Likelihood inference for Markov switching GARCH(1,1) models using sequential Monte Carlo

Damien C. H. Wee[a], Feng Chen[a,*] and William T. M. Dunsmuir[a]

[a]*University of New South Wales, Sydney, Australia*

ARTICLE INFO

Keywords:
Regime switching
Missing data
Intractable likelihood
Particle filter
Time series
Volatility modelling

ABSTRACT

Markov switching (MS-)GARCH(1,1) models allow for structural changes in volatility dynamics between a finite number of regimes. Since the regimes are not observed, computation of the likelihood requires integrating over an exponentially increasing number of regime paths, which is intractable. An existing smooth likelihood estimation procedure for sequential Monte Carlo (SMC), that is currently limited to hidden Markov models with a one-dimensional state variable, is modified to enable likelihood estimation and maximisation for MS-GARCH(1,1) models, a model which requires two dimensions, volatility and regime, to evolve its hidden state process. Furthermore, the modified SMC procedure is shown to be easily adapted to fitting MS-GARCH(1,1) models even when there are missing observations. The proposed methodology is validated with simulated data and is also illustrated with analysis of two financial time series, the daily returns on the S&P 500 index and on the Henry Hub natural gas spot price, with the latter series containing a gap caused by shutdown in response to hurricane Rita in 2005.

1. Introduction

Since its introduction by Bollerslev [9] the GARCH model has been widely and successfully applied, primarily for modelling the volatility of financial returns, but also in other applications such as high frequency wind speed [cf. 11]. The GARCH(1,1) model is the following system of recurrence equations

$$Y_i = \sigma_i z_i + \mu, \tag{1}$$

$$\sigma_i^2 = \omega + \alpha(Y_{i-1} - \mu)^2 + \beta\sigma_{i-1}^2, \tag{2}$$

for $i = 1, \dots, N$, where $(Y_i)_{i=1,\dots,N}$ is termed the observation series, $(z_i)_{i=1,\dots,N}$ is an independently and identically distributed (i.i.d) noise series with density d and finite second moment, with $\omega > 0, \alpha \geq 0, \beta \geq 0$ and $\mu \in \mathbb{R}$ being model parameters. Nelson [28] showed that (1)-(2) has a unique, strictly stationary and ergodic solution if and only if

$$\mathbb{E}(\log(\alpha z_1^2 + \beta)) < 0. \tag{3}$$

It has been argued that artificially high persistence (measured as $\alpha + \beta$) obtained in empirical studies utilising the standard GARCH(1,1) specification (1)-(2) can be avoided by allowing the GARCH parameters to evolve over time; see for instance [13], [26] and [27]. Indeed, studies such as [31], [22] and [29] have found that stock returns possess different characteristics during expansionary and contractionary phases of the business cycle. Of course, the standard GARCH model with its fixed parameters can have trouble accounting for this stylised fact.

One could somehow pre-process the underlying data set to take out the effects of business cycles or other structural changes. However, as phenomena such as business cycles are random in nature, the fitted model obtained after deterministically pre-processing out the effects of business cycles is not useful for forecasting purposes unless one has in addition a mechanism to incorporate the random nature of future business cycles.

A popular method, first utilised by Hamilton [21], for incorporating structural changes of a random nature into a model structure, is to assume the world can exist in one of a finite set of regimes and assume the data generating process has the same model structure across all regimes but that each regime puts in effect its own set of model

*Corresponding author
feng.chen@unsw.edu.au (F. Chen); w.dunsmuir@unsw.edu.au (W.T.M. Dunsmuir)
https://web.maths.unsw.edu.au/~fengchen (F. Chen);
https://research.unsw.edu.au/people/professor-william-t-m-dunsmuir (W.T.M. Dunsmuir)
ORCID(s): 0000-0002-9646-3338 (F. Chen); 0000-0002-7932-4408 (W.T.M. Dunsmuir)

parameters. A finite state discrete time Markov chain is then used to model the regime evolution. Enriching the standard GARCH specification, in the spirit of [21], yields what Francq and Zakoian [17] and Bauwens et al. [5], among others, call the Markov switching (MS-)GARCH model. Formally, let $(R_i)_{i=1,\dots,N}$ be an unobserved discrete time ergodic homogeneous Markov chain on a finite space $\mathcal{R} := \{1,\dots,J\}$ corresponding to J different regimes. Then the MS-GARCH(1,1) is defined as the following system

$$Y_i = \sigma_i z_i + \mu(R_i), \tag{4}$$

$$\sigma_i^2 = \omega(R_i) + \alpha(R_i)(Y_{i-1} - \mu(R_{i-1}))^2 + \beta(R_i)\sigma_{i-1}^2, \tag{5}$$

where $\omega(.), \alpha(.), \beta(.)$ and $\mu(.)$ are functions taking values on the respective finite sets $\omega(\mathcal{R}) := \{\omega_1,\dots,\omega_J\} \in (0,\infty)^J, \alpha(\mathcal{R}) := \{\alpha_1,\dots,\alpha_J\} \in [0,\infty)^J, \beta(\mathcal{R}) := \{\beta_1,\dots,\beta_J\} \in [0,\infty)^J$ and $\mu(\mathcal{R}) := \{\mu_1,\dots,\mu_J\} \in \mathbb{R}^J$. Denote the $J \times J$ transition matrix which governs the regime evolution by $\mathbf{P}$ with elements $P_{k,l} := \mathbb{P}(R_i = l | R_{i-1} = k)$. Let $\pi_k := \mathbb{P}(R_1 = k), k = 1,\dots,J$ be the stationary distribution of the regime Markov chain. Francq et al. [16] showed that a strictly stationary solution to (4)-(5) exists if and only if

$$\sum_{k=1}^{J} \pi_k \mathbb{E}(\log(\alpha_k z_1^2 + \beta_k)) < 0. \tag{6}$$

Therefore, for strict stationarity it need not be the case that under all regimes the stability condition (3) of [28] on the GARCH parameters hold. Thus, there is flexibility to preserve strict stationarity while allowing regimes with unstable GARCH parameters, in the sense of persistences in excess of 1, to exist. It was shown in [5] that the higher the unconditional probabilities of being in regimes with stable GARCH parameters, the higher the persistence of unstable regimes can assume whilst maintaining the condition (6).

As one does not observe the regimes, computation of the likelihood requires integrating over all possible regime paths. This path dependency renders the likelihood intractable as the number of possible paths grow exponentially with the observation periods making maximum likelihood estimation (MLE) infeasible.

Several alternative specifications of a GARCH model incorporating Markov switching effects, devised with intention of preserving tractability of the likelihood function, have been proposed. The models by Gray [19], Dueker [14] and Klaassen [25] avoid path dependency by collapsing the conditional variances in each regime into a single variance at each time point, thereby removing from the variance evolution dependency on the full history of regimes. The model by Haas et al. [20] avoids path dependency by separating the GARCH dynamics from the regime process.

Despite these alternative specifications where MLE is feasible, the specification (4)-(5) as stated by Klaassen [25] and Haas et al. [20] is the most natural application of the GARCH(1,1) model in a regime switching context. As such, much research, as we will summarise shortly, into estimation methods for the specification (4)-(5) which bypass exact evaluation of the likelihood have been proposed. Hereafter by MS-GARCH we are referring to the specification (4)-(5) with condition (6).

A number of Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures have been developed for these MS-GARCH models. These methods circumvent the path dependency problem by including the whole regime path in the parameter space. The first of these procedures was developed by Bauwens et al. [5] which employed a Gibbs sampling algorithm to sample the regime at each time point individually. This single-move Gibbs sampler showed slow convergence, motivating the work of [4], which develop a particle Gibbs sampler to instead sample jointly the whole regime path history. Further research into multi-move sampling techniques for Bayesian inference of MS-GARCH models was conducted in [7].

In terms of frequentist methods for this class of model, the first was developed by Augustyniak [2] who employed the Monte Carlo Expectation Maximisation algorithm, simulating regimes from the single-move Gibbs sampler of Bauwens et al. [5]. Later, a non-simulation based frequentist approach was developed by Augustyniak et al. [3] based on a collapsing procedure that generalises those of [19], [14] and [25].

Other developments towards estimation are the works of Francq and Zakoian [17] who propose a generalised method of moments approach, Bildirici and Ersin [6] who utilise neural networks, and Elliott et al. [15] who propose a Viterbi based technique for sampling the regimes.

In employing a particle filter approach, Bauwens et al. [4] was able to obtain, through simulation, an estimated likelihood function. However, the standard bootstrap particle filter (a.k.a sequential Monte Carlo (SMC)) likelihood estimates, although consistent and unbiased [cf. 12], sample the particles in such a way that makes the likelihood estimates prone to being discontinuous functions of the parameters [cf. 30], thus inhibiting their direct use in numerical optimisation to obtain the MLE, which is why Bauwens et al. [4] resorted to a MCMC approach.

The purpose of this paper is twofold. Firstly, we show that it is in fact possible to modify the standard SMC likelihood estimation procedure [18] to yield estimated likelihood surfaces that are indeed amenable to numerical optimisation, thereby providing a computationally feasible method for parameter estimation of MS-GARCH(1,1) models through simulated maximum likelihood. This SMC procedure adds to the class of frequentist methods developed for these MS-GARCH(1,1) models, utilising a superior sampling mechanism through particle filtering than the single-move Gibbs sampler employed in [2] and not resorting to the auxiliary model structure used in [3].

Secondly, we extend our modified SMC approach to deal with parameter estimation of MS-GARCH(1,1) series when one has missing observations – a real world problem, so far largely unaddressed in the current literature, although encountered frequently for instance with most high-frequency economic or financial time series that are not measured on holidays or weekends, even though "*economic activity continues as a product of political and social phenomena that will have a direct influence on the next value of the studied index*" [10]. Even ignoring gaps due to holidays or weekends, financial instruments with low liquidity can have missing values on trading days. Such a problem is faced when modelling bond yields from corporations or governments of smaller emerging nations [cf. 24]. Outside of financial applications, the GARCH model has been applied to model wind speed volatility, in which case mechanical failures of measurement equipment can indeed lead to gaps in wind data.

Organisation of this paper is as follows. Section 2 describes the exponentially increasing in time state space of the volatility, which results from an unobserved regime process and renders exact calculation of the likelihood of the MS-GARCH(1,1) model infeasible. Our modified SMC procedure, for parameter estimation of MS-GARCH(1,1) models through simulated maximum likelihood is introduced in Section 3. Extension of the SMC procedure, for a solution to parameter estimation, when in addition to the unobserved regime process, one only has partial observation of the MS-GARCH(1,1) series, is outlined in Section 4. Section 5.1 compares through simulation studies the performance of the SMC estimation method to the estimates obtained using the approximating model structure of [3]. The additional functionality of the SMC method, as a parameter estimation method when faced with varying degrees of missingness in the MS-GARCH(1,1) series, is tested in Section 5.2. Two real world applications of the SMC method are illustrated. Section 6.1 analyses the volatility of stock returns allowing for structural changes brought about by phenomena such as changing business cycles. Section 6.2 studies the volatility of natural gas spot price returns, a commodity subjected to cyclical fluctuations in consumer demand and also a real world example when unscheduled trading interruptions, such as natural disasters, result in gaps in the observation series. The paper closes with some concluding remarks in Section 7.

2. Likelihood

Hereafter, for ease of exposition, we restrict our discussion to the case of only two possible regimes. Denote the (unobserved) regime path up to time i as $R_{1:i} := \{R_1, \dots, R_i\}$, the observation history up to time i as $Y_{1:i} := \{Y_1, \dots, Y_i\}$ (and for notational convenience $Y_{1:0} \equiv 0$) and define the functions

$$h_i(Y_{1:i-1}, R_{1:i}) := \omega(R_i) + \alpha(R_i)\Big(Y_{i-1} - \mu(R_{i-1})\Big)^2 + \beta(R_i)h_{i-1}(Y_{1:i-2}, R_{1:i-1}), \tag{7}$$

for $i = 2, \dots, N$ with $h_1(0, R_1) := \mathbb{E}(\sigma_1^2|R_1)$. From [17], $\mathbb{E}(\sigma_1^2|R_1 = r), r = 1, 2$ are obtained by solving the stationary equations

$$\begin{bmatrix} P_{1,1}\omega_1 & P_{2,1}\omega_1 \\ P_{1,2}\omega_2 & P_{2,2}\omega_2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} 1 - P_{1,1}(\alpha_1 + \beta_1) & -P_{2,1}(\alpha_1 + \beta_1) \\ -P_{1,2}(\alpha_2 + \beta_2) & 1 - P_{2,2}(\alpha_2 + \beta_2) \end{bmatrix} \begin{bmatrix} \pi_1 \mathbb{E}(\sigma_1^2|R_1 = 1) \\ \pi_2 \mathbb{E}(\sigma_1^2|R_1 = 2) \end{bmatrix}, \tag{8}$$

where $\pi_1 = \frac{P_{2,1}}{P_{1,2}+P_{2,1}}$ and $\pi_2 = \frac{P_{1,2}}{P_{1,2}+P_{2,1}}$.

For general i, there are 2^i possible realisations of $R_{1:i}$ which we denote by $R_{1:i}^{(k)}, k = 1, \dots, 2^i$. These in turn, given $Y_{1:i-1}$, result in 2^i possible realisations of σ_i^2 given by $h_i(Y_{1:i-1}, R_{1:i}^{(k)}) =: \sigma_i^{2(k)}, k = 1, \dots, 2^i$ that occur, conditional on $Y_{1:i-1}$, with probability $\mathbb{P}(R_{1:i} = R_{1:i}^{(k)}|Y_{1:i-1}) = \pi_{R_1^{(k)}} A_1^{(k)} \dots A_{i-1}^{(k)} =: w_i^{(k)}$, where

$$A_{i-1}^{(k)} := P_{R_{i-1}^{(k)}, R_i^{(k)}} \frac{p(Y_{i-1}|R_{1:i-1} = R_{1:i-1}^{(k)}, Y_{1:i-2})}{p(Y_{i-1}|Y_{1:i-2})}, \tag{9}$$

with $R_j^{(k)}$, $j = 1, \dots, i$ being the j-th component of $R_{1:i}^{(k)}$. Thus, we have

$$\mathbb{P}(\sigma_i^2, R_i | Y_{1:i-1}) = \sum_{k=1}^{2^i} w_i^{(k)} \delta_{\{\sigma_i^{2(k)}, R_i^{(k)}\}}(\sigma_i^2, R_i), \tag{10}$$

where $\delta_{\{a,b\}}$ is the bivariate Dirac measure at the point (a, b). The quantity (10) is a sum of size 2^i. When the number of observations N is not too large, one could compute the likelihood via

$$p(Y_{1:N}) = \prod_{i=2}^{N} p(Y_i|Y_{1:i-1})p(Y_1) = \prod_{i=2}^{N} \Big(\sum_{\sigma_i^2, R_i} p(Y_i|\sigma_i^2, R_i)\mathbb{P}(\sigma_i^2, R_i|Y_{1:i-1}) \Big) p(Y_1). \tag{11}$$

However, in most practical situations N will be too large to evaluate (11) using (10). In the next section, a modified SMC approach is presented which effectively trims, so to speak, the branches (σ_i^2, R_i) such that for all i, they never exceed 2^q for some chosen q. Trimming is conducted by resampling from the set of branches, weighting them by how likely they would result in the given observation and in a manner such that the resultant approximated likelihood surface is amenable to numerical optimisation. For the SMC algorithm that will be proposed, the relevant state variable is (σ_i^2, R_i) rather than the sequence of regimes, where it is this state variable that will be approximated along with the probabilities $\mathbb{P}(\sigma_i^2, R_i|Y_{1:i-1})$.

3. Modified SMC approach

The SMC approach we propose is as follows. Calculate the likelihood exactly up to time q. Then at time q we note that we have the exact quantities

$$\mathbb{P}(R_q = r|Y_{1:q}) = \frac{\sum_{j=1}^{2^{q-1}} W_{q,j,r}}{\sum_{s=1}^{2} \sum_{j=1}^{2^{q-1}} W_{q,j,s}} \quad \text{and} \quad \mathbb{P}(\sigma_q^2|Y_{1:q}, R_q = r) = \sum_{j=1}^{2^{q-1}} w_{q,r}^{(j)} \delta_{\{\sigma_{q,r}^{2(j)}\}}(\sigma_q^2), \tag{12}$$

for each $r = 1, 2$, where

$$W_{q,j,r} := p(Y_q|\sigma_q^2 = \sigma_{q,r}^{2(j)}, R_q = r)\mathbb{P}(R_{1:q} = \{R_{1:q-1}^{(j)}, r\}|Y_{1:q-1}), \tag{13}$$

$$w_{q,r}^{(j)} := W_{q,j,r} / \sum_{k=1}^{2^{q-1}} W_{q,k,r} \tag{14}$$

and

$$\sigma_{q,r}^{2(j)} := h_q(Y_{1:q-1}, \{R_{1:q-1}^{(j)}, r\}), \tag{15}$$

for $j = 1, \dots, 2^{q-1}$. The idea going forward is to approximate $\mathbb{P}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i})$, $i = q, \dots, N-1$, for which the pair $(\sigma_{i+1}^2, R_{i+1})$ can take on 2^{i+1} possible values by an approximate measure $\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i})$ for which $(\sigma_{i+1}^2, R_{i+1})$ is restricted to 2^q possible values, denote these by $(\sigma_{i+1}^{2(j)}, R_{i+1}^{(j)})$, $j = 1, \dots, 2^q$. Essentially we trim the tree, never allowing the possible branch nodes $(\sigma_{i+1}^2, R_{i+1})$, $i = q, \dots, N-1$, to exceed 2^q. This is achieved using resampling ideas from SMC. Then from these $\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i})$ we will obtain the approximate likelihood components $\hat{p}(Y_{i+1}|Y_{1:i})$ via

$$p(Y_{i+1}|Y_{1:i}) = \sum_{\sigma_{i+1}^2, R_{i+1}} p(Y_{i+1}|\sigma_{i+1}^2, R_{i+1})\mathbb{P}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i}) \tag{16}$$

$$\approx \sum_{j=1}^{2^q} p(Y_{i+1}|\sigma_{i+1}^2 = \sigma_{i+1}^{2(j)}, R_{i+1} = R_{i+1}^{(j)})\hat{\mathbb{P}}(\sigma_{i+1}^2 = \sigma_{i+1}^{2(j)}, R_{i+1} = R_{i+1}^{(j)}|Y_{1:i}) =: \hat{p}(Y_{i+1}|Y_{1:i}). \tag{17}$$

Now, note that

$$\mathbb{P}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i}) = \sum_{\sigma_i^2, R_i} \mathbb{P}(\sigma_{i+1}^2, R_{i+1}|\sigma_i^2, R_i, Y_i)\mathbb{P}(\sigma_i^2, R_i|Y_{1:i}) \tag{18}$$

and

$$\sigma_{i+1}^2|\sigma_i^2, R_i, Y_i = \begin{cases} h(Y_i, \sigma_i^2, R_i, 1) & \text{w.p. } P_{R_i,1}, \\ h(Y_i, \sigma_i^2, R_i, 2) & \text{w.p. } P_{R_i,2}, \end{cases} \tag{19}$$

where $h(y, \sigma^2, r, k) := \omega_k + \alpha_k(y - \mu_r)^2 + \beta_k\sigma^2$. Thus to obtain $\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i})$ for which the couple $(\sigma_{i+1}^2, R_{i+1})$ is restricted to 2^q possible values, due to the regime branching mechanism (19), one can first obtain $\hat{\mathbb{P}}(\sigma_i^2, R_i|Y_{1:i})$ for which (σ_i^2, R_i) is restricted to 2^{q-1} possible values, then replace $\mathbb{P}(\sigma_i^2, R_i|Y_{1:i})$ with $\hat{\mathbb{P}}(\sigma_i^2, R_i|Y_{1:i})$ in (18) to obtain the required $\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i})$. We begin at the first step $i = q$ and discuss several ways $\hat{\mathbb{P}}(\sigma_q^2, R_q|Y_{1:q})$ could be obtained.

1. If we resampled $K = 2^{q-1}$ pairs of (σ_q^2, R_q) from the discrete distribution

$$\mathbb{P}(\sigma_q^2, R_q|Y_{1:q}) = \sum_{r=1}^{2} \mathbb{P}(R_q = r|Y_{1:q}) \sum_{j=1}^{2^{q-1}} w_{q,r}^{(j)} \delta_{\{\sigma_{q,r}^{2(j)}, r\}}(\sigma_q^2, R_q), \tag{20}$$

   this is the bootstrap resampling of [18], which however, generally will not admit SMC likelihood surfaces that are amenable to numerical optimisation [see 30]. The continuous resampling procedure of [30] is only viable when the domain of the hidden state is a connected interval of the real line and thus cannot be applied to resample the two-dimensional hidden state pair (σ_q^2, R_q).
2. If we instead resample $K_1 = \lfloor K\mathbb{P}(R_q = 1|Y_{1:q})\rfloor$ ($\lfloor . \rfloor$ being the floor function) draws of σ_q^2 from a continuous approximation to $\mathbb{P}(\sigma_q^2|Y_{1:q}, R_q = 1)$ and then $K_2 = K - K_1$ draws of σ_q^2 from a continuous approximation to $\mathbb{P}(\sigma_q^2|Y_{1:q}, R_q = 2)$, the problem however, will be that K_1 and K_2 via $\mathbb{P}(R_q|Y_{1:q})$ are dependent on $\theta := (\omega_j, \alpha_j, \beta_j, \mu_j, P_{j,k})_{j,k=1,2}$ and drastically different σ_q^2 could be sampled by switching drawing from one $\mathbb{P}(\sigma_q^2|Y_{1:q}, R_q)$ to the other under a small parameter change, again resulting in discontinuities of the approximated likelihood.
3. To prevent K_1 and K_2 varying across parameter changes, a solution is to set K_1 and K_2 independent of θ. Eliminating the source of discontinuity in procedure 2, this procedure produces sufficiently smooth SMC likelihood surfaces amenable to numerical optimisation. Taking for instance the fair split $K_1 = K_2 = K/2$, this is effectively importance sampling according to (20) drawing the R_q from $\mathbb{Q}(R_q = 1) = \mathbb{Q}(R_q = 2) = 1/2$ instead of $\mathbb{P}(R_q = r|Y_{1:q}), r = 1, 2$.

Proceeding then as per procedure 3 with the fair split, the recipe is to initialise $\hat{\mathbb{P}}(\sigma_q^2|Y_{1:q}, R_q) = \mathbb{P}(\sigma_q^2|Y_{1:q}, R_q)$ and $\hat{\mathbb{P}}(R_q|Y_{1:q}) = \mathbb{P}(R_q|Y_{1:q})$ and then proceed from $\hat{\mathbb{P}}(\sigma_i^2|Y_{1:i}, R_i)$ and $\hat{\mathbb{P}}(R_i|Y_{1:i})$, $i = q, \dots, N-1$ to the next $\hat{\mathbb{P}}(\sigma_{i+1}^2|Y_{1:i+1}, R_{i+1})$ and $\hat{\mathbb{P}}(R_{i+1}|Y_{1:i+1})$ obtaining along the way the required $\hat{\mathbb{P}}(\sigma_i^2, R_i|Y_{1:i})$. To begin, for each $r = 1, 2$ draw σ_i^2, $K/2 = 2^{q-2}$ times from the continuous approximation (details of how to construct this provided later in Section 3.1) to the discrete distribution $\hat{\mathbb{P}}(\sigma_i^2|Y_{1:i}, R_i = r)$, denote these draws by $\hat{\sigma}_{i,r}^{2(j)}$, $j = 1, \dots, 2^{q-2}$. Then construct the empirical distribution

$$\hat{\mathbb{P}}(\sigma_i^2, R_i|Y_{1:i}) := \sum_{r=1}^{2} \frac{\hat{\mathbb{P}}(R_i = r|Y_{1:i})}{K/2} \sum_{j=1}^{2^{q-2}} \delta_{\{\hat{\sigma}_{i,r}^{2(j)}, r\}}(\sigma_i^2, R_i), \tag{21}$$

for which replacing $\mathbb{P}(\sigma_i^2, R_i|Y_{1:i})$ with (21) in (18) and using (19) we arrive at

$$\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i}) := \sum_{j=1}^{2^{q-1}} \bar{w}_{i+1,r}^{(j)} \delta_{\{\sigma_{i+1,r}^{2(j)}, r\}}(\sigma_{i+1}^2, R_{i+1}), \tag{22}$$

for $r = 1, 2$, where

$$\bar{w}_{i+1,r}^{(j)} := \begin{cases} \frac{P_{1,r}\hat{\mathbb{P}}(R_i=1|Y_{1:i})}{K/2}, & \text{for } j = 1, \dots, 2^{q-2}, \\ \frac{P_{2,r}\hat{\mathbb{P}}(R_i=2|Y_{1:i})}{K/2}, & \text{for } j = 2^{q-2}+1, \dots, 2^{q-1}, \end{cases} \tag{23}$$

and

$$\sigma^{2(j)}_{i+1,r} := \begin{cases} h(Y_i, \hat{\sigma}^{2(j)}_{i,1}, 1, r), & \text{for } j = 1, \dots, 2^{q-2}, \\ h(Y_i, \hat{\sigma}^{2(j-2^{q-2})}_{i,2}, 2, r), & \text{for } j = 2^{q-2}+1, \dots, 2^{q-1}. \end{cases} \tag{24}$$

From (17) and (22) we then obtain the approximated likelihood component

$$\hat{p}(Y_{i+1}|Y_{1:i}) = \sum_{r=1}^{2} \sum_{j=1}^{2^{q-1}} \bar{w}^{(j)}_{i+1,r} p(Y_{i+1}|\sigma^2_{i+1} = \sigma^{2(j)}_{i+1,r}, R_{i+1} = r). \tag{25}$$

Note that if z_1 has density d then

$$p(Y_{i+1}|\sigma^2_{i+1}, R_{i+1}) = \frac{1}{\sigma_{i+1}} d\Big(\frac{Y_{i+1} - \mu_{R_{i+1}}}{\sigma_{i+1}}\Big). \tag{26}$$

Next, using (22) and (25), $\hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1})$ is obtained via

$$\mathbb{P}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1}) \approx \frac{p(Y_{i+1}|\sigma^2_{i+1}, R_{i+1} = r)\hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i})}{\hat{p}(Y_{i+1}|Y_{1:i})} \tag{27}$$

$$= \sum_{j=1}^{2^{q-1}} \tilde{w}^{(j)}_{i+1,r} \delta_{\{\sigma^{2(j)}_{i+1,r}, r\}}(\sigma^2_{i+1}, R_{i+1}) =: \hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1}), \tag{28}$$

where

$$\tilde{w}^{(j)}_{i+1,r} := \bar{w}^{(j)}_{i+1,r} p(Y_{i+1}|\sigma^2_{i+1} = \sigma^{2(j)}_{i+1,r}, R_{i+1} = r) / \sum_{s=1}^{2} \sum_{k=1}^{2^{q-1}} \bar{w}^{(k)}_{i+1,s} p(Y_{i+1}|\sigma^2_{i+1} = \sigma^{2(k)}_{i+1,s}, R_{i+1} = s). \tag{29}$$

Furthermore,

$$\mathbb{P}(R_{i+1} = r|Y_{1:i+1}) = \sum_{\sigma^2_{i+1}} \mathbb{P}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1}) \tag{30}$$

$$\approx \sum_{\sigma^2_{i+1}} \hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1}) = \sum_{j=1}^{2^{q-1}} \tilde{w}^{(j)}_{i+1,r} =: \hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1}). \tag{31}$$

Therefore,

$$\hat{\mathbb{P}}(\sigma^2_{i+1}|Y_{1:i+1}, R_{i+1} = r) := \frac{\hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1})}{\hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1})} = \sum_{j=1}^{2^{q-1}} \hat{w}^{(j)}_{i+1,r} \delta_{\{\sigma^{2(j)}_{i+1,r}\}}(\sigma^2_{i+1}), \tag{32}$$

where

$$\hat{w}^{(j)}_{i+1,r} := \tilde{w}^{(j)}_{i+1,r} / \sum_{k=1}^{2^{q-1}} \tilde{w}^{(k)}_{i+1,r}. \tag{33}$$

The required quantities $\hat{\mathbb{P}}(R_{i+1}|Y_{1:i+1})$ and $\hat{\mathbb{P}}(\sigma^2_{i+1}|Y_{1:i+1}, R_{i+1})$ to repeat the process for the next time step $i+1$ have been provided. After iterating through to the last time point, the SMC likelihood approximation is then given by

$$\hat{\mathcal{L}}(Y_{1:N}) := p(Y_{1:q}) \prod_{i=q}^{N-1} \hat{p}(Y_{i+1}|Y_{1:i}). \tag{34}$$

3.1. Continuous approximation to $\hat{\mathbb{P}}(\sigma_i^2|Y_{1:i}, R_i)$

The aforementioned steps required resampling from continuous approximations to the discrete distributions

$$\mathbb{P}(\sigma_q^2|Y_{1:q}, R_q = r) = \sum_{j=1}^{2^{q-1}} w_{q,r}^{(j)} \delta_{\{\sigma_{q,r}^{2(j)}\}}(\sigma_q^2) \quad \text{and} \quad \hat{\mathbb{P}}(\sigma_i^2|Y_{1:i}, R_i = r) = \sum_{j=1}^{2^{q-1}} \hat{w}_{i,r}^{(j)} \delta_{\{\sigma_{i,r}^{2(j)}\}}(\sigma_i^2), \tag{35}$$

for $i = q+1, \dots, N-1$. We detail here how these continuous approximations, for smooth SMC likelihoods, are constructed.

3.1.1. At $i = q$

For $i = q$ the weights $w_{q,r}^{(j)}$, see (14)-(13), for a given r, are in addition to σ_q^2 also functions of $R_{1:q-1}$. Following [30], in this case it is necessary to first smooth the weights $w_{q,r}^{(j)}$ such that $\sigma_{q,r}^{2(j)}$ that are close together have weights close in magnitude; this can be achieved with a kernel smoothing approach. Proceeding, we calculate

$$\dot{w}_{q,r}^{(j)} := \frac{\sum_{k=1}^{2^{q-1}} w_{q,r}^{(k)} \phi\Big((\sigma_{q,r}^{2(k)} - \sigma_{q,r}^{2(j)})/\hbar\Big)}{\sum_{k=1}^{2^{q-1}} \phi\Big((\sigma_{q,r}^{2(k)} - \sigma_{q,r}^{2(j)})/\hbar\Big)} \quad \text{and} \quad \ddot{w}_{q,r}^{(j)} := \dot{w}_{q,r}^{(j)} / \sum_{k=1}^{2^{q-1}} \dot{w}_{q,r}^{(k)}, \tag{36}$$

for $j = 1, \dots, 2^{q-1}$, where $\phi(.)$ is the standard Gaussian density and $\hbar := c/2^{q-1}$ for a very small number c. Sorting $(\sigma_{q,r}^{2(j)}, \ddot{w}_{q,r}^{(j)})$, $j = 1, \dots, 2^{q-1}$ in ascending order by first element, denoting the sorted series by $(\bar{\sigma}_{q,r}^{2(j)}, \breve{w}_{q,r}^{(j)})$, $j = 1, \dots, 2^{q-1}$, a continuous approximation to the cumulative distribution function of $\sigma_q^2|Y_{1:q}, R_q$ is then given by

$$\bar{\mathbb{P}}(\sigma_q^2 \le x|Y_{1:q}, R_q = r) := \gamma_0 \mathbb{I}(x \ge \bar{\sigma}_{q,r}^{2(1)}) + \sum_{k=1}^{K-1} \gamma_k H\Big(\frac{x - \bar{\sigma}_{q,r}^{2(k)}}{\bar{\sigma}_{q,r}^{2(k+1)} - \bar{\sigma}_{q,r}^{2(k)}}\Big) + \gamma_K \mathbb{I}(x \ge \bar{\sigma}_{q,r}^{2(K)}), \tag{37}$$

where $\gamma_0 := \breve{w}_{q,r}^{(1)}/2$, $\gamma_k := (\breve{w}_{q,r}^{(k+1)} + \breve{w}_{q,r}^{(k)})/2$, $k = 1, \dots K-1$ and $\gamma_K := \breve{w}_{q,r}^{(K)}/2$, with $H(z) := \max\Big(0, \min(z, 1)\Big)$. The σ_q^2 are resampled by means of the inversion method applied to the continuous distribution (37). As explained in [30], with the traditional [18] method, the source of discontinuity of the approximated likelihood function across parameter changes is due to application of the inversion method to discrete distribution functions.

3.1.2. At $i > q$

Now, in the case of $i = q+1, \dots, N-1$ note that the weights $\hat{w}_{i,r}^{(j)}$, see (33), (29) and (23), are functions of σ_i^2 and R_{i-1}. Taking advantage of the fact that R_{i-1} can take on only two possible values, a computationally cheaper operation than kernel smoothing is available. Note that the weights $\hat{w}_{i,r}^{(j)}$, $j = 1, \dots, 2^{q-2}$ correspond to $R_{i-1} = 1$ (see (23)) and differ only through $\sigma_{i,r}^{2(j)}$, where the variation in $\hat{w}_{i,r}^{(j)}$ is continuous in $\sigma_{i,r}^{2(j)}$, that is if $\sigma_{i,r}^{2(k)}$ and $\sigma_{i,r}^{2(l)}$ for some $k, l = 1, \dots, 2^{q-2}$ are close in proximity, so too are their respective weights $\hat{w}_{i,r}^{(k)}$ and $\hat{w}_{i,r}^{(l)}$. Likewise, the weights $\hat{w}_{i,r}^{(j)}$, $j = 2^{q-2}+1, \dots, 2^{q-1}$, which correspond to $R_{i-1} = 2$, possess the same attribute. Therefore, we can continuously approximate the individual (sub-distribution) functions,

$$A(x) := \sum_{j=1}^{2^{q-2}} \hat{w}_{i,r}^{(j)} \mathbb{I}(\sigma_{i,r}^{2(j)} \le x) \quad \text{and} \quad B(x) := \sum_{j=2^{q-2}+1}^{2^{q-1}} \hat{w}_{i,r}^{(j)} \mathbb{I}(\sigma_{i,r}^{2(j)} \le x), \tag{38}$$

and then combine both continuous approximations to give a continuous function.

Sort $(\sigma_{i,r}^{2(j)}, \hat{w}_{i,r}^{(j)})$, $j = 1, \dots, 2^{q-2}$ in ascending order by first element and denote the sorted series by $(\sigma_{i,r,1}^{2(j)}, w_{i,r,1}^{(j)})$, $j = 1, \dots, 2^{q-2}$. Similarly, sort $(\sigma_{i,r}^{2(j)}, \hat{w}_{i,r}^{(j)})$, $j = 2^{q-2}+1, \dots, 2^{q-1}$ in ascending order by first element and denote the sorted series by $(\sigma_{i,r,2}^{2(j)}, w_{i,r,2}^{(j)})$, $j = 1, \dots, 2^{q-2}$. The continuous approximations to $A(x)$ and $B(x)$ are given respectively by $\bar{A}(x) = D_1(x)$ and $\bar{B}(x) = D_2(x)$ where

$$D_s(x) := \gamma_{0,s} \mathbb{I}(x \ge \sigma_{i,r,s}^{2(1)}) + \sum_{k=1}^{\bar{K}-1} \gamma_{k,s} H\Big(\frac{x - \sigma_{i,r,s}^{2(k)}}{\sigma_{i,r,s}^{2(k+1)} - \sigma_{i,r,s}^{2(k)}}\Big) + \gamma_{\bar{K},s} \mathbb{I}(x \ge \sigma_{i,r,s}^{2(\bar{K})}), \tag{39}$$

with $\bar{K} := 2^{q-2}, \gamma_{0,s} := w_{i,r,s}^{(1)}/2, \gamma_{\bar{K},s} := w_{i,r,s}^{(\bar{K})}/2$ and $\gamma_{k,s} := (w_{i,r,s}^{(k+1)} + w_{i,r,s}^{(k)})/2$, $k = 1, \dots \bar{K} - 1$, $s = 1, 2$.

Now, the function $\bar{C}(x) := \bar{A}(x) + \bar{B}(x)$ will have four point masses at $x = \sigma_{i,r,1}^{2(1)}, \sigma_{i,r,2}^{2(1)}, \sigma_{i,r,1}^{2(\bar{K})}$ and $\sigma_{i,r,2}^{2(\bar{K})}$. Based on the Pitt and Malik [30] procedure, the two point masses at $\sigma_{\min}^{2(1)} := \min(\sigma_{i,r,1}^{2(1)}, \sigma_{i,r,2}^{2(1)})$ and $\sigma_{\max}^{2(\bar{K})} := \max(\sigma_{i,r,1}^{2(\bar{K})}, \sigma_{i,r,2}^{2(\bar{K})})$ are admissible, however the two point masses at $\sigma_{\max}^{2(1)} := \max(\sigma_{i,r,1}^{2(1)}, \sigma_{i,r,2}^{2(1)})$ and $\sigma_{\min}^{2(\bar{K})} := \min(\sigma_{i,r,1}^{2(\bar{K})}, \sigma_{i,r,2}^{2(\bar{K})})$ may be problematic for smooth SMC likelihoods. Note that $\bar{A}^{-1}(u)$ is continuous in the interval $[0, \sum_{j=1}^{2^{q-2}} w_{i,r,1}^{(j)}]$ and $\bar{B}^{-1}(u)$ is continuous in the interval $[0, \sum_{j=1}^{2^{q-2}} w_{i,r,2}^{(j)}]$. In the event $\sigma_{\max}^{2(1)} \leq \sigma_{\min}^{2(\bar{K})}$, $\bar{C}^{-1}(u)$ is continuous in the interval $[0, 1]$ (see LHS Figure 1), thus the point masses of $\bar{C}(x)$ at $x = \sigma_{\max}^{2(1)}$ and $\sigma_{\min}^{2(\bar{K})}$ are admissible under these circumstances.

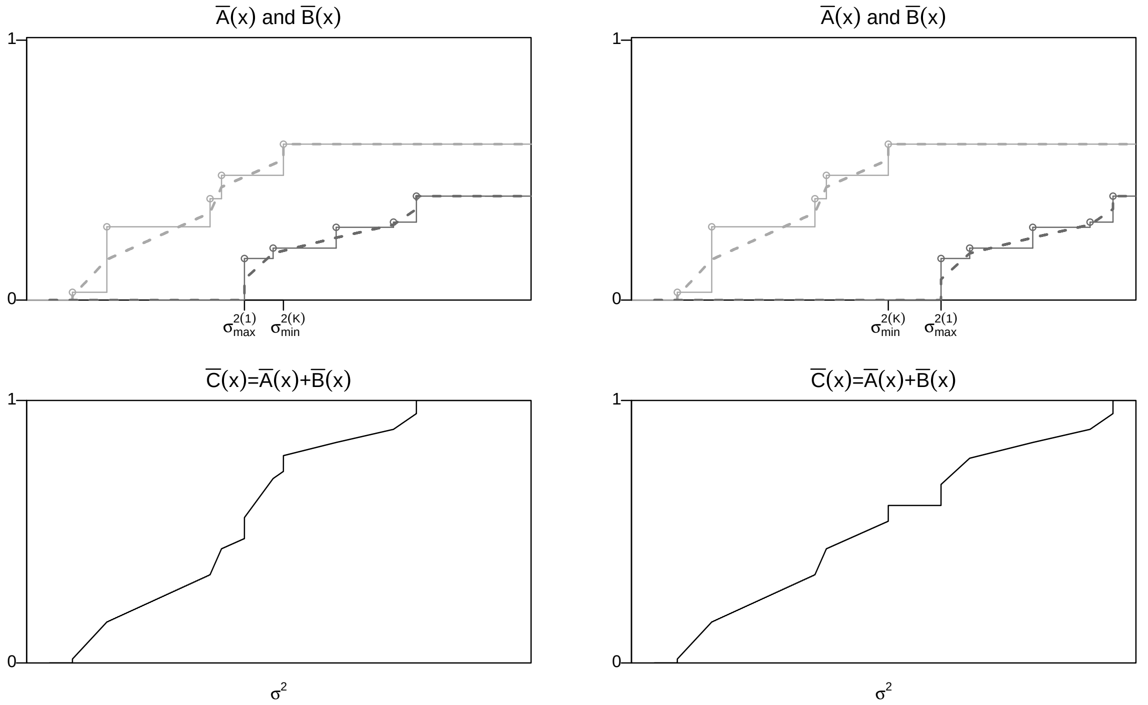

Figure 1: For the two top plots, the light-grey dashed line construct is the continuous approximation $\bar{A}(x)$ to the light-grey solid line function $A(x)$ and the dark-grey dashed line construct is the continuous approