On the conjugacy of off-line and on-line
Sequential Monte Carlo Samplers

Working Paper Research

by Arnaud Dufays

September 2014    No 263

National Bank
OF BELGIUM

Eurosystem

**Statement of purpose:**

The purpose of these working papers is to promote the circulation of research results (Research Series) and analytical studies (Documents Series) made within the National Bank of Belgium or presented by external economists in seminars, conferences and conventions organised by the Bank. The aim is therefore to provide a platform for discussion. The opinions expressed are strictly those of the authors and do not necessarily reflect the views of the National Bank of Belgium.

**Orders**

For orders and information on subscriptions and reductions: National Bank of Belgium, Documentation - Publications service, boulevard de Berlaimont 14, 1000 Brussels

Tel +32 2 221 20 33 - Fax +32 2 21 30 42

The Working Papers are available on the website of the Bank: http://www.nbb.be

## Abstract

Sequential Monte Carlo (SMC) methods are widely used for filtering purposes of non-linear economic or financial models. Nevertheless the SMC scope encompasses wider applications such as estimating static model parameters so much that it is becoming a serious alternative to Markov-Chain Monte-Carlo (MCMC) methods. Not only SMC algorithms draw posterior distributions of static or dynamic parameters but additionally provide an estimate of the normalizing constant. The tempered and time (TNT) algorithm, developed in the paper, combines (off-line) tempered SMC inference with on-line SMC inference for estimating many slightly different distributions. The method encompasses the Iterated Batch Importance Sampling (IBIS) algorithm and more generally the Re-sample Move (RM) algorithm. Besides the number of particles, the TNT algorithm self-adjusts its calibrated parameters and relies on a new MCMC kernel that allows for particle interactions. The algorithm is well suited for efficiently back-testing models. We conclude by comparing in-sample and out-of-sample performances of complex volatility models.

**Corresponding author**:

Arnaud Dufays,   École Nationale de la Statistique et de l'Administration Économique, CREST, 15, Boulevard Gabriel Péri, 92245 Malakoff, France.
email: arnaud.dufays@ensae.fr.
Website: https://sites.google.com/site/websiteofarnauddufays/

**TABLE OF CONTENTS**

# 1    Introduction

Sequential Monte Carlo (SMC) algorithm is a simulation-based procedure used in Bayesian framework for drawing distributions. Its core idea relies on an iterated application of the importance sampling technique to a sequence of distributions converging to the distribution of interest[1]. For many years, on-line inference was the most relevant applications of SMC algorithms. Indeed, one powerful advantage of sequential filtering consists in being able to update the distributions of the model parameters in light of new coming data (hence the term on-line) allowing for important time saving compared to off-line methods such as the popular Markov-Chain Monte-Carlo (MCMC) procedure that requires a new estimation based on all the data at each new observation entering in the system. Other SMC features making it very interesting are an intuitive implementation based on the importance sampling technique (Geweke (1989), Smith and Gelfand (1992) and Gordon, Salmond, and Smith (1993)) and a direct computation of the marginal likelihood (i.e. the normalizing constant of the distribution, see e.g. Chib, Nardari, and Shephard (2002)).

Recently, the SMC algorithms have been applied to inference of static parameters (field in which the MCMC algorithm excels). Neal (1998) provides a relevant improvement in this direction by building a SMC algorithm, named annealed importance sampling (AIS), that sequentially evolves from the prior distribution to the posterior distribution using a tempered function, which basically consists in gradually introducing the likelihood information into the sequence of distributions by means of an increasing function. To preclude particle degeneracies, he uses MCMC kernels at each SMC iteration. Few years later, Chopin (2002) proposes an Iterated Batch Importance Sampling (IBIS) SMC algorithm, a particular case of the Re-sample Move (RM) algorithm of Gilks and Berzuini (2001), which sequentially evolves over time and adapts the posterior distribution using the previous approximate distribution. Again, an MCMC move (and a re-sampling step) is used for diversifying the particles. The SMC sampler (see Del Moral, Doucet, and Jasra (2006)) unifies, among others, these SMC algorithms in a theoretical framework. It is shown that the methods of Neal (1998) and Gilks and Berzuini (2001) arise as special cases with a specific choice of the 'backward kernel function' introduced in their paper. These researches have been followed by empirical works (see Jasra, Stephens, and Holmes (2007),Jasra, Stephens, Doucet, and Tsagaris (2011) and

---

[1]whereas the sequence does not have to evolve over the time domain

Jeremiah, Sisson, Marshall, Mehrotra, and Sharma (2011)) where it is demonstrated that the SMC mixing properties often dominate MCMC methods based on a single Markov-chain.

A relevant drawback of the SMC samplers lies in the many user-defined parameters which can be difficult to tune for non-experts. Jasra, Stephens, Doucet, and Tsagaris (2011) develop an adaptive SMC sampler that exploits the information of the history of the samples to automatically compute the user-defined parameters on the fly. However the asymptotic properties of such algorithms were not understood until recently. Del Moral, Doucet, and Jasra (2012) and Beskos, Jasra, and Thiery (2013) have done many progress in this area, providing a theoretical justification for a class of adaptive SMC methods[2]. Nowadays papers are devoted to build self-adapting SMC samplers by automatically tuning MCMC kernels (e.g. Fearnhead and Taylor (2013)), by marginalizing the state vector (in a state space specification) using the particle MCMC framework (e.g. Fulop and Li (2013) and Chopin, Jacob, and Papaspiliopoulos (2013)), to construct efficient SMC samplers for parallel computations (see Durham and Geweke (2012)) or to simulate from complex multi-modal posterior distributions (e.g. Herbst and Schorfheide (2012)).

In this paper, we are interested in Bayesian settings where many estimations of slightly different parameter distributions are required such as for testing the forecast ability of a model. For example, in a model comparison context the main methodology consists in repeating estimations of the parameters given an evolving number of observations. In circumstances where the Bayesian parameter estimation is highly demanding as it is usually the case for complex models and where the number of available observations is huge, this iterative methodology can be too burdensome. For instance, Markov-Switching (MS) GARCH models, used in the empirical applications, may require several hours for one MCMC estimation (e.g. Bauwens, Dufays, and Rombouts (2013)). Recursive forecast exercise on many observations is therefore out of reach. In order to easily compare models, we provide a generic SMC algorithm that reduces the time requirement compared to MCMC methods by using previous posterior parameter distributions in order to estimate the next one, that automatically tunes the tempered function exhibited in the AIS algorithm and that limits the particle degeneration

---

[2]especially covering the adaptive Metropolis MCMC kernel and the adaptive tempered function of Jasra, Stephens, Doucet, and Tsagaris (2011).

observed in SMC algorithms evolving through time. The proposed tempered and time (TNT) algorithm, based on the SMC sampler, exhibits the AIS, the IBIS and the RM samplers as particular cases. It innovates by switching over tempered and time domains for estimating posterior distributions. For example, it firstly iterates from the prior to the posterior distributions by means of a sequence of tempered posterior distributions. It then updates in the time dimension the slightly different posterior distributions by sequentially adding new observations, each SMC step providing all the forecast summary statistics relevant for comparing models. The TNT algorithm combines the tempered approach of Neal (1998) with the IBIS algorithm (IBIS) of Chopin (2002) if the model parameters are static or with the RM method of Gilks and Berzuini (2001) if their support evolves with the SMC updates. Since all these methods are built on the same SMC steps (re-weighting, re-sampling and rejuvenating) and the same SMC theory, the combination is achieved without efforts.

The proposed methodology exhibits several advantages over numerical alternatives :

- *SMC algorithms that directly iterate on time domain (Gilks and Berzuini (2001) and Chopin (2002))* sometimes exhibit high particle discrepancies. Although the problem is more acute for model where the parameter space evolves through time, it remains an issue for models with static parameter at the very first SMC steps. To quote Chopin (2002) (p 546) :

  > Note that the particle system may degenerate strongly in the very early stages, when the evolving target distribution changes the most[...].

  The combination of off-line/on-line SMC algorithms allows for limiting this particle discrepancies observed at the early stage since the first estimated posterior distribution is achieved by off-line method taking into account more than a few observations. One advantage of using a sequence of tempered distributions to converge to the posterior distribution consists in the number of SMC steps that can be used. Compared to SMC algorithms that directly iterate on time domain where the sequence of distributions is obviously defined by the number of data, the tempered approach allows for choosing this sequence of distribution and for targeting the posterior distribution of interest by using as much bridging distributions as needed. Moreover, thanks to the unified SMC sampler theory, the proposed SMC algorithm preserves the appealing 'black box' struc-

ture of the IBIS one.

- *Compared to MCMC approaches*, after only one estimation, all the relevant quantities for testing in-sample and out-of-sample performances of a model are computed (including the marginal likelihood and the predictive marginal likelihood). Many estimators have been proposed for computing the marginal likelihood from MCMC outputs (e.g. Chib (1995), Chib and Jeliazkov (2001),Meng and Wong (1996)) but they all require substantial coding efforts and are model-dependent which is not the case for SMC algorithms. Lastly, the SMC algorithm is easily parallelized, a difficult feature to achieve in MCMC frameworks.

A relevant drawback of SMC algorithms lies in the number of user-defined parameters to be tuned for optimizing the estimation procedure. The paper comments on each parameter and proposes generic solutions for all of them. Among the choices, a special attention is paid to the MCMC kernel, which is the most critical one. We innovate by introducing into the SMC technique the generic Metropolis algorithm called DREAM (Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009)) and its discrete extension, namely the D-DREAM (see Bauwens, Dufays, and De Backer (2011)). Moreover, relying on the method of Atchadé and Rosenthal (2005), we improve the DREAM proposal distributions by adapting the size of the jump from one SMC iteration to another. By doing so, we also solve the issue of the number of MCMC moves at each SMC step.

As empirical exercises, we study stationary and non-stationary volatility models and look at their forecast abilities on the S&P 500 daily percentage returns. We emphasize that change-point (CP-) GARCH models are easily inferred using the TNT samplers. Furthermore, we document a Bayesian estimation of the (Discrete) Markov-switching Multifractal (MSM) model (see Calvet and Fisher (2004)). To our knowledge, MSM models are uniquely estimated using the maximum likelihood technique or the generalized method of moments (see Lux (2006)). While the former is limited by the number of latent factors involved in the model, the latter does not produce accurate estimates. By moving to the Bayesian framework, we rule out these restrictions. The application compares the GARCH (Bollerslev (1986)), the CP-GARCH process with the MSM model.

4

The paper is organized as follows. Section 2 presents the SMC algorithm as well as its theoretical derivation. We then show on simulations the desirable properties of the proposed method compared to SMC alternatives in section 3. Finally we study the MSM forecast performances compared to GARCH and CP-GARCH models in section 4. Section 5 concludes.

## 2 Off-line and On-line inferences

We first theoretically and practically introduce the tempered and time (TNT) framework. To ease the discussion, let consider a standard state space model:

$$y_t = f(\theta, s_t, \omega_t) \tag{1}$$

$$s_t = g(\theta, s_{t-1}, v_t) \tag{2}$$

where $s_t$ is a random variable driven by a Markov chain and the functions f(-) and g(-) are deterministic given their arguments. The observation $y_t$ belongs to the set $y_{1:T} = \{y_1, ..., y_T\}$ with $T$ denoting the sample size and is assumed to be independent conditional to the state $s_t$ and $\theta$ with distribution $f(y_t|\theta, s_t)$. The innovations $\omega_t$ and $v_t$ are mutually independent and stand for the noise of the observation/state equations. The model parameters included in $\theta$ do not evolve over time (i.e. they are static). Let denote the set of parameters at time $t$ by $x_t = \{\theta, s_{1:t}\}$ defined on the measurable space $E_t$.

