\pi^2$ and $B(K) = 5/8$. 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

$$h_t^*(N) = \sigma_t \left(\frac{5\pi^{9/2}}{48\, N} \right)^{1/5} , \tag{3.6}$$

with σ_t estimated by the sample standard deviation of the simulated observations $\{\tilde{y}_t^{(n)}\}_{n=1}^N$.

4 Application to a Multifractal Financial Model

4.1 Specification

We now apply our methodology to the multifractal equilibrium model of equity returns developed in Calvet & Fisher (2007, 2008). The model contains three levels of information corresponding to nature, an agent and the statistician, as illustrated in Figure 1. Nature selects a vector M_t that drives the volatility of the economy. The agent observes a signal s_t and uses Bayes' rule to impute the conditional distribution of nature's vector M_t. We refer to this distribution as the agent's belief Π_t. The agent sets the return on the stock as a function of her signals and beliefs. The statistician observes the stock return and uses the ABC filter to track the economy's hidden

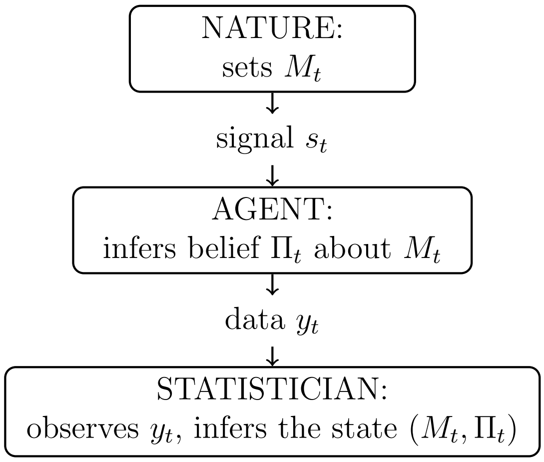

Figure 1: Information structure.

state (M_t, Π_t). We now provide a more detailed description of the specification we use in the Monte Carlo simulations.

State model. Nature's vector $M_t = (M_{k,t})_{1 \le k \le \overline{k}} \in \{m_0, 2 - m_0\}^{\overline{k}}$ takes $d = 2^{\overline{k}}$ possible values denoted $m^1, \ldots, m^d$. The transition matrix $A = (a_{ij})$ has elements

$$a_{ij} = \mathrm{pr}(x_{t+1} = m^j \mid x_t = m^i) = \prod_{k=1}^{\overline{k}} \left\{ (1 - \frac{\gamma_k}{2}) \mathbb{1}_{m_k^i = m_k^j} + \frac{\gamma_k}{2} \mathbb{1}_{m_k^i \neq m_k^j} \right\} \tag{4.1}$$

where $\gamma_k = 1 - (1 - \gamma_1)^{b^{k-1}}$ for all $k = 1, \ldots, \overline{k}$.

The agent receives a conditionally Gaussian signal $s_t = (s_{1,t}, \ldots, s_{\overline{k}+2,t})^{\mathrm{T}} \sim \mathcal{N}\{\mu(M_t), \Sigma(M_t)\}$, where $\mu(M_t) = \{\mu_D(M_t), 0, M_{1,t}, \cdots, M_{\overline{k},t}\}^{\mathrm{T}}$ and

$$\Sigma(M_t) = \begin{pmatrix} \sigma_D^2(M_t) & \rho_{1,2}\sigma_D(M_t) & \\ \rho_{1,2}\sigma_D(M_t) & 1 & \\ & & \sigma_\delta^2 I_{\overline{k}} \end{pmatrix}.$$

The first component of the signal, $s_{1,t}$, corresponds to log dividend growth; it has stochastic volatility $\sigma_D(M_t) = \overline{\sigma}_D \left(\prod_{k=1}^{\overline{k}} M_{k,t} \right)^{1/2}$ and drift $\mu_D(M_t) = g_D - \sigma_D^2(M_t)/2$.

Table 1: Dimension of the state space

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

The second component of the signal drives consumption. The remaining components provide signals about the components of nature's vector. The noise parameter σ_δ controls the quality of the information received by the agent and is the main parameter of interest. Numerical values of g_D, $\overline{\sigma}_D$ and $\rho_{1,2}$, are given in the Appendix.

The agent is fully informed about nature's vector M_t if $\sigma_\delta = 0$. Otherwise, she uses Bayes' rule to infer the probability distribution of M_t conditional on current and past signals:

$$\Pi_t^j = \mathrm{pr}(M_t = m^j \mid s_1, \dots, s_t) \propto n_j(s_t) \sum_{i=1}^{d} a_{ij} \Pi_{t-1}^i \,,$$

for all $j = 1, \dots, d$, where n_j denotes the probability density function of $\mathcal{N}\{\mu(m^j), \Sigma(m^j)\}$. We will refer to $\Pi_t = (\Pi_t^j)_{j=1}^d$ as the *belief vector*. By construction, Π_t belongs to the $(d-1)$-dimensional unit simplex $\Delta_+^{d-1} = \{\Pi \in \mathbb{R}_+^d \mid \sum_{j=1}^d \Pi_j = 1\}$.

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

$$x_t = (M_t, \Pi_t) \in \{m^1, \dots, m^d\} \times \Delta_+^{d-1}, \tag{4.2}$$

where M_t is nature's vector and Π_t is the agent's belief. The state space is therefore $\mathcal{X} = \{m^1, \dots, m^d\} \times \Delta_+^{d-1}$. In the special case where $\sigma_\delta = 0$, the agent is fully informed about the state of nature, the belief vector is a vertex of the unit simplex, and the state space reduces to the finite set $\mathcal{X} = \{m^1, \dots, m^d\}$. The topological dimensions of the state spaces under full and incomplete information are summarized in Table 1.

Observation model. The statistician observes the return on the stock computed by the agent:

$$y_t = \log\left(\frac{1 + q^{\mathrm{T}}\Pi_t}{q^{\mathrm{T}}\Pi_{t-1}}\right) + s_{1,t}\,, \tag{4.3}$$

where $q \in \mathbb{R}^d_{++}$ is a constant vector defined in the appendix. The scalar product $q^{\mathrm{T}}\Pi_t$ is the stock's price dividend ratio at date t. Let $\mathcal{L}(\sigma_\delta \mid y_{1:T})$ denote the loglikelihood of the model. Under full information ($\sigma_\delta = 0$), the likelihood is available analytically (Calvet & Fisher, 2007). When instead information is incomplete ($\sigma_\delta > 0$), the observation density $f_Y(y_t \mid M_t, \Pi_t, y_{1:t-1})$ is unavailable and the computation of the likelihood can be implemented by ABC .

4.2 Likelihood estimation

ABC *likelihood estimates.* The top panel of Figure 2 illustrates a simulated sample of returns $y_{1:T}$ of size $T = 1,000$ periods generated from the structural model with $\overline{k} = 3$ volatility components and signal noise parameter $\sigma_\delta = 0$. The loglikelihood function $\mathcal{L}(0 \mid y_{1:T})$ is available analytically and equal to 3895·99. To illustrate the good accuracy of ABC filters we use in this section the quasi-Cauchy kernel presented in section 3.2 with the plug-in bandwidth (3.6). The plug-in bandwidth $h_t^*(N)$ series for a filter with $N = 10^5$ particles is drawn in the bottom panel of Figure 2. The bandwidth automatically increases when large observations occur.

We now investigate the convergence of the filter when the number of particles N increases. The left panel of Figure 3 illustrates boxplots of 100 estimates of the loglikelihood function obtained with various values of the filter size $N = 10^3$, 10^4, 10^5, 10^6. The ABC -estimated loglikelihood increases with the filter size, as one expects from Jensen's inequality and, for $N \geq 10^6$, gets very close to the true loglikelihood represented by the horizontal line.

Defeat of the curse of dimensionality. In the right panel of Figure 3, we use ABC to compute the loglikelihood $\mathcal{L}(\sigma_\delta \mid y_{1:T})$ under incomplete information for σ_δ =0·1 and σ_δ =0·5 and the maintained dataset $y_{1:T}$ for sample sizes varying from $N = 10^3$ to $5\,10^6$. The ABC filter is now applied to a structural model with state

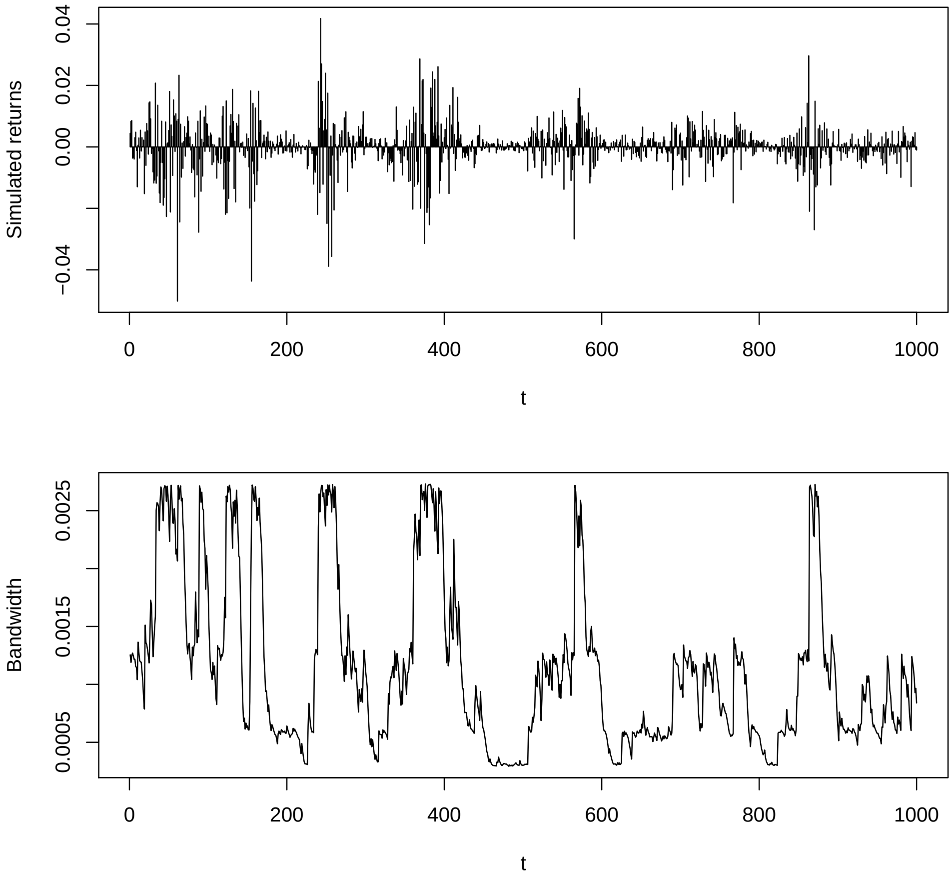

Figure 2: This figure illustrates a simulated series of 1000 daily returns (top panel) generated under the multifractal learning model with $\sigma_\delta = 0$ and bandwidth (bottom panel) of the ABC filter with quasi-Cauchy kernel and optimal bandwidth and $N = 10^5$ particles.

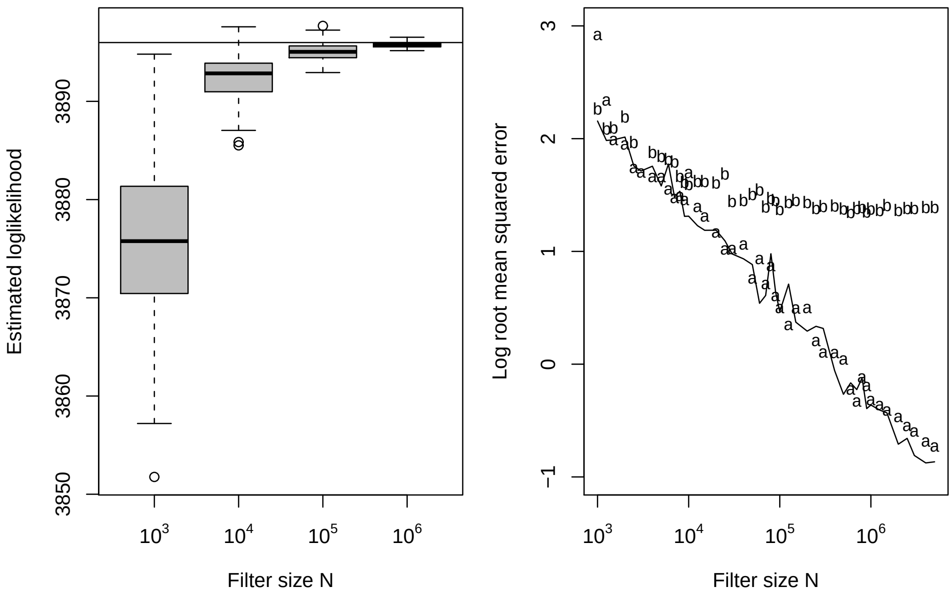

Figure 3: Likelihood estimation via the accurate ABC filter with the quasi-Cauchy kernel and plug-in bandwidth. The left panel illustrates boxplots of the full-information loglikelihood estimates for various filter sizes. Each boxplot is based on 100 estimates and the horizontal line illustrates the true loglikelihood. The right panel illustrates the log root mean squared error of the ABC filter obtained with a correctly specified model ($\sigma_\delta = 0$, continuous line) and misspecified models with a higher state space dimension: (a) σ_δ =0·1; and (b) 0·5. All filters are applied to the dataset reported in Figure 2.

space dimension $n_X = 7$. We compare the convergence of the ABC estimation with increased state space dimension to the ABC estimation using the correctly specified full-information model (σ_δ) where $n_X = 0$. To illustrate convergence results of Theorem 1, we use the root mean squared error:

$$RMSE = \left[\frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{1}{N} \sum_{n=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t \mid y_{1:t-1}) \right\}^2 \right]^{1/2}, \tag{4.4}$$

where $f_Y(y_t \mid y_{1:t-1})$ is the analytically available conditional density function of the full-information model. Right panel of Figure 3 plots the log root mean squared error in (4.4) as a function of the filter size N for ABC with correctly specified full-information model and misspecified learning models with increased state space dimension. Convergence of ABC is nearly identical for $\sigma_\delta = 0$ and $\sigma_\delta = 0{\cdot}1$, even though the state space is much larger under incomplete information than under full information. These findings confirm the result of Theorem 1 that the convergence rate of ABC is independent of the dimension of the state space. Moreover, a least squares regression fit of the log root mean squared error (under the true model) on log sample size results in a slope parameter of $-0{\cdot}365$ which is close to the convergence rate of the root mean squared error $-0{\cdot}4$ in Theorem 1.

4.3 Accurate model selection

ABC methods have been recently criticized for their poor performance in model selection (Marin et al., 2012). We now illustrate the inaccuracy in selecting the correct model when using standard ABC methods, but show that using a kernel and bandwidth satisfying assumptions 1 and 3 ABC is very powerful in selecting the correct model.

ABC importance weights (1.2) are usually defined with the uniform kernel and with either a fixed ε or a quantile-based adaptive ε_t. We now estimate the full information likelihood $\mathcal{L}(0 \mid y_{1:T})$ with $\overline{k} = 3$ and parameter values as in the previous section for the simulated sample in Figure 2 using the two benchmark uniform

ABC filters. The left panel of Figure 4 illustrates the log root mean squared error (4.4) for the estimated $\mathcal{L}(0 \mid y_{1:T})$ with a fixed ε ABC approach for $N = 10^5$ and various $\varepsilon \geq$0·00011. Below this threshold, the fixed-ε ABC encounters weight degeneracy problems. To avoid weight-degenaracy problems, Del Moral et al. (2011) and Jasra et al. (2011) use adaptive ε_t selected to make sure that the proportion of alive particles is equal to a given percentage $0 < \alpha < 1$:

$$\sum_{n=1}^{N} \mathbb{1}_{\|\tilde{y}_t^{(n)} - y_t\| < \varepsilon_t} = \alpha N. \tag{4.5}$$

The proportion α is typically chosen to be large ($\alpha \geq 0{\cdot}5$), to obtain reliable approximations. Right panel of Figure 4 illustrates the log RMSE for the estimated $\mathcal{L}(0 \mid y_{1:T})$ with an adaptive $\varepsilon_t(\alpha)$ for various $0 < \alpha < 1$ and $N = 10^5$. The log RMSE of the quasi-Cauchy kernel-based ABC with plug-in bandwidth $h_t^*(N)$ (value corresponding to $N = 10^5$ in Figure 3) is reported by a dashed line in both panels of Figure 4, which shows that our method outperforms the accuracy of the estimated likelihood of both benchmark methods for all choices of ε and α.

We now turn to model selection. We consider 100 samples simulated from an incomplete information model with $\overline{k} = 3$ components and $\sigma_\delta = 1$. We then estimate the loglikelihood $\mathcal{L}(1 \mid y_{1:T})$ for $\overline{k} = 1, \ldots, 5$ using the quasi-Cauchy kernel-based ABC with plug-in bandwidth $h_t^*(N)$, and adaptive ABC benchmarks with $\alpha =$0·1, 0·5, 0·9. ABC with a fixed bandwidth encounters weigth degeneracy problems for any reasonably small ε when $\overline{k} = 1$ and we do not pursue it here. For the other ABC methods, Table 2 reports the number of times each specification has the highest loglikelihood estimate. Note we are comparing state spaces of respective dimensions 1, 3, 7, 15, and 31. With the recommended values $\alpha =$0·5 and 0·9, the adaptive ABC benchmark consistently selects the wrong model $\overline{k} = 5$, i.e. the model with the highest state space dimension. With $\alpha =$0·1, the correct model is selected 49% of the cases. On the other hand, ABC with quasi-Cauchy kernel and plug-in bandwidth $h_t^*(N)$ selects the correct model in all 100 cases.

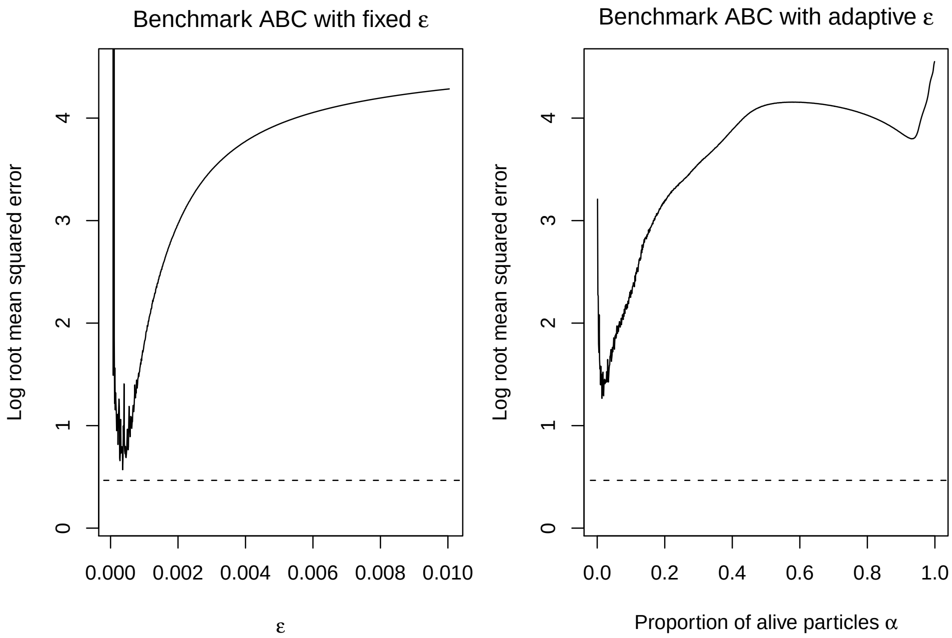

Figure 4: Benchmark ABC methods with $N = 10^5$ particles. The left panel illustrates how the log root mean squared error of a fixed-ε ABC filter varies with the tolerance level ε. The right panel illustrates the precision of an adaptive-ε ABC filter as function of the quantile α. The horizontal dashed line represents the log root mean squared error of the accurate ABC filter with the quasi-Cauchy kernel and plug-in bandwidth. All the filters are applied to the dataset reported in Figure 2.

Table 2: Model selection

	Number of times model with $\overline{k}$ components is selected					Proportion of correct model selections
Method	$\overline{k}=1$	$\overline{k}=2$	$\overline{k}=3$	$\overline{k}=4$	$\overline{k}=5$	(in %)
Quasi-Cauchy kernel, plug-in bandwidth	0	0	100	0	0	100%
Benchmark ABC with adaptive ε_t, α =0·10	0	9	49	1	41	49%
Benchmark ABC with adaptive ε_t, α =0·50	0	0	0	0	100	0%
Benchmark ABC with adaptive ε_t, α =0·90	0	0	0	0	100	0%

A Appendix

A.1 Proofs

Proof of Theorem 1. Since the path $y_{1:T}$ 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 1 *Assume there exists $U_{t-1}(N)$ such that for any bounded measurable function $\Phi : \mathcal{X} \to \mathbb{R}$,*

$$E\Big(\Big[\frac{1}{N}\sum_{n=1}^{N}\Phi(x_{t-1}^{(n)}) - E\{\Phi(x_{t-1}) \mid y_{1:t-1}\}\Big]^2\Big) \leq U_{t-1}(N)\|\Phi\|^2. \tag{A.1}$$

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

$$E\Big(\Big[\frac{1}{N}\sum_{n=1}^{N}\Phi(\tilde{x}_t^{(n)})K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t \mid y_{1:t-1})E\{\Phi(x_t) \mid y_{1:t}\}\Big]^2\Big) \leq U_t^*(N)\|\Phi\|^2\,.$$

Proof of Lemma 1. Consider $a_{t-1}(x_{t-1}) = \int \Phi(\tilde{x}_t)K_{h_t}(y_t - \tilde{y}_t)g(d\tilde{x}_t, d\tilde{y}_t \mid x_{t-1}, y_{1:t-1})$. The function a_{t-1} is bounded above by $\kappa_t\,\|\Phi\|$ since

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

The difference $Z = N^{-1}\sum_{n=1}^{N}\Phi(\tilde{x}_t^{(n)})K_{h_t}(y_t - \tilde{y}_t^{(n)}) - f_Y(y_t \mid y_{1:t-1})E\{\Phi(x_t) \mid y_{1:t}\}$ is the sum of the following three terms:

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

Let $\mathcal{X}_{t-1}^{(N)} = (x_{t-1}^{(1)}, \dots, x_{t-1}^{(N)})$ 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:

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

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 $g(\cdot \mid x_{t-1}^{(n)}, y_{1:t-1})$; 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:

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

By applying the variable change $u = (y_t - \tilde{y}_t)/h_t$, the integral $\int K_{h_t}(y_t - \tilde{y}_t)^2 d\tilde{y}_t = B(K)/(h_t^{n_Y})$. We infer that $E(Z_1^2) \leq \|\Phi\|^2 B(K)\kappa_t/(N h_t^{n_Y})$. Since the function $a_{t-1}(x_{t-1})$ is bounded above by $\kappa_t\,\|\Phi\|$, we infer from (A.1) that: $E(Z_2^2) \leq U_{t-1}(N)\kappa_t^2\,\|\Phi\|^2$.

Finally, we observe that $f_Y(y_t \mid y_{1:t-1})E\{\Phi(x_t) \mid y_{1:t}\} = \int \Phi(x_t) f_Y(y_t \mid x_t, y_{1:t-1})\lambda(dx_t \mid y_{1:t-1})$, and therefore

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

Since $\left|\int K(u)\{f_Y(y_t - h_t u \mid x_t, y_{1:t-1}) - f_Y(y_t \mid x_t, y_{1:t-1})\}du\right| \leq \kappa_t' A(K) h_t^2$, we have

$|Z_3| \le \kappa_t' A(K) h_t^2 \|\Phi\|$ and therefore $E(Z_3^2) \le \kappa_t'^2 A(K)^2 h_t^4 \|\Phi\|^2$. We conclude that the lemma holds. $\square$

Given this result, the proof of (3.2) 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 (3.2) holds with $U_0(N) = 1/N$.

We now assume that the property (3.2) holds at date $t-1$. The estimation error $X = N^{-1} \sum_{n=1}^{N} \Phi(x_t^{(n)}) - E\{\Phi(x_t) \mid y_{1:t}\}$ is the sum of:

$$
\begin{aligned}
X_1 &= \frac{1}{N} \sum_{n=1}^{N} \Phi(x_t^{(n)}) - \sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{x}_t^{(n)}). \\
X_2 &= \Big\{ \sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{x}_t^{(n)}) \Big\} \Big\{ \frac{f_Y(y_t \mid y_{1:t-1}) - N^{-1} \sum_{n'=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n')})}{f_Y(y_t \mid y_{1:t-1})} \Big\}, \\
X_3 &= \frac{1}{N f_Y(y_t \mid y_{1:t-1})} \sum_{n=1}^{N} \Phi(\tilde{x}_t^{(n)}) K_{h_t}(y_t - \tilde{y}_t^{(n)}) - E\{\Phi(x_t) \mid y_{1:t}\}.
\end{aligned}
$$

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 $E[X_1^2 \mid \{(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})\}_{n=1}^{N}] \le \|\Phi\|^2/N$, and therefore $E(X_1^2) \le \|\Phi\|^2/N$. When we use stratified, residual or combined stratified-residual resampling in step 3, the inequality $E(X_1^2) \le \|\Phi\|^2/N$ remains valid, and smaller upper bounds can also be derived (see chapter 7 in Cappé et al. (2005) for a detailed discussion of sampling variance.)

Conditional on $\{(\tilde{x}_t^{(n)}, \tilde{y}_t^{(n)})\}_{n=1}^{N}$, X_2 and X_3 are deterministic variables. Also,

$$
E(X^2) = E(X_1^2) + E\{(X_2 + X_3)^2\} \le E(X_1^2) + 2E(X_2^2) + 2E(X_3^2).
$$

We note that $|X_2| \le \|\Phi\| f_Y(y_t \mid y_{1:t-1})^{-1} |f_Y(y_t \mid y_{1:t-1}) - \sum_{n'=1}^{N} K_{h_t}(y_t - \tilde{y}_t^{(n')})/N|$. Using the induction hypothesis at date $t-1$, we apply Lemma A1 with $\Phi \equiv 1$

and obtain that $E(X_2^2)$ is bounded above by $U_t^*(N)\|\Phi\|^2 f_Y(y_t \mid y_{1:t-1})^{-2}$. Lemma A1 implies that $E(X_3^2)$ is also bounded above by $U_t^*(N)\|\Phi\|^2 f_Y(y_t \mid y_{1:t-1})^{-2}$. We conclude that $E(X^2) \leq U_t(N)\|\Phi\|^2$, where $U_t(N) = 4U_t^*(N) f_Y(y_t \mid y_{1:t-1})^{-2} + N^{-1}$, or equivalently

$$U_t(N) = 4f_Y(y_t \mid y_{1:t-1})^{-2}\Big\{2\kappa_t'^2 A(K)^2 h_t^4 + B(K)\kappa_t/(N h_t^{n_Y}) + 2U_{t-1}(N)\kappa_t^2\Big\} + N^{-1}. \tag{A.2}$$

This establishes part (3.2) of the theorem. From (3.2) and Lemma A1 with $\Phi \equiv 1$, (3.1) 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.2) is minimal if $h_t = N^{-1/(n_Y+4)}\{\kappa_t n_Y B(K)/(8\kappa_t'^2 A(K)^2)\}^{1/(n_Y+4)}$. More generally, if the bandwidth sequence is of the form $h_t(N) = h_t(1)/N^{-1/(n_Y+4)}$, then

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

where $u_{1,t} = 4f(y_t \mid y_{1:t-1})^{-2}\{2\kappa_t'^2 h_t(1)^4 A(K)^2 + B(K)\kappa_t h_t(1)^{-n_Y}\}$ and $u_{2,t} = 8\kappa_t^2 f(y_t \mid y_{1:t-1})^{-2}$. By a simple recursion, $U_t(N)$ is of order $N^{-4/(n_Y+4)}$ for all t. □

Proof of Proposition 1. We first investigate the bias.

$$\begin{aligned} E\{K_{h_t}(y_t - \tilde{y}_t^{(n)}) \mid x_{t-1}^{(n)}, y_{1:t-1}\} &= \int K_{h_t}(y_t - \tilde{y}_t) f_Y(\tilde{y}_t \mid x_{t-1}^{(n)}, y_{1:t-1}) d\tilde{y}_t \\ &= \int K(u) f_Y(y_t - h_t u \mid x_{t-1}^{(n)}, y_{1:t-1}) du. \end{aligned} \tag{A.3}$$

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

$$E\{K_{h_t}(y_t - \tilde{y}_t^{(n)}) \mid x_{t-1}^{(n)}, y_{1:t-1}\} = f_Y(y_t \mid x_{t-1}^{(n)}, y_{1:t-1}) + \frac{h_t^2}{2}\int K(u) u^{\mathrm{T}} \frac{\partial^2 f_Y}{\partial y_t \partial y_t^{\mathrm{T}}}(y_t \mid x_{t-1}^{(n)}, y_{1:t-1}) u \, du + \mathcal{O}(h_t^3)\,.$$

By construction, $E_K(u) = \int uK(u)du = 0$. We observe that

$$\int K(u)u^{\mathrm{T}} \frac{\partial^2 f_Y}{\partial y_t \partial y_t^{\mathrm{T}}}(y_t \mid x_{t-1}^{(n)}, y_{1:t-1}) u \, du = \mathrm{tr}\Big\{ \frac{\partial^2 f_Y}{\partial y_t \partial y_t^{\mathrm{T}}}(y_t \mid x_{t-1}^{(n)}, y_{1:t-1}) \mathrm{var}_K(u) \Big\}.$$

The bias is therefore

$$E(\hat{f}_t \mid \mathcal{X}_{t-1}^{(N)}, y_{1:t-1}) - f_t = \frac{h_t^2}{2} \mathrm{tr}\{H_t(f) \mathrm{var}_K(u)\} + \mathcal{O}(h_t^3)\,.$$

The variance $\mathrm{var}\{K_{h_t}(y_t - \tilde{y}_t^{(n)}) \mid x_{t-1}^{(n)}, y_{1:t-1}\}$ is equal to

$$\begin{aligned}
\int [K_{h_t}(y_t - \tilde{y}_t)]^2 f_Y(\tilde{y}_t \mid x_{t-1}^{(n)}, y_{1:t-1}) d\tilde{y}_t - \Big\{ E\Big[K_{h_t}(y_t - \tilde{y}_t^{(n)}) \mid x_{t-1}^{(n)}, y_{1:t-1} \Big] \Big\}^2 = \\
= \frac{1}{h_t^{n_Y}} \int [K(u)]^2 f_Y(y_t - h_t u \mid x_{t-1}^{(n)}, y_{1:t-1}) du + \mathcal{O}(1) \\
= \frac{B(K)}{h_t^{n_Y}} f_Y(y_t \mid x_{t-1}^{(n)}, y_{1:t-1}) + \mathcal{O}(h_t^{-n_Y+1})\,.
\end{aligned}$$

We conclude that (3.4) holds. □

A.2 Simulation parameters

We choose the parameter values used in Calvet & Fisher (2007): m_0 =1·7, $\gamma_{\bar{k}}$ =0·06, $b = 2$, $g_D = 5\,10^{-5}$, $\overline{\sigma}_D$ =0·70%, and $\rho_{1,2}$ =0·6. The vector q is given by $q^{\mathrm{T}} = (I_d - B)^{-1}\iota - \iota$, where $B = (b_{ij})_{1 \leq i,j \leq d}$ is the matrix with components $b_{ij} = a_{i,j} \exp\{g_D - \alpha\, \rho_{1,2}\, \sigma_C\, \sigma_D(m^j)\}$ and $\iota = (1, \ldots, 1)^{\mathrm{T}}$. We set σ_C =0·189% and choose the risk aversion coefficient α so that the average price-dividend ratio is $d^{-1} \sum_{i=1}^d q^i = 6000$ in daily units.

References

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

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

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

Cappé, O., Moulines, E. & Rydén, T. (2005). *Inference in Hidden Markov Models*. New York: Springer-Verlag.

Chopin, N. (2004). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Annals of Statistics* **32**, 2385–2411.

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

Del Moral, P., Doucet, A. & Jasra, A. (2011). An Adaptative sequential Monte Carlo method for approximate bayesian computation. *Statistics and Computing* , forthcoming.

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

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

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

Hall, P., Lahiri, S. N. & Truong, Y. K. (1995). On Bandwidth Choice for Density Estimation with Dependent Data. *The Annals of Statistics* **23**, 2241–2263.

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

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

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

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

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

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

Marin, J.-M., Pudlo, P., Robert, C. P. & Ryder, R. (2012). Approximate Bayesian Computation methods. *Statistics and Computing* , forthcoming.

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

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

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

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