We are interested in estimating many posterior distributions starting from $\pi(x_\tau|y_{1:\tau})$, where $\tau << T$, until $T$. The SMC algorithm approximates these posterior distributions with a large (dependent) collection of $M$ weighted random samples $\{W_t^i, x_t^i\}_{i=1}^M$ where $W_t^i > 0$ and $\sum_{i=1}^M W_t^i = 1$ such that as $M \to \infty$, the empirical distribution converges to the posterior distribution of interest, meaning that for any $\pi(x_t|y_{1:t})$-integrable function g : $E_t \to \Re$
:

$$\sum_{i=1}^M W_t^i g(x_t^i) \to \int_{E_t} g(x_t)\pi(x_t|y_{1:t})dx_t \quad \text{almost surely.}$$

The TNT method combines an enhanced Annealed Importance sampling[3] (AIS, see Neal (1998)) with the Re-sample Move (RM) SMC inference of Gilks and Berzuini (2001)[4]. For

---

[3]enhanced in the sense that the AIS incorporates a re-sampling step

[4]The IBIS algorithm being a particular case.

building the TNT algorithm, we rely on the theoretical paper of Del Moral, Doucet, and Jasra (2006) that unifies the two SMC inferences into one SMC framework called 'SMC sampler'. The TNT algorithm first estimates an initial posterior distribution, namely $\pi(x_\tau|y_{1:\tau})$, by an enhanced AIS (E-AIS) algorithm and then switches from the tempered domain to the time domain and sequentially updates the approximated distributions from $\pi(x_\tau|y_{1:\tau})$ to $\pi(x_T|y_{1:T})$ by adding one by one the new observation. As discussed in section 2.3, the other domain change (from time to tempered function) also exhibits some advantages. The TNT algorithm therefore switches over these two domains during the entire inference. We now begin by mathematically deriving the validity of the SMC algorithms under the two different domains and by showing that they are particular cases of the SMC sampler. The practical algorithm steps are given afterward (see subsection 2.3).

## 2.1 E-AIS : the tempered domain

The first phase, carried out by an E-AIS, creates a sequence of probability measures $\{\pi_n\}_{n=0}^p$ that are defined on measurable spaces $\{E_n, \xi_n\}$, where $E_n = E_{n+1} = E \ni x_\tau$, $n \in \{0, 1, ..., p\}$ is a counter and does not refer to 'real time', $p$ denotes the number of posterior distribution estimations and $\pi_p$ coincides with the first posterior distribution of interest $\pi(x_\tau|y_{1:\tau})$. The sequence distribution, used as bridge distributions, is defined as $\pi_n(x_n|y_{1:\tau}) = \gamma(y_{1:\tau}|x_n)^{\phi(n)} f(x_n)/Z_n$ where $Z_n = \int_E \gamma(y_{1:\tau}|x_n)^{\phi(n)} f(x_n) dx_n$ denotes the normalizing constant, $\gamma(y_{1:\tau}|x_n)$ and $f(x_n)$ respectively are the likelihood function and the prior density of the model. Through an increasing function $\phi(n)$ respecting the bound conditions $\phi(0) = 0$ and $\phi(p) = 1$, the E-AIS artificially builds a sequence of distributions that converges to the posterior distribution of interest.

_Remark 1:_ The E-AIS makes up a sequence of random variables $\{x_n\}_{n=0}^p$ that exhibit the same support $E$ also shared by $x_\tau$. The random variable $x_\tau$ coincides with $x_p$ since $\phi(p) = 1$.

The E-AIS is merely a sequential importance sampling technique where the draws of a proposal distribution $\eta_n$ with MCMC kernel combinations are used to approximate the next posterior distribution $\pi_{n+1}$, the difficulty lying in specifying the sequential proposal distribution. Del Moral, Doucet, and Jasra (2006) theoretically develop a coherent framework for

helping to choose a generic sequence of proposal distributions.

In the SMC sampler framework, we augment the support of the posterior distribution ensuring that the targeted posterior distribution marginally arises :

$$\tilde{\pi}_n(x_{1:n}) = \pi_n(x_n) \prod_{k=2}^{n} L_k(x_{k-1}|x_k)$$

$$= \frac{\gamma_n(x_n)}{Z_n} \prod_{k=2}^{n} L_k(x_{k-1}|x_k)$$

where $\gamma_n(x_n) = \gamma(y_{1:\tau}|x_n)^{\phi(n)} f(x_n)$, $Z_n = \int_E \gamma(y_{1:T}|x_n)^{\phi(n)} f(x_n) dx_n$ is the normalizing constant, and $L_k(x_{k-1}|x_k)$ is a backward MCMC kernel such that $\int_E L_k(x_{k-1}|x_k) dx_{k-1} = 1$. By defining a sequence of proposal distributions as

$$\eta_n(x_{1:n}) = f(x_1) \prod_{k=2}^{n} K_k(x_k|x_{k-1}),$$

where $K_k(x_k|x_{k-1})$ is an MCMC kernel with stationary distribution $\pi_k$ such that it verifies $\pi_k(x_k) = \int_E K_k(x_k|x_{k-1})\pi_k(x_{k-1}) dx_{k-1}$, we derive a recursive equation of the importance weight :

$$w_n(x_{1:n}) = \frac{\gamma_n(x_n) \prod_{k=2}^{n} L_k(x_{k-1}|x_k)}{Z_n f(x_1) \prod_{k=2}^{n} K_k(x_k|x_{k-1})}$$

$$= w_{n-1}(x_{1:n-1}) \frac{Z_{n-1}\gamma_n(x_n)L_n(x_{n-1}|x_n)}{Z_n\gamma_{n-1}(x_{n-1})K_n(x_n|x_{n-1})}.$$

For a smoothly increasing tempered function, we can argue that $\pi_{n-1}$ will be close to $\pi_n$. We therefore define the backward kernel by detailed balance argument as

$$L_n(x_{n-1}|x_n) = \frac{\pi_n(x_{n-1})}{\pi_n(x_n)} K_n(x_n|x_{n-1}). \tag{3}$$

It gives the following weights :

$$w_n(x_{1:n}) = w_{n-1}(x_{1:n-1}) \frac{Z_{n-1}\gamma_n(x_{n-1})}{Z_n\gamma_{n-1}(x_{n-1})} \tag{4}$$

$$\propto w_{n-1}(x_{1:n-1})\gamma(y_{1:\tau}|x_{n-1})^{\phi_n-\phi_{n-1}}$$

The normalizing constant $Z_n$ is approximated as

$$\frac{Z_n}{Z_{n-1}} \approx \sum_{i=1}^{M} W_{n-1}^i \frac{\gamma_n(x_{n-1})}{\gamma_{n-1}(x_{n-1})}$$

where $W_{n-1}^i = w_{n-1}(x_{1:n-1}^i)/\sum_{j=1}^M w_{n-1}(x_{1:n-1}^j)$, i.e. the normalized weight.

The E-AIS requires to tune many parameters : an increasing function $\phi(n)$, MCMC kernels of invariant distribution $\pi_n(.)$, a number of particles $M$, of iterations $p$, of MCMC steps $J$. Tuning these parameters can be difficult. Some guidance are given in Herbst and Schorfheide (2012) for DSGE models. For example, they propose a quadratic tempered function $\phi(n)$. It slowly increases for small values of $n$ and the step becomes larger and larger as $n$ tends to $p$. In this paper, the TNT algorithm generically adapts the different user-defined parameters and belongs to the class of adaptive SMC algorithms. It automatically adjusts the tempered function with respect to an efficiency measure as it was proposed by Jasra, Stephens, Doucet, and Tsagaris (2011). By doing so, we preclude the difficult choice of the function $\phi(n)$ and the number of iteration $p$. The number of MCMC steps $J$ will be controlled by the acceptance rate exhibited by the MCMC kernels. The choice of MCMC kernels and the number of particles are discussed later (see section 2.5).

## 2.2  The Re-sample Move algorithm : the time domain

Once we have a set of particles that approximates the first posterior distribution of interest $\pi(x_\tau|y_{1:\tau})$, a second phase takes place. Firstly, let assume that the support of $x_t$ does not evolve over time (i.e. $x_t \in E \quad \forall \quad t$). In this context, the SMC sampler framework shortly reviewed here for the tempered domain still applies. Let define the following distributions :

$$
\begin{aligned}
\pi_t(x_t) &= \pi(x_t|y_{1:t}) \\
\tilde{\pi}_t(x_{1:t}) &= \pi_t(x_t) \prod_{k=2}^t L_k(x_{k-1}|x_k) \\
\eta_t(x_{1:t}) &= f(x_1) \prod_{k=2}^t K_k(x_k|x_{k-1}) \\
L_k(x_{k-1}|x_k) &= \frac{\pi_k(x_{k-1})}{\pi_k(x_k)} K_k(x_k|x_{k-1})
\end{aligned}
$$

Then the weight equation of the SMC sampler are equal to :

$$
w_t(x_{1:t}) \quad = \quad w_{t-1}(x_{1:t-1}) \frac{\pi_t(x_{t-1})}{\pi_{t-1}(x_{t-1})} \tag{5}
$$

which is exactly the weight equation of the IBIS algorithm (see Chopin (2002), step 1, page 543).

Let now consider the more difficult case where a subset of the support of $x_t$ evolves with $t$ such as $x_t = \{x_{t-1}, s_t\} = \{\theta, s_{1:t-1}, s_t\}$ (see state space model equations (1),(2)) meaning that $\forall t \in [1, T], x_t \in E_t$ and $E_{t-1} \subset E_t$. The previous method cannot directly be applied (due to the backward kernel) but with another choice of the kernel functions, the SMC sampler also operates. Let define the following distribution :

$$\pi_t(x_t) = \pi(x_t|y_{1:t}) \tag{6}$$

$$\tilde{\pi}_t(x_{1:t}, x^*_{2:t}) = \pi_t(x_t) \prod_{k=2}^{t} L_k(x_{k-1}, x^*_k|x_k) \tag{7}$$

$$\eta_t(x_{1:t}, x^*_{2:t}) = f(x_1) \prod_{k=2}^{t} \tilde{K}_k(x_k, x^*_k|x_{k-1}) \tag{8}$$

$$\tilde{K}_k(x_k, x^*_k|x_{k-1}) = \tilde{q}_k(x^*_k|x_{k-1}) K_k(x_k|x^*_k) \tag{9}$$

$$L_k(x_{k-1}, x^*_k|x_k) = q_k(x_{k-1}|x^*_k) K_k(x^*_k|x_k) \tag{10}$$

$$K_k(x^*_t|x_t) = \frac{\pi_k(x^*_k) K_k(x_k|x^*_k)}{\pi_k(x_k)} \quad \text{by detailed balance argument} \tag{11}$$

To deal with the time-varying dimension of $x_t$, we augment the support of the artificial sequence of distributions by new random variables (see $x^*_{2:t}$ in equation (7)) while ensuring that the posterior distribution of interest $\pi_t(x_t)$ marginally arises. Sampling from the proposal distribution $\eta_t(x_{1:t}, x^*_{2:t})$ is achieved by drawing from the prior distribution and then by sequentially sampling from distributions $\tilde{K}_k(x_k, x^*_k|x_{k-1})$, which is composed by a user-defined distribution $\tilde{q}_k(x^*_k|x_{k-1})$ and an MCMC kernel exhibiting $\pi_k(x_k)$ as invariant distribution. Under this framework, the weight equation of the SMC sampler becomes :

$$w_t(x_{1:t}, x^*_{2:t}) = w_{t-1}(x_{1:t-1}, x^*_{2:t-1}) \frac{\pi_t(x^*_t) q_t(x_{t-1}|x^*_t)}{\pi_{t-1}(x_{t-1}) \tilde{q}_t(x^*_t|x_{t-1})} \tag{12}$$

By setting the distributions $q_t(x_{t-1}|x^*_t) = \delta_{\{x_{t-1}=\theta^*, s_{1:t-1}\}}$, where $\delta_i$ denotes the probability measure concentrated at $i$, and $\tilde{q}_t(x^*_t|x_{t-1}) = \nu_t(s^*_t|x_{t-1}) \delta_{\{\theta^*, s_{1:t-1}=x_{t-1}\}}$, we recover the weight equation of Gilks and Berzuini (2001) (see eq. (20), page 135) :

$$w_t(x_{1:t}, x^*_{2:t}) = w_{t-1}(x_{1:t-1}, x^*_{2:t-1}) \frac{\pi_t(x_{t-1}, s^*_t)}{\pi_{t-1}(x_{t-1}) \nu_t(s^*_t|x_{t-1})} \tag{13}$$

9

Like in Gilks and Berzuini (2001), only the distribution $\nu_t(.)$ has to be specified. For example, it can be set either to the prior distribution or the full conditional posterior distribution (if the latter exhibits a closed form).

<u>Remark 2:</u> The division $\frac{\pi(x_t|y_{1:t})}{\pi(x_{t-1}|y_{1:t-1})}$ appearing in weight equations ((5),(13)) can be reduced to $\frac{Z_{t-1}}{Z_t}f(y_t|x_t,y_{1:t-1})f(s_t|x_{t-1})$ which highly limits the computational cost of the weights.

## 2.3   The TNT algorithm

The algorithm initializes the $M$ particles using the prior distributions, sets each initial weight $\{W_0^i\}_{i=1}^M$ to $W_0^i = \frac{1}{M}$ and then iterates from $n = 1, \ldots, p, p+1, \ldots, p+(T-\tau)+1$ as follows

- **Correction step:** $\forall i \in [1, M]$, Re-weight each particle with respect to the $n$th posterior distribution

  - If in tempered domain ( $n \leq p$ ) :

  $$\tilde{w}_n^i = \gamma(y_{1:\tau}|x_{n-1}^i)^{\phi(n)-\phi(n-1)} \tag{14}$$

  - If in time domain ( $n > p$ ) and the parameter space does not evolve over time (i.e. $E_{n-1} = E_n$) :

  $$\tilde{w}_n^i = \frac{\gamma(y_{1:\tau+n-p}|x_{n-1}^i)}{\gamma(y_{1:\tau+n-p-1}|x_{n-1}^i)} = f(y_{\tau+n-p}|x_{n-1}^i, y_{1:\tau+n-p-1}) \quad \text{(see remark 2)} \tag{15}$$

  - If in time domain ( $n > p$ ) and the parameter space increases (i.e. $E_{n-1} \subset E_n$) :
    Set $x_n^i = \{x_{n-1}^i, s_n^i\}$ with $s_n^i \sim \nu_n(.|x_{n-1}^i)$.

  $$\tilde{w}_n^i = \frac{\gamma(y_{1:\tau+n-p}|x_n^i)f(x_n^i)}{\gamma(y_{1:\tau+n-p-1}|x_{n-1}^i)f(x_{n-1}^i)\nu_n(s_n^i|x_{n-1}^i)} \tag{16}$$

  Compute the unnormalized weights : $\tilde{W}_n^i = \tilde{w}_n^i W_{n-1}^i$.
  Normalize the weights : $W_n^i = \frac{\tilde{W}_n^i}{\sum_{j=1}^M \tilde{W}_n^j}$

- **Re-sampling step:** Compute the Effective Sample Size (ESS) as

$$ESS = \frac{1}{\sum_{i=1}^M (W_n^i)^2}$$

10

If $ESS < \kappa$ where $\kappa$ is a user-defined threshold then re-sample the particles and reset the weight uniformly.

- **Mutation step:** $\forall i \in [1, M]$, run J steps of an MCMC kernel with invariant distribution $\pi_n(x_n|y_{1:\tau})$ for $n \le p$ and $\pi(x_{\tau+n-p}|y_{1:\tau+n-p})$ for $n > p$.

<u>Remark 3:</u> According to the algorithm derivation, note that the mutation step is not required at each SMC iteration.

When the parameter space does not change over time (i.e. tempered or time domains with $E_{n-1} = E_n$), the algorithm reduces to the SMC sampler with a specific choice of the backward kernel (see equation (3), more discussions in Del Moral, Doucet, and Jasra (2006)) that implies that $\pi_{n-1}(.)$ must be close to $\pi_n(.)$ for non-degenerating estimations. The backward kernel is introduced for avoiding the choice of an importance distribution at each iteration of the SMC sampler. This specific choice of backward kernel does not work for model where the parameter space increases with the sequence of posterior distributions (hence the use of a second weighting scheme when $n > p$, see equation (12)) but the algorithm also reduces to a SMC sampler with another backward kernel choice (see (10)). In empirical applications, we first estimate an off-line posterior distribution with fixed parameters and then by just switching the weight equation, we sequentially update the posterior distributions by adding new information. This two phases preclude the particle degeneration that may occur at the early stage of the IBIS algorithm (when the splitting time $\tau >> 0$). The tempered function $\phi(n)$ allows for converging to the first posterior distribution targeted as slowly as we want. Indeed, as we are not constraint by the time domain, we can sequentially iterate as much as needed to get rid of degeneracy problems. The choice of the tempered function $\phi(n)$ is therefore relevant. In the spirit of a black-box algorithm as is the IBIS one, the section 2.4 shows how the TNT algorithm automatically adapts it at each SMC iteration.

Switching from the time to the tempered domains can also be beneficial. During the second phase (i.e. updating the posterior distribution through time), one may observe high particle discrepancies especially when the space of the parameters evolves over time[5]. In that

---

[5]Theorem 1 in Chopin (2002) ensures that with a sufficiently large number of particles $M$, any relative precision of the importance sampling can be obtained if the number of observations already covered is large

case, instead of directly approximate the posterior distribution, we could rather apply the importance weight on a tempered version of the targeted posterior distribution. Then, we update the tempered posterior distribution by weight equations (14) until reaching the posterior distribution of interest. Since all the three weight equations belong to the same SMC sampler framework, one can switch from one domain to the other without any coding effort. In practice, once the algorithm lies in phase two, at each correction step, we maximize the effective sample size with respect to a tempered value $\bar{\phi}$. More precisely, the correction step becomes (assuming for example that the parameter space does not evolve over time) :

1. Find $\bar{\phi} = \text{argmax}_{\bar{\phi}} \frac{1}{\sum_{i=1}^{M}(W_n^i)^2}$ where $W_n^i = \frac{\tilde{W}_n^i}{\sum_{j=1}^{M} \tilde{W}_n^j}$ is the normalized weights and the unnormalized $\tilde{W}_n^i$ depends on $\bar{\phi}$ as

$$\tilde{W}_n^i = W_{n-1}^i \tilde{w}_n^i = \frac{\gamma(y_{1:\tau+n-p}|x_{n-1}^i)^{\bar{\phi}}}{\gamma(y_{1:\tau+n-p-1}|x_{n-1}^i)} \tag{17}$$

2. Compute the normalized weights $\{W_n^i\}_{i=1}^{M}$ under the value of $\bar{\phi}$.

For any $\bar{\phi} < 1$, the sampler approximates a posterior distribution in the tempered domain. The weight equation (17) is compatible with the E-AIS algorithm since the targeted distribution is $\gamma(y_{1:\tau+n-p}|x_n)^{\bar{\phi}} f(x_n)/Z_n$ where $Z_n$ stands for the normalizing constant. However the efficiency gains of this strategy can be small since only the numerator of the weight equation (17) depends on $\bar{\phi}$. In practice, one can run an entire E-AIS on the data $y_{1:\tau+n-p}$ when a degeneracy issue is detected (i.e. the ESS falls below a user-defined value $\kappa_1 < \kappa$) and cannot be undertaken using weight equation (17). The adaption of the tempered function (discussed in the next section) makes the E-AIS very fast since it reduces the number of iteration $p$ at its minimum given the ESS threshold $\kappa$. Controlling for degeneracy issue is therefore automated and a minimal number of effective sample size is ensured at each SMC iteration.

## 2.4 Adaption of the tempered function

Previous works on SMC samplers usually provide a tempered function $\phi(n)$ obtained by empirical trials[6], making these functions model-dependent. Jasra, Stephens, Doucet, and

enough.

[6] a piecewise cooling linear function for Del Moral, Doucet, and Jasra (2006) and a quadratic function for Herbst and Schorfheide (2012).

Tsagaris (2011) propose a generic choice of $\phi(n)$ that only requires a few more codes. The E-AIS correction step (see equation (14)) of iteration $n$ is modified as follows

1. Find $\bar{\phi}_n$ such that

$$\bar{\phi}_n = \operatorname{argmax}_{\bar{\phi}_n} \frac{1}{\sum_{i=1}^M (W_n^i)^2}$$
$$\text{with } \kappa > \frac{1}{\sum_{i=1}^M (W_n^i)^2}$$

Where $\kappa$ is the threshold of the re-sampling step, $W_n^i = \frac{\tilde{W}_n^i}{\sum_{j=1}^M \tilde{W}_n^j}$ is the normalized weights and the unnormalized $\tilde{W}_n^i$ depends on $\bar{\phi}_n$ as

$$\tilde{w}_n^i = \frac{\gamma(y_{1:\tau}|x_{n-1}^i)^{\bar{\phi}_n}}{\gamma(y_{1:\tau}|x_{n-1}^i)^{\bar{\phi}_{n-1}}}$$

2. Compute the normalized weights $\{W_n^i\}_{i=1}^M$ under the value of $\bar{\phi}_n$.

Roughly speaking, we find the value $\bar{\phi}_n$ that maximizes the effective sample size (ESS) criterion while keeping it just below the threshold $\kappa$ of the re-sampling step.

This adaption implies that at each E-AIS iteration, the ESS is always ensured to be close to $\kappa$, a property that was not at all ensured for fixed ex-ante tempered functions. Note also that the algorithm applies the re-sampling step at each iteration and that the number of them are reduced to a minimum with respect to the choice of $\kappa$. Because the tempered function is adapted on the fly using the SMC history, the usual SMC asymptotic results do not apply. Del Moral, Doucet, and Jasra (2012) and Beskos, Jasra, and Thiery (2013) provide asymptotic results by assuming that the adapted tempered function converges to the optimal one (if it exists).

## 2.5 Choice of MCMC kernels

The MCMC kernel is the most computational demanding step of the algorithm and determines the posterior support exploration, making its choice very relevant. Chopin (2002) emphasizes that the IBIS algorithm is designed to be a true 'black box' (i.e. whose the sequential steps are not model-dependent), reducing the task of the practitioner to only supply the likelihood function and the prior densities. For this purpose, a natural choice of MCMC kernel is the

Metropolis-Hastings with an independent proposal distribution whose summary statistics are derived using the particles of the previous SMC step and the weight of the current step. The IBIS algorithm uses an independent normal proposal. It is worth noting that this 'black box' structure is still applicable in this framework that combines SMC iterations on tempered and time domains.

Nevertheless the independent Metropolis-Hastings kernel may perform poorly at the early stage of the algorithm if the parameter posterior distribution is well behaved and at any time otherwise. We rather propose to use an adaptive Metropolis algorithm of random walk type (which can be theoretically justified by Beskos, Jasra, and Thiery (2013)) where the symmetric distribution is characterized by particles of the $n-1$ SMC iteration and a fixed (at step $n$) scale parameter $c_{n-1}$. Well-suited for uni- and multi-modal distributions, the symmetric DiffeRential Evolution Adaptive Metropolis (DREAM) distribution of Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) that is fully adapted has been proven to dominate most of other RW alternatives. Relying on multiple MCMC, the DREAM proposal updates the parameters of one Markov-chain by using a combination of the locations of the other Markov-chains. This method can be transferred to the TNT framework as follows :

1. Propose new parameters $\tilde{x}^i$ for particle $i$ at SMC iteration $n$, according to the rule (omitting subscripts referring to SMC iterations) :

$$\tilde{x}^i = x^i + \gamma(\delta, d)(\sum_{g=1}^{\delta} x^{r_1(g)} - \sum_{h=1}^{\delta} x^{r_2(h)}) + \zeta, \qquad (18)$$

with $\forall g, h = 1, 2, ..., \delta$, $i \neq r_1(g)$, $r_2(h)$; $r_1(.)$ and $r_2(.)$ stand for random integers uniformly distributed on the support $[1, M]_{-i}$ and it is required that $r_1(g) \neq r_2(h)$ when $g = h$; $\zeta \sim N(0, \eta_x^2 I)$; and $d$ is the number of elements in $x$. Since we take advantage of the locations of multiple particles, $\eta_x^2$ is set to a small value and its presence ensures the ergodicity of the MCMC algorithm. In the seminal paper, $\gamma(\delta, d)$ is chosen as $2.38/\sqrt{2\delta d}$ since it constitutes the asymptotic optimal choice for normal posterior distribution as demonstrated in Roberts and Rosenthal (2001).[7] As the posterior distribution is rarely Normal, we prefer adapting it from one SMC iteration to another so as the scale parameter is fixed during the entire MCMC moves of each SMC step. The adapting procedure is detailed below.

---

[7]in the sense of $M \to \infty$

2. Accept the proposed $\tilde{x}^i$ according to a Metropolis ratio since the proposal distribution is symmetric.

In the empirical application, the DREAM procedure is used for continuous as well as for discrete random variables. For the latter case, we rely on the discretized version of the DREAM algorithm developed in Bauwens, Dufays, and De Backer (2011). Only the equation (18) slightly changes to account for discrete parameters. The modification is given by

$$\tilde{x}^i = x^i + \text{round}[\gamma(\delta, d)(\sum_{g=1}^{\delta} x^{r_1(g)} - \sum_{h=1}^{\delta} x^{r_2(h)}) + \zeta], \tag{19}$$

where the round function stands for the nearest integer operator.

Two relevant issues should be discussed. First, the MCMC kernel makes interacting the particles, which rules out the desirable parallel property of the SMC. To keep this advantage, we apply the kernel on subsets of particles instead of on all the particles and we perform paralelization between the subsets. Secondly and more importantly, the SMC theory derived in Section 2 does not allow for particle interaction. Proposition 1 ensures that the TNT sampler also works under a DREAM-type MCMC kernel.

**Proposition 1.** Consider a SMC sampler with a given number of particles $M$ and MCMC kernels allowing for interacting particles via the proposal distribution (18). Then, it yields a standard SMC sampler with particle weights given by equation (4).

**Proof.** See Appendix A.

**Adaption of the scale parameter $\gamma(\delta, d)$**

Since the chosen backward MCMC kernel in the algorithm derivation implies that the consecutive distributions approximated by the TNT sampler are very similar, we can analyze the mixing properties of the previous MCMC kernel to adapt the scale parameter $c_{n-1}$. Atchadé and Rosenthal (2005) present a simple recursive algorithm to fix the value $c_{n-1}$ in order to achieve a specified acceptance rate in an MCMC framework. At the end of the $n-1$ SMC step, we adapt the scale parameter $c_{n-1}$ (initialized at one) as follows :

$$c_{n-1} = p(c_{n-2} + \frac{\alpha_{n-1} - \alpha_{\text{targeted}}}{(n-1)^{0.6}}) \tag{20}$$

where the function $p(.)$ is such that $p(c) = c$ if $c \in [A_0, A_1]$, $p(c) = A_0$ if $c < A_0$ and $p(c) = A_1$ if $c > A_1$, the parameter $\alpha_{n-1}$ stands for the acceptance rate of the MCMC kernel of the $n-1$ SMC step and $\alpha_{\text{targeted}}$ is a user-defined acceptance rate. The function $p(.)$ precludes the degeneracy of the recursive equation and if the optimal scale parameter lies in the compact set $[A_0, A_1]$, the equation will converge to it (in an MCMC context). In the empirical exercises, we fix the rate $\alpha_{\text{targeted}}$ to $\frac{1}{3}$ implying that every three MCMC iterations, all the particles have been approximately rejuvenated. The number of MCMC iteration $J$ is therefore fixed to 3. It is worth emphasizing that the denominator $(n-1)^{0.6}$ has been chosen as proposed in Atchadé and Rosenthal (2005) but its presence, which ensures the ergodicity property in an MCMC context, is not relevant in our SMC framework since at each rejuvenation step, the variance $c_i$ is fixed for the entire MCMC move.

When the parameter space evolves over time, the MCMC kernel can become model dependent since sampling the state vector using a filtering method is often the most efficient technique in terms of mixing. In special cases where the forward-backward algorithm (Rabiner (1989)) or the Kalman filter (Kalman (1960)) operate, the state can be filtered out. By doing so, we come back to the framework with static parameter space. For non linear state space model, recent works of Chopin, Jacob, and Papaspiliopoulos (2013) and Fulop and Li (2013) rely on the particle MCMC framework of Andrieu, Doucet, and Holenstein (2010) for integrating out the state vector. We believe that switching from the tempered domain to the time one as well as employing the DREAM MCMC kernel could even more increase the efficiency of these sophisticated SMC samplers. For example, the particle discrepancies of the early stage inherent to the IBIS algorithm is present in all the empirical simulations of Fulop and Li (2013) whereas with the TNT sampler, we can ensure a minimum ESS bound during the entire procedure.

## 3  Simulations

We carry out two simulations for comparing the performance of the TNT algorithm to the standard on-line SMC algorithms. To this purpose, we consider two volatility models : the standard generalized auto-regressive conditional heteroskedastic (GARCH) model of Bollerslev (1986) (with static parameters) and the Discrete Markov-switching Multifractal (MSM)

model (see Calvet and Fisher (2004)) belonging to the stochastic volatility models (exhibiting dynamic parameters). These two models are non-trivially estimated but remain tractable by maximum likelihood principle to provide comparison with the maximum likelihood estimates. Table 1 reviews the user-defined parameters that are employed in all the next simulations.

| Parameters | TNT algorithm | IBIS and RM SMC |
|---|---|---|
| Threshold ESS $\kappa$ | $\frac{3M}{4}$ | $\frac{3M}{4}$ |
| Sec. threshold ESS $\kappa_1$ | $\frac{M}{2}$ | — |
| Nb. Particles $M$ | $[200, 500, 1.000, 10.000]$ | $[200, 500, 1.000, 10.000]$ |
| Acc. rate $\alpha_{\text{targeted}}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |
| Nb. MCMC $J$ | 3 | 3 |

Table 1: Chosen parameters for the standard sequential SMC and the TNT algorithms.

## 3.1 Static parameters : The GARCH model

Let assume a financial time series $y_{1:T} = \{y_1, ..., y_T\}$. The GARCH model is specified as follows :

$$y_t = \mu + \epsilon_t \quad \text{with} \quad \epsilon_t | y_{1:t-1} \sim \text{i.i.d. } N(0, \sigma_t^2) \tag{21}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{22}$$

where the mean parameter $\mu \in \Re$ and the GARCH parameters $\omega \in \Re^+, \alpha, \beta \in [0, 1]^2$. We impose the standard stationary condition $\alpha + \beta < 1$ through the prior distributions[8] and gather the parameters in the set $\theta = \{\mu, \omega, \alpha, \beta\}$.

One simulated series of 3000 observations from the data generating process (DGP) documented in table 2 will be used to compare the different numerical methods. We perform an exercise by starting at $\tau = 1500$ and by adding one by one the other observations.

In Table 3, we compare the TNT algorithms with the IBIS one. It documents the number of particles, the maximum auto-correlation time (Act., computed by Batch means (see Geyer

---

[8]$\mu \sim N(0, 10), \quad \omega \sim U[0, 1.5], \quad \alpha \sim U[0, 0.3]$ and $\beta | \alpha \sim U[0, 1 - \alpha]$.

| $\mu$ | $\omega$ | $\alpha$ | $\beta$ |
|---|---|---|---|
| 0 | 0.1 | 0.07 | 0.9 |

Table 2: Data generating process of the GARCH model.

(1992))) on the GARCH parameters of the last SMC iteration, the elapsed time for computing the first posterior distribution until the end (from $\tau = 1500$ to 3000), the marginal log-likelihood (MLL) estimators obtained for the entire set of observations and the average of the acceptance rate observed at the end of each SMC iteration.

Regarding the auto-correlation times, all the SMC algorithms exhibit extremely low levels which is one of the relevant advantages of these algorithms. Considering now the elapsed time, the 'E-AIS with IBIS' algorithm (meaning that it does not allow any switch back from time to tempered domain) and the TNT one are clearly faster than the IBIS procedure. This superior performance comes from the tempered domain used for estimating the first posterior distribution. Not only starting by iterating on the tempered domain speeds up the estimation procedure but also limits the particle deterioration since the first posterior distribution is directly targeted using all the available observations. To document this improvement, Figure 1a shows the ESS of the IBIS algorithm from time one until time $\tau$. This Figure should be put in perspective with Figure 1b that clearly indicates no particle discrepancies and less SMC iterations. The TNT algorithm furthermore allows for returning in the tempered domain for keeping high the ESS level from one SMC step to another. The Figure 2 exhibits the value of the tempered function $\phi$ for $M = 1000$ over the entire simulation but after the first E-AIS on $\tau = 1500$. We observe that the algorithm comes back to the tempered domain several times. The cumulative gains in ESS are given in Table 3. Moving from time to tempered domain highly increases the particle diversification. Keeping studying the Table 3, in term of MLL estimations, the algorithms are equivalent. Finally, note that the adapting equation (20) for obtaining a chosen acceptance rate performs very well. The average acceptance rate of each SMC algorithm has reached the user-defined one.

We end this study on static parameter estimation by providing the ML estimates and their respective standard deviations as well as the posterior means and the posterior deviations
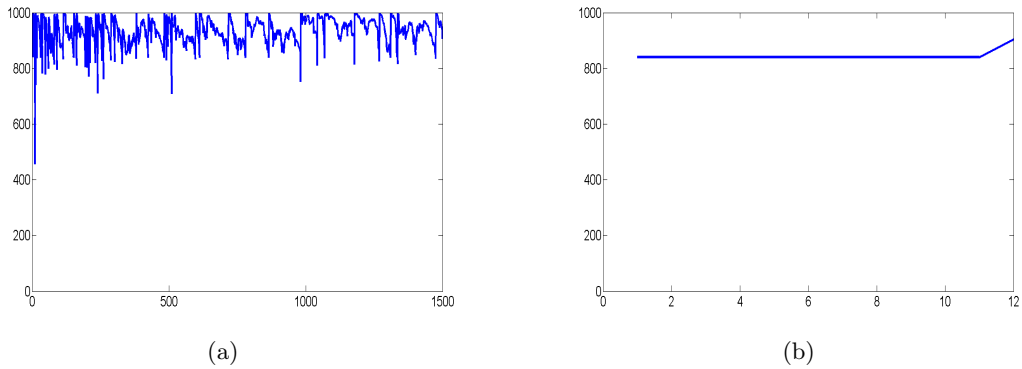
18

|     (a)     |     (b)     |

Figure 1: ESS over SMC iterations of the IBIS algorithm (left) and the E-AIS (right) for estimating the first posterior distribution of interest ($\tau = 1500$ and $M = 1000$). For the IBIS algorithm, several drops occur at the beginning of the sampler whereas the AIS method controls the ESS level at each iteration.


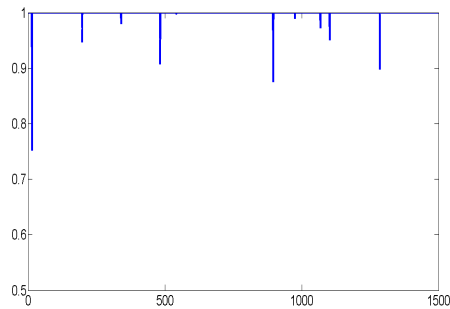
Figure 2: Tempered values after the first E-AIS for a number of particle equal to 1000.

of the SMC methods based on the entire generated data. Table 4 displays the different estimators. All the different methods lead to similar parameter values.

| Algorithms | IBIS | | | |
|---|---|---|---|---|
| Nb. Particles | 200 | 500 | 1000 | 10000 |
| Max. Act. | 3.67 | 1.53 | 5.28 | 3.42 |
| Elapsed time (in hour) | 0.03 | 0.07 | 0.14 | 1.39 |
| MLL | -5957.78 | -5957.80 | -5957.83 | -5957.87 |
| Avg. Acc. rate | 0.33 | 0.33 | 0.33 | 0.33 |
| Algorithms | E-AIS with IBIS | | | |
| Nb. Particles | 200 | 500 | 1000 | 10000 |
| Max. Act. | 8.03 | 3.49 | 2.72 | 7.09 |
| Elapsed time (in hour) | 0.02 | 0.04 | 0.08 | 0.76 |
| MLL | -5957.82 | -5957.73 | -5957.66 | -5957.80 |
| Avg. Acc. rate | 0.32 | 0.32 | 0.32 | 0.32 |
| Algorithms | TNT | | | |
| Nb. Particles | 200 | 500 | 1000 | 10000 |
| Avg. Max. Act. | 1.29 | 7.11 | 4.04 | 2.29 |
| Elapsed time (in hour) | 0.02 | 0.04 | 0.08 | 0.76 |
| MLL | -5957.85 | -5957.70 | -5957.86 | -5957.78 |
| Cum. ESS Gain | 26.41 | 51.67 | 183.75 | 1282.37 |
| Avg. Acc. rate | 0.31 | 0.30 | 0.32 | 0.32 |

Table 3: Summary statistics of the different SMC algorithms.

|  | $\mu$ | $\omega$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| ML estimates | | | | |
| | 0.02 | 0.16 | 0.08 | 0.87 |
| | (0.03) | (0.04) | (0.01) | (0.02) |
| IBIS estimates | | | | |
| | 0.02 | 0.19 | 0.09 | 0.86 |
| | (0.03) | (0.05) | (0.01) | (0.03) |
| E-AIS with IBIS estimates | | | | |
| | 0.02 | 0.19 | 0.09 | 0.86 |
| | (0.03) | (0.05) | (0.01) | (0.03) |
| TNT estimates | | | | |
| | -0.01 | 0.19 | 0.09 | 0.86 |
| | (0.03) | (0.05) | (0.01) | (0.03) |

Table 4: Maximum likelihood estimates and their respective standard errors compared to the parameter posterior mean of the SMC algorithms and their respective standard errors for $M = 10000$.

## 3.2 Dynamic parameters : The Discrete MSM Model

The univariate MSM model relies on $\bar{k}$ first order Markovian processes where $\bar{k}$ is determined by model selection. These processes are combined with the unconditional variance of the series to generate the variance dynamic. More precisely, defining a financial time series $y_{1:T} = \{y_1, ..., y_T\}$, the model is specified as follows :

$$y_t \;=\; \mu + \sigma \prod_{i=1}^{\bar{k}} M_{i,t}^{\frac{1}{2}} \epsilon_t \quad \text{with} \quad \epsilon_t \sim \text{ i.i.d. } N(0,1) \tag{23}$$

where the first-order Markovian processes are independently driven by the transition probabilities :

$$M_{i,t} \;=\; M_{i,t-1} \text{ with probability } 1 - \gamma_i$$
$$M_{i,t} \;\sim\; M \qquad \text{with probability } \gamma_i$$

$M$ denotes the shared stationary distribution of the Markov processes and satisfies the non-restrictive assumption $E(M) = 1$ and $E(M^2) > 0$ imposed by the fractal theory. It is usually set to a discrete distribution or a Log-Normal one. Choosing a continuous distribution enhances the model flexibility but also renders unfeasible the estimation by Maximum Likelihood (MLE). The model can still be estimated by particle filter (Calvet, Fisher, and Thompson (2006)) and by Generalized Methods of Moments (see Lux (2006)). Lux investigated the Log-Normal distribution and found no evidence with respect to the discrete distribution. This paper focuses on the discrete one so the distribution $M$ only counts two different states that appear with equal probability: $s_1 = m_0$ and $s_2 = 2 - m_0$ where $m_0$ is a random parameter. The binomial distribution satisfies $E(M) = 1$ and $E(M^2) > 0$ for any values of $m_0$. To identify the parameter as well as to ensure positiveness of the variance, the parameter $m_0$ belongs to the set $]0, 1[$ .

The transition probabilities $\gamma_i$ that characterize different frequencies of the Markovian processes are specified as

$$\gamma_i = 1 - (1 - \gamma_1)^{b^{i-1}} \qquad \forall \ \ 1 < i \le \bar{k}$$

If $b > 1$, the probabilities grow with $k$ and verify the inequalities: $0 \le \gamma_1 \le \ldots \le \gamma_{\bar{k}} \le 1$. The Markovian processes are thus generated from constraint transition probabilities and evolve

at different frequencies. The frequencies ensure the long memory of the process $y_t$ in level as $\bar{k} \to \infty$ (for a proof, see Calvet and Fisher (2004)).

The model only contains five parameters : the unconditional mean $\mu$ and variance $\sigma^2$, the parameter $m_0$ of the discrete distribution $M$ and the parameters $\{\gamma_{\bar{k}}, b\}$ that determine the transition probabilities. However, the SMC algorithms will also include the $\bar{k}$ latent factors for easing the estimation, a technique called data augmentation. Therefore, the parameter space increases over time and the IBIS algorithm is no longer practical.

To fully specify the model, the table 5 displays the prior distributions of the MSM parameters. All the distributions are almost uninformative.

| | | |
|---|---|---|
| $\mu \sim \text{N}(0,10)$ | $\sigma^{-2} \sim G(3, 0.5)$ | $m_0 \sim U[0,1]$ |
| $\gamma_{\bar{k}} \sim \text{U}[0,1]$ | $b \sim U[1,30]$ | |

Table 5: Prior Distributions of the MSM parameters. The distribution $N(a, b)$ denotes the normal distribution with expectation $a$ and variance $b$, $G(a, b)$ stands for the Gamma distribution with density function : $f(\sigma^{-2}|a, b) = \frac{1}{\Gamma(a)b^a} \sigma^{-2(a-1)} e^{-\frac{1}{\sigma^2 b}}$ and $U[a, b]$ is the Uniform distribution.

We simulate a time series of $T = 1000$ with DGP given by table 6. The exercise considers parameter estimations from $\tau = 300$ until the end of the sample. The MSM MCMC scheme used in the SMC algorithms is documented in Appendix B where, as already mentioned, we augment the parameter set with $\{M_{i,1:T}\}_{i=1}^{\bar{k}}$ for easing the estimation as well as for being able to estimate MSM models involving any number $\bar{k}$ of Markovian processes. By doing so, we improve over the ML approach of Calvet and Fisher (2004) which works for values of $\bar{k}$ below or equal to 10.

| $\mu$ | $\sigma^2$ | $m_0$ | $b$ | $\gamma_{\bar{k}}$ | $\bar{k}$ |
|---|---|---|---|---|---|
| 0 | 2 | 0.8 | 3 | 0.5 | 4 |

Table 6: Data generating process of the MSM model.

In the context of evolving parameter space, the alternative to the E-AIS algorithm is the RM one (see Gilks and Berzuini (2001), equivalent to apply the TNT algorithm on the time domain only). Figure 3a shows the ESS criterion of the RM algorithm for estimating the parameters of the data with $\tau = 300$ (and a number of particles $M = 1000$). Two relevant differences with Figure 3b should be put forward. Firstly, the RM algorithm requires many more SMC iterations (300 compared to 14 for the E-AIS sampler). We also observe some relevant ESS drops. For example, a minimum of 334 diversified particles is achieved at RM iteration 209. For precluding such degeneracies, the RM algorithm should be run with more particles.
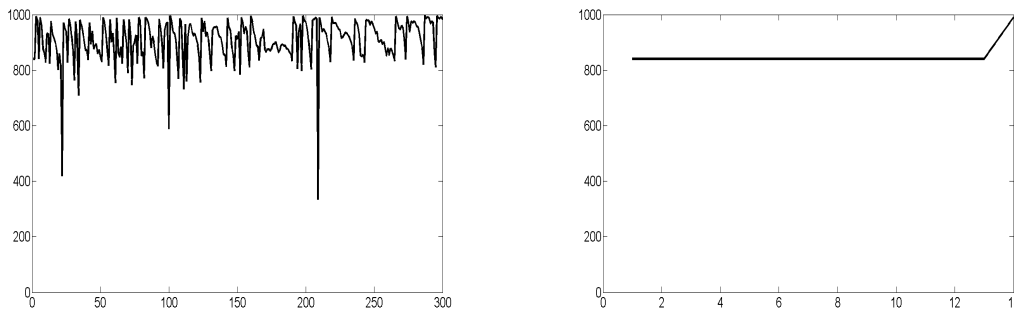


Figure 3: ESS over SMC iterations of the RM algorithm (left) and the E-AIS (right) for estimating the first posterior distribution of interest ($\tau = 300$ and $M = 1000$). Sharp ESS drops occur during the simulation.

We carry out a comparison between RM, E-AIS with RM and TNT algorithms on the MSM data (with $\tau = 300$ and $T = 1000$). As in the previous exercise, the E-AIS with RM algorithm uses the E-AIS algorithm for estimating the first posterior distribution of interest and afterward, it applies the RM algorithm to update observation by observation the approximated distribution. Figure 4 exhibits the ESS levels of each algorithm obtained during the simulation with a number of particles equal to 1000. The TNT algorithm ensures that no ESS below $\kappa_1 = 500$ happens. If it is the case, an E-AIS is run in order to avoid this level of degeneracy. For the two other algorithms, many relevant drops are depicted. At these iterations, the algorithms only depend on a small set of particles for exploring the posterior distribution which can sometimes results in misleading inferences. Consider now the Table 7 that documents the following summary statistics : the number of particles used for the parameter estimation, the

24

elapsed time in hours and the estimated MLL. For the TNT algorithm, we also document the cumulative ESS gain due to switches to the tempered domain as well as the number of re-run E-AIS sampler when the ESS should be below $\kappa_1$. Looking at the elapsed time category, the E-AIS with RM algorithm is the fastest SMC. It dominates the RM algorithm because the first posterior distribution is obtained by tempered function instead of iterating on the time domain. The TNT algorithm is slower than the two other SMC algorithms due to the re-launch of the E-AIS sampler at several times in order to preclude degeneracy issues (see *Nb of E-AIS* on the Table). Note that the cumulative ESS gain encourages the use of the TNT algorithm since the levels are staggeringly high. Finally, the MLL estimates from all the different methods are similar. Regarding the accuracy of the different methods, the Table 8 shows the MLE and the SMC parameter estimates given the entire data set. The SMC algorithms exhibit accurate posterior means and standard deviations that are comparable to the MLE.
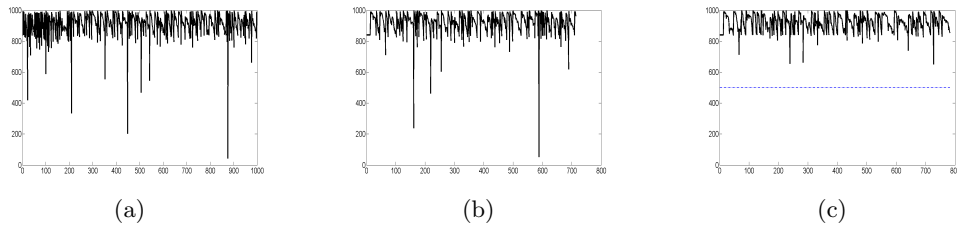


Figure 4: Effective Sample Size over SMC iterations of the RM algorithm on Figure (a), of the E-AIS combined with RM algorithm on Figure (b) and of the TNT algorithm on Figure (c) .

| Algorithms | RM | | | |
|---|---|---|---|---|
| Nb. Particles | 200 | 500 | 1000 | 10000 |
| Elapsed time (in hour) | 0.07 | 0.18 | 0.36 | 2.95 |
| MLL | -1757.00 | -1755.25 | -1755.89 | -1756.22 |
| Algorithms | E-AIS with RM | | | |
| Nb. Particles | 200 | 500 | 1000 | 10000 |
| Elapsed time (in hour) | 0.05 | 0.15 | 0.31 | 2.78 |
| MLL | -1758.05 | -1756.71 | -1757.27 | -1756.53 |
| Algorithms | TNT | | | |
| Nb. Particles | 200 | 500 | 1000 | 10000 |
| Elapsed time (in hour) | 0.10 | 0.22 | 0.41 | 4.15 |
| MLL | -1757.44 | -1757.20 | -1757.01 | -1756.27 |
| Cum. ESS Gain | 709.81 | 1759.70 | 3414.57 | 41554.26 |
| Nb of E-AIS | 3 | 2 | 2 | 3 |

Table 7: Summary statistics of the different SMC algorithms for MSM data.

| $\mu$ | $\sigma^2$ | $m_0$ | $b$ | $\gamma_{\bar{k}}$ |
|---|---|---|---|---|
| | | ML estimates | | |
| -0.01 | 1.95 | 0.78 | 3.76 | 1.00 |
| (0.04) | (0.10) | (0.22) | (1.48) | (20.16) |
| | | RM estimates | | |
| -0.00 | 1.99 | 0.79 | 5.34 | 0.74 |
| (0.04) | (0.26) | (0.05) | (7.62) | (0.18) |
| | | E-AIS with RM estimates | | |
| -0.01 | 1.99 | 0.79 | 4.64 | 0.74 |
| (0.04) | (0.28) | (0.04) | (7.03) | (0.19) |
| | | TNT estimates | | |
| -0.01 | 1.98 | 0.80 | 5.16 | 0.77 |
| (0.04) | (0.27) | (0.05) | (7.31) | (0.17) |

Table 8: Maximum likelihood estimates and their respective standard errors compared to the parameter posterior mean of the SMC algorithms and their respective standard errors for $M = 1000$.

# 4 Empirical applications

The TNT algorithm allows for comparing complex models through marginal likelihoods and predictive marginal likelihoods. We propose to examine the predictive performances of the MSM model in relation to the standard GARCH and the Change-point (CP-) GARCH ones. The subsection (4.1) briefly describes the CP-GARCH specification while the next one (4.2) focuses on model comparisons.

## 4.1 Change-point GARCH model

As alternative to the MSM model, we investigate the CP-GARCH model (see e.g. Bauwens, Dufays, and De Backer (2011)) that allows for abrupt switching in the parameter values of the GARCH model. We consider the following process :

$$y_t = \mu + \epsilon_t \quad \text{with} \quad \epsilon_t|y_{1:t-1} \sim N(0, \sigma_t^2) \tag{24}$$

$$\sigma_t^2 = \omega_{s_t} + \alpha_{s_t}\epsilon_{t-1}^2 + \beta_{s_t}\sigma_{t-1}^2 \tag{25}$$

where $s_t$ is a latent variable taking discrete values in $[1, K+1]$, with K being the number of turning points, and is driven by a Markov-chain characterized by a transition matrix $P_{CP}$ :

$$P_{CP} = \begin{pmatrix} p_{1,1} & 1-p_{1,1} & 0 & ... & 0 \\ 0 & p_{2,2} & 1-p_{2,2} & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The parameter $p_{i,j}$ stands for the probability of moving from the state $i$ to the state $j$. The specification does not admit recurrent regimes and counts at most K+1 ones due to the absorbing state at the end of the transition matrix. The number of structural breaks $K$ is determined by marginal likelihood.

The model exhibits a path dependence problem since the variance at a specific time $t$ depends on the lagged variance (which is unobservable and requires the entire path $s_{1:t-1}$ to be computed). In this context, the standard estimation procedure of Chib (1998) turns out to be infeasible. For carrying out the estimation, we adapt the MCMC method of Bauwens, Dufays, and De Backer (2011) to the TNT algorithm. The MCMC move alternates between DREAM proposals for continuous parameters and the discrete version of it (given in equation

(19)) for sampling the break date parameters. More details about the estimations are given in appendix C.

The table 9 documents the prior distributions of the model parameters.

---

**Mean parameter :**

$$\mu \sim N(0, 0.01)$$

**GARCH parameters** $\forall i \in [1, 2]$ **:**

$$\omega_i \sim U[0, 1] \quad \alpha_i | \beta_i \sim U[0, 1 - \beta_i] \quad \beta_i \sim U[0.5, 1]$$

**Transition parameters** $\forall i \in [1, 2]$ **:**

$$\{p_{i,i}\} \sim \text{Beta}(a_p = T, b_p = 1)$$

---

Table 9: Prior Distributions of the CP parameters. The distribution $N(a, b)$ denotes the normal distribution with expectation $a$ and variance $b$ and U[a,b] stands for the uniform distribution.

## 4.2 Model comparison on the S&P 500

The study bears on the S&P 500 daily percentage returns spanning from May 20, 1999 to April 25, 2011 (3000 observations). We estimate the different models using the TNT algorithm and we fix the value $\tau = 1500$ which controls the very first swing from the tempered to the time domain. In the next subsections we detail the results obtained on the entire sample and then we pay attention to the predictive performance of the different models from May 06, 2005 (1500th observation) by looking at the marginal log-likelihoods (MLLs) and the predictive marginal likelihoods over time. The MSM model is estimated with a number of latent vectors equal to 8 which is sufficient in financial applications (see Calvet and Fisher (2004), Lux (2006)). The number of considered regimes for the CP-GARCH process goes up to 4.

### 4.2.1 Results on the entire sample

Focusing on the whole sample and the MLL values of the different models, it appears that the MSM model dominates the CP-GARCH model whatever the number of regimes. As emphasized by Table 10 that summarizes the MLLs, even the best CP-GARCH model (i.e. with four regimes) performs poorly compared to the MSM one. The Graphic 5 shows the return upper bound of the 95% confidence interval implied by two models as well as the detected break dates for the CP-GARCH model with four regimes. The turning points occurring in June 09, 2003, in February 22, 2007 and in March 12, 2007 can also be observed on the Graphic by a change of volatility dynamics. The very short regime lasting one month in 2007 is characterized by a spectacular negative return that has not been seen since 2003 and puts forward the CP-GARCH ability of capturing outliers.

| CP-GARCH | K=0 | K=1 | K=2 | K=3 | MSM model |
|----------|------|------|------|------|-----------|
| MLL | -4504.94 | -4502.79 | -4498.80 | -4498.80 | **-4472.70** |

Table 10 : S&P 500 daily percentage returns : MLL of the different models. The highest value is bolded.

### 4.2.2 Model comparison over time

For the 1501 estimated posterior distributions (from $\pi(x_\tau|y_{1:\tau})$ to $\pi(x_T|y_{1:T})$), the TNT algorithm provides an estimator of the MLL for each model. We can therefore observe the MLL progress over the period. A sharp decrease means that the model cannot easily capture the new observation. Figures 6 show the log-Bayes factors with respect to the GARCH model for the best CP-GARCH and the MSM ones. We remind that the log-Bayes factor (BF) is computed as the difference of the MLL of two models. Following the informal rule of Kass and Raftery (1995), if the logarithm of the Bayes factor exceeds 3, we have strong evidence in favor of the model with the highest value.

We first observe that the MSM model is always highly better than the benchmark one (i.e. GARCH model with normal innovations). According to Kass and Raftery, we have strong
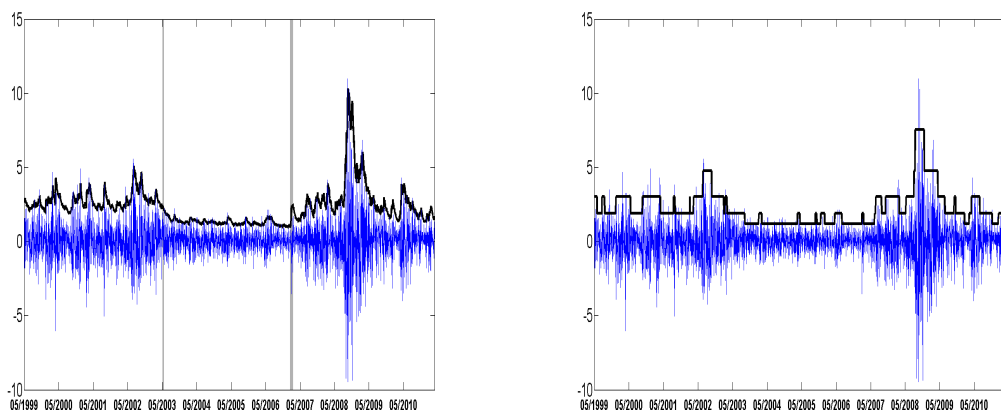
Figure 5: S&P 500 daily percentage returns from 1999 to 2011. On the left plot, the upper bound of the 95% confidence interval of the time series implied by the CP-GARCH with four regimes as well as the detected break dates. The right Graphic displays the upper bound of the 95% confidence interval implied by the MSM model with 8 latent vectors.

evidence in favor of the MSM model over the GARCH model for the entire period. Regarding the other alternative, the Graphics document that the CP-GARCH model, though outperforms the standard GARCH one, competes badly with the MSM one.

Simultaneous sharp increases can be depicted on the plots. At these dates, the standard GARCH dynamic is definetely not able to foresee the next return. The most staggering one happens in February 27, 2007. On Graphic 5, we can observe that this day belongs to the short regime occurring just before the financial crisis. In order to connect this outlier to the financial crisis, it was the very day where Freddie Mac company announced that it will no longer buy the most risky subprime mortgages and mortgage-related securities.

We now turn to the model forecast abilities in terms of predictive marginal log-likelihoods (PMLLs). The PMLL, which corresponds to a ratio of marginal likelihoods (i.e. $\frac{f(Y_{1:t+h})}{f(Y_{1:t})}$), is computed on three different horizons : the shortest one ($h = 1$), a mid-term prediction ($h = 5$) and long term one ($h = 50$). Table 11 documents the average of predictive marginal log-likelihoods over the period as well as their respective standard deviation. Although the results are more mitigated than the previous exercise, we can still observe that the MSM
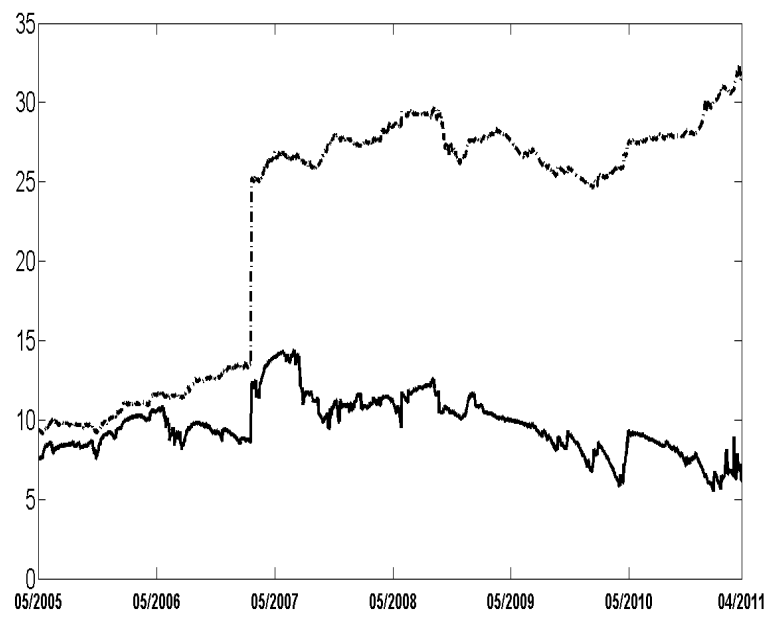
31

Figure 6: S&P 500 log-Bayes factors (log-BF) over time of the volatility models in relation to the GARCH one. In black, the log-BF of the CP-GARCH model and in dashed black, the MSM model. A positive value provide evidence in favor of the considered model compared to the GARCH one.

values are always higher than those of its competitors for any chosen horizon. Interestingly, the CP-GARCH model and the GARCH one seem to perform equally although the former process exhibits more complexity and requires advanced estimation tools. Nevertheless none of these conclusions are significant.

| Horizon | h=1 | h=5 | h=50 |
|---------|-----|-----|------|
| GARCH | -1.47 | -7.34 | -74.01 |
| | (1.14) | (3.49) | (27.11) |
| CP-GARCH | -1.47 | -7.34 | -74.06 |
| | (1.09) | (3.42) | (27.31) |
| MSM | **-1.45** | **-7.26** | **-73.28** |
| | (1.02) | (3.37) | (27.69) |

Table 11: Average of the PMLLs of the different models for several horizons. The highest values are bolded.

## 5   Conclusion

We develop an off-line and on-line SMC algorithm (called TNT) well-suited for situations where a relevant number of similar distributions has to be estimated. The method encompasses the off-line AIS of Neal (1998), the on-line IBIS algorithm of Chopin (2002) and the RM method of Gilks and Berzuini (2001) that all arise as special cases in the SMC sampler theory (see Del Moral, Doucet, and Jasra (2006)). The TNT algorithm benefits from the conjugacy of the tempered and the time domains to avoid particle degeneracies observed in the on-line methods. Moreover, the algorithm is automated for choosing the sequence of posterior distributions using in the tempered domain and we introduce, in the SMC context, a new adaptive RW MCMC scheme which combines the DREAM algorithm with the self-tuning parameter method of Atchadé and Rosenthal (2005), making the TNT algorithm fully generic (besides the user-defined number of particles). Simulations show that the TNT algorithms perform as good as the other on-line SMC but without exhibiting any particle degeneracies. The TNT sampler is an on-line algorithm that allows for easily comparing complex models.

The paper therefore highlights this feature by comparing the MSM model with the GARCH and the CP-GARCH models on the S&P 500 daily percentage returns. The marginal log-likelihood clearly indicates evidence in favor of the MSM model compared to the other processes. The results based on the predictive marginal log-likelihoods are more mitigated.

We believe that the TNT algorithm could be adapted to recent SMC algorithms such as Fearnhead and Taylor (2013) and Fulop and Li (2013) since they propose advanced SMC samplers based on the IBIS and the E-AIS samplers.

# A   Proof of Proposition 1

Using the notation $x_{1:n}^{1:M} = \{x_1^1, ..., x_1^M, x_2^1, ..., x_n^M\}$ which stands for NxM random variables and assuming that $x_i^j \in E \ \ \forall i, j$ as in the E-AIS method (tempered domain) or the IBIS one (time domain), we consider the augmented posterior distribution :

$$\tilde{\pi}_n(x_{1:n}^{1:M}) \quad = \quad [\prod_{i=1}^{M} \pi_n(x_n^i)] \prod_{k=2}^{n} L_k(x_{k-1}^1 | x_k^{1:M}) \prod_{q=2}^{M} L_k(x_{k-1}^q | x_{k-1}^{1:q-1}, x_k^{q:M})$$

If the backward kernels $L_k(.|.)$ denote proper distributions, the product of the distribution of interest marginally arises :

$$\tilde{\pi}_n(x_n^{1:M}) \quad = \quad \int [\prod_{i=1}^{M} \pi_n(x_n^i)] \prod_{k=2}^{n} L_k(x_{k-1}^1 | x_k^{1:M}) \prod_{q=2}^{M} L_k(x_{k-1}^q | x_{k-1}^{1:q-1}, x_k^{q:M}) dx_{1:n-1}^{1:M}$$

$$= \quad [\prod_{i=1}^{M} \pi_n(x_n^i)].$$

The SMC sampler with DREAM MCMC kernels leads to a proposal distribution of the form :

$$\eta_n(x_n^{1:M}) \quad = \quad [\prod_{i=1}^{M} f(x_1^i)] \prod_{k=2}^{n} K_k(x_k^1 | x_{k-1}^{1:M}) \prod_{q=2}^{M} K_k(x_k^q | x_k^{1:q-1}, x_{k-1}^{q:M})$$

where $K_k(.|.)$ denotes the DREAM subkernel with invariant distribution $\pi_k(.)$. Sampling one draw from this proposal distribution is achieved by firstly drawing $M$ realizations from the prior distribution and then applying the DREAM algorithm (N-1)xM times. As proven in Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) and Bauwens, Dufays, and

De Backer (2011), the DREAM algorithm leads to the detailed balance equation :

$$[\prod_{i=1}^{M} \pi_k(x_{k-1}^i)]K_k(x_k^1|x_{k-1}^{1:M}) \prod_{q=2}^{M} K_k(x_k^q|x_k^{1:q-1}, x_{k-1}^{q:M})$$

$$= [\prod_{i=1}^{M} \pi_k(x_k^i)]K_k(x_{k-1}^1|x_k^{1:M}) \prod_{q=2}^{M} K_k(x_{k-1}^q|x_{k-1}^{1:q-1}, x_k^{q:M})$$

Using this relation, we specify the backward kernel as

$$L_k(x_{k-1}^1|x_k^{1:M}) \prod_{q=2}^{M} L_k(x_{k-1}^q|x_{k-1}^{1:q-1}, x_k^{q:M})$$

$$= \frac{[\prod_{i=1}^{M} \pi_k(x_{k-1}^i)]K_k(x_k^1|x_{k-1}^{1:M}) \prod_{q=2}^{M} K_k(x_k^q|x_k^{1:q-1}, x_{k-1}^{q:M})}{[\prod_{i=1}^{M} \pi_k(x_k^i)]}.$$

The sequential importance sampling procedure generates weights given by

$$\tilde{w}_n(x_{1:n}^{1:M}) \equiv \frac{\tilde{\pi}_n(x_{1:n}^{1:M})}{\eta_n(x_n^{1:M})} = \tilde{w}_{n-1}(x_{1:n-1}^{1:M}) \frac{[\prod_{i=1}^{M} \pi_n(x_{n-1}^i)]}{[\prod_{i=1}^{M} \pi_{n-1}(x_{n-1}^i)]}$$

$$= \prod_{i=1}^{M} w_n(x_{1:n}^i),$$

resulting in a product of independent weights exactly equal to the product of SMC sampler weights (see equation (4)).

# B    Appendix : Markov-switching Multifractal model estimation

In this appendix, we document how to simulate the MSM parameters from the joint posterior distribution using an MCMC framework. Let gather all the MSM parameters into the set $\theta = \{\mu, \sigma^2, m_0, \gamma_{\bar{k}}, b\}$. For estimating MSM model with any number of state vector $\bar{k}$, we augment the posterior distribution with the $\bar{k}$ Markov chain processes $\bar{M}_{1:t} = \{M_{i,1:t}\}_{i=1}^{\bar{k}}$. The MCMC scheme exhibits as invariant distribution the posterior distribution $\pi_n(\theta, \bar{M}_{1:t}|y_{1:t}) = \frac{f(y_{1:t}|\theta, \bar{M}_{1:t})^{\phi(n)}f(\bar{M}_{1:t}|\theta)f(\theta)}{Z_n}$ where $Z_n$ is the normalizing constant.

Given a set of parameters $\{\theta, \bar{M}_{1:t}\}$, one MCMC iteration operates as follows :

1. For $i = 1, ..., \bar{k}$, sample $M_{i,1:t} \sim \pi_n(M_{i,1:t}|y_{1:t}, \theta, \bar{M}_{1:t} \backslash M_{i,1:t})$ by forward-backward algorithm (see Rabiner (1989)).

2. Randomly set the block size $r$ of $\theta : r \sim U[1, 5]$

3. Sample $\theta$ by blocks of size $r$ using the DREAM algorithm.

## C    Appendix : estimation of CP-GARCH using SMC sampler

We detail here the CP-GARCH estimation carried out by SMC sampler. An efficient Metropolis-Hastings algorithm resting on the DREAM proposal as well as its discrete version has been proposed in Bauwens, Dufays, and De Backer (2011). This technique constitutes the rejuvenate step of the SMC sampler. To understand the MCMC moves, observe that it exists a one-to-one mapping between the state vector $s_{1:T}$ and discrete break parameters $T = \{\tau_0, \tau_1, ..., \tau_K\}$ where $\tau_0 = 1$ and $\tau_i \; \forall i \in [1, K]$ denotes the very observation where the process jumps from state $i$ to state $i + 1$.

Let denote $\theta = \{\mu, \omega_1, \alpha_1, \beta_1, ..., \omega_{K+1}, \alpha_{K+1}, \beta_{K+1}\}$, the set of GARCH parameters. We now provide the MCMC scheme to estimate a CP-GARCH model relying on DREAM and D-DREAM proposals :

1. Sample $\tau_1, ..., \tau_K$ by blocks of random size using D-DREAM algorithm (see equation (19)).

2. $\forall i \in [1, K]$, sample $p_{ii} \sim \text{Beta}(a_p + \tau_i - \tau_{i-1}, b_p + 1)$.

3. Sample $\theta$ by blocks of random size using the DREAM algorithm (see equation (18)).

## References

ANDRIEU, C., A. DOUCET, AND R. HOLENSTEIN (2010): "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society B*, 72, 269–342.

ATCHADÉ, Y., AND J. ROSENTHAL (2005): "On adaptive Markov chain Monte Carlo algorithms," *Bernoulli*, 11(5), 815–828.

BAUWENS, L., A. DUFAYS, AND B. DE BACKER (2011): "Estimating and forecasting structural breaks in financial time series," *CORE discussion paper, 2011/55.*

BAUWENS, L., A. DUFAYS, AND J. ROMBOUTS (2013): "Marginal Likelihood for Markov Switching and Change-point GARCH Models," *Journal of Econometrics*, 178 (3), 508–522.

BESKOS, A., A. JASRA, AND A. THIERY (2013): "On the Convergence of Adaptive Sequential Monte Carlo Methods," *Available at http://arxiv.org/pdf/1306.6462v2.pdf.*

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

CALVET, L. E., AND A. J. FISHER (2004): "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes," *Journal of Financial Econometrics*, 2(1), 49–83.

CALVET, L. E., A. J. FISHER, AND S. B. THOMPSON (2006): "Volatility comovement: a multifrequency approach," *Journal of Econometrics*, 131(1-2), 179–215.

CHIB, S. (1995): "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

——— (1998): "Estimation and comparison of multiple change-point models," *Journal of Econometrics*, 86, 221–241.

CHIB, S., AND I. JELIAZKOV (2001): "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

CHIB, S., F. NARDARI, AND N. SHEPHARD (2002): "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108(2), 281–316.

CHOPIN, N. (2002): "A Sequential Particle Filter Method for Static Models," *Biometrika*, 89, 539–551.

CHOPIN, N., P. E. JACOB, AND O. PAPASPILIOPOULOS (2013): "SMC2: an efficient algorithm for sequential analysis of state space models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 397–426.

DEL MORAL, P., A. DOUCET, AND A. JASRA (2006): "Sequential Monte Carlo samplers," *The Royal Statistical Society: Series B(Statistical Methodology)*, 68, 411–436.

——— (2012): "On adaptive resampling strategies for sequential Monte Carlo methods," *Bernoulli*, 18(1), 252–278.

DURHAM, G., AND J. GEWEKE (2012): "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," *http://www.censoc.uts.edu.au/pdfs/gewekepapers/gpu2full.pdf*.

FEARNHEAD, P., AND B. TAYLOR (2013): "An adaptive Sequential Monte Carlo Sampler," *Bayesian Analysis*, 8 (2), 411–438.

FULOP, A., AND J. LI (2013): "Efficient learning via simulation: A marginalized resample-move approach," *Journal of Econometrics*, 176(2), 146 – 161.

GEWEKE, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

GEYER, C. J. (1992): "Practical Markov Chain Monte Carlo," *Statistical Science*, 7(4), 473–511.

GILKS, W. R., AND C. BERZUINI (2001): "Following a moving target - Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 127–146.

GORDON, N., D. SALMOND, AND A. F. M. SMITH (1993): "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F*, 140(2), 107–113.

HERBST, E., AND F. SCHORFHEIDE (2012): "Sequential Monte carlo sampling for DSGE models," *Working Paper No12-27, Federal reserve Bank of Philadelphia*.

JASRA, A., D. A. STEPHENS, A. DOUCET, AND T. TSAGARIS (2011): "Inference for Lévy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo," *Scandinavian Journal of Statistics*, 38, 1–22.

JASRA, A., D. A. STEPHENS, AND C. C. HOLMES (2007): "On population-based simulation for static inference," *Statistics and Computing*, 17(3), 263–279.

JEREMIAH, E., S. SISSON, L. MARSHALL, R. MEHROTRA, AND A. SHARMA (2011): "Bayesian calibration and uncertainty analysis of hydrological models: A comparison of adaptive Metropolis and sequential Monte Carlo samplers," *Water Resources Research*, 47(7), n/a–n/a.

KALMAN, R. E. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D), 35–45.

KASS, R., AND A. RAFTERY (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

LUX, T. (2006): "The Markov-Switching Multifractal Model of asset returns: GMM estimation and linear forecasting of volatility," Economics Working Papers 2006,17, Christian-Albrechts-University of Kiel, Department of Economics.

MENG, X.-L., AND W. WONG (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity : A theoretical Exploration," *Statistica Sinica*, 6, 831–860.

NEAL, R. M. (1998): "Annealed Importance Sampling," *Statistics and Computing*, 11, 125–139.

RABINER, L. R. (1989): "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, pp. 257–286.

ROBERTS, G. O., AND J. S. ROSENTHAL (2001): "Optimal scaling for various Metropolis-Hastings algorithms," *Statistical Science*, 16, 351–367.

SMITH, A. F. M., AND A. E. GELFAND (1992): "Bayesian Statistics without Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46, 84–88.

VRUGT, J. A., C. J. F. TER BRAAK, C. G. H. DIKS, B. A. ROBINSON, J. M. HYMAN, AND D. HIGDON (2009): "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptative Randomized Subspace Sampling," *International Journal of Nonlinear Sciences and Numerical Simulations*, 10, 271–288.

**NATIONAL BANK OF BELGIUM - WORKING PAPERS SERIES**

The Working Papers are available on the website of the Bank: http://www.nbb.be.

209. "Wage and employment effects of a wage norm: The Polish transition experience" by A. de Crombrugghe and G. de Walque, *Research series*, February 2011.

210. "Estimating monetary policy reaction functions: A discrete choice approach" by J. Boeckx, *Research series*, February 2011.

211. "Firm entry, inflation and the monetary transmission mechanism" by V. Lewis and C. Poilly, *Research series*, February 2011.

212. "The link between mobile telephony arrears and credit arrears" by H. De Doncker, *Document series*, March 2011.

213. "Development of a financial health indicator based on companies' annual accounts", by D. Vivet, *Document series*, April 2011.

214. "Wage structure effects of international trade: Evidence from a small open economy", by Ph. Du Caju, F. Rycx and I. Tojerow, *Research series*, April 2011.

215. "Economic importance of the Belgian ports: Flemish maritime ports, Liège port complex and the port of Brussels - Report 2009", by C. Mathys, *Document series,* June 2011.

216. "Verti-zontal differentiation in monopolistic competition", by F. Di Comite, J.-F. Thisse and H. Vandenbussche, *Research series*, October 2011.

217. "The evolution of Alexandre Lamfalussy's thought on the international and European monetary system (1961-1993)" by I. Maes, *Research series*, November 2011.

218. "Economic importance of air transport and airport activities in Belgium – Report 2009", by X. Deville and S. Vennix, *Document series*, December 2011.

219. "Comparative advantage, multi-product firms and trade liberalisation: An empirical test", by C. Fuss and L. Zhu, *Research series*, January 2012.

220. "Institutions and export dynamics", by L. Araujo, G. Mion and E. Ornelas, *Research series*, February 2012.

221. "Implementation of EU legislation on rail liberalisation in Belgium, France, Germany and the Netherlands", by X. Deville and F. Verduyn, *Document series*, March 2012.

222. "Tommaso Padoa-Schioppa and the origins of the euro", by I. Maes, *Document series*, March 2012.

223. "(Not so) easy come, (still) easy go? Footloose multinationals revisited", by P. Blanchard, E. Dhyne, C. Fuss and C. Mathieu, *Research series*, March 2012.

224. "Asymmetric information in credit markets, bank leverage cycles and macroeconomic dynamics", by A. Rannenberg, *Research series*, April 2012.

225. "Economic importance of the Belgian ports: Flemish maritime ports, Liège port complex and the port of Brussels - Report 2010", by C. Mathys, Document series, July 2012.

226. "Dissecting the dynamics of the US trade balance in an estimated equilibrium model", by P. Jacob and G. Peersman, *Research series*, August 2012.

227. "Regime switches in volatility and correlation of financial institutions", by K. Boudt, J. Daníelsson, S.J. Koopman and A. Lucas, *Research series*, October 2012.

228. "Measuring and testing for the systemically important financial institutions", by C. Castro and S. Ferrari, *Research series*, October 2012.

229. "Risk, uncertainty and monetary policy", by G. Bekaert, M. Hoerova and M. Lo Duca*, Research series*, October 2012.

230. "Flights to safety", by L. Baele, G. Bekaert, K. Inghelbrecht and M. Wei, *Research series*, October 2012.

231. "Macroprudential policy, countercyclical bank capital buffers and credit supply: Evidence from the Spanish dynamic provisioning experiments", by G. Jiménez, S. Ongena, J.-L. Peydró and J. Saurina, *Research series*, October 2012.

232. "Bank/sovereign risk spillovers in the European debt crisis", by V. De Bruyckere, M. Gerhardt, G. Schepens and R. Vander Vennet, *Research series*, October 2012.

233. "A macroeconomic framework for quantifying systemic risk", by Z. He and A. Krishnamurthy, *Research series*, October 2012.

234. "Fiscal policy, banks and the financial crisis", by R. Kollmann, M. Ratto, W. Roeger and J. in't Veld, *Research series*, October 2012.

235. "Endogenous risk in a DSGE model with capital-constrained financial intermediaries", by H. Dewachter and R. Wouters, *Research series*, October 2012.

236. "A macroeconomic model with a financial sector", by M.K. Brunnermeier and Y. Sannikov, *Research series*, October 2012.

237. "Services versus goods trade: Are they the same?", by A. Ariu, *Research series*, December 2012.

238. "Importers, exporters, and exchange rate disconnect", by M. Amiti, O. Itskhoki and J. Konings, *Research series*, December 2012.

239. "Concording EU trade and production data over time", by I. Van Beveren, A.B. Bernard and H. Vandenbussche, *Research series*, December 2012.

240. "On the origins of the Triffin dilemma: Empirical business cycle analysis and imperfect competition theory", by I. Maes, *Research series*, December 2012.

241. "The Influence of the Taylor rule on US monetary policy", by P. Ilbas, Ø. Røisland and T. Sveen, *Research series*, January 2013.

242. "Economic importance of the Belgian ports: Flemish maritime ports, Liège port complex and the port of Brussels - Report 2011", by C. Mathys, *Document series,* July 2013.

243. "The fragility of two monetary regimes: The European Monetary System and the Eurozone", by P. De Grauwe and Y. Ji, *Research series*, October 2013.

244. "Funding liquidity, market liquidity and TED spread: A two-regime model", by K. Boudt, E. C.S. Paulus and D. W.R. Rosenthal, *Research series*, November 2013.

245. "Robustifying optimal monetary policy using simple rules as cross-checks", by P. Ilbas, Ø. Røisland and T. Sveen, *Research series*, November 2013.

246. "Household and firm leverage, capital flows and monetary policy in a small open economy", by M. Pirovano, *Research series*, November 2013.

247. "The BIS and the Latin American debt crisis of the 1980s", by P. Clement and I. Maes, *Research series*, December 2013.

248. "The importance of the right amount of business resources for firms' exporting behavior", by I. Paeleman, C. Fuss and T. Vanacker, *Research series*, December 2013.

249. "The role of financial frictions during the crisis: An estimated DSGE model", by R. Merola, *Research series*, December 2013.

250. "Bank reactions after capital shortfalls", by C. Kok and G. Schepens, *Research series*, December 2013.

251. "Why firms avoid cutting wages: Survey evidence from European firms", by P. Du Caju, T. Kosma, M. Lawless, J. Messina and T. Rõõm, *Research series*, December 2013.

252. "The distribution of debt across euro area countries: The role of individual characteristics, institutions and credit conditions", by O. Bover, J. M. Casado, S. Costa, Ph. Du Caju, Y. McCarthy, E. Sierminska, P. Tzamourani, E. Villanueva and T. Zavadil, *Research series*, December 2013.

253. "Micro-based evidence of EU competitiveness: The CompNet database", by CompNet Task Force, *Research series*, March 2014.

254. "Information in the yield curve: A macro-finance approach", by H. Dewachter, L. Iania and M. Lyrio, *Research series*, March 2014.

255. "The Single supervisory mechanism or 'SSM', part one of the Banking Union", by E. Wymeersch, *Research series*, April 2014.

256. "Nowcasting Belgium", by D. de Antonio Liedo, *Research series*, April 2014.

257. "Human capital, firm capabilities and productivity growth", by I. Van Beveren and S. Vanormelingen, *Research series*, May 2014.

258. "Monetary and macroprudential policies in an estimated model with financial intermediation", by P. Gelain and P. Ilbas, *Research series*, May 2014.

259. "A macro-financial analysis of the euro area sovereign bond market", by H. Dewachter, L. Iania, M. Lyrio and M. de Sola Perea, *Research series*, June 2014.

260. "Economic importance of the Belgian ports: Flemish maritime ports, Liège port complex and the port of Brussels - Report 2012", by C. Mathys, *Document series*, June 2014.

261. "European competitiveness: A semi-parametric stochastic metafrontier analysis at the firm level", by M. Dumont, B. Merlevede, G. Rayp and M. Verschelde, *Document series*, July 2014.

262. "Employment, hours and optimal monetary policy", by M. Dossche, V. Lewis and C. Poilly, *Research series*, September 2014.

263. "On the conjugacy of off-line and on-line Sequential Monte Carlo Samplers", by A. Dufays, *Research series*, September 2014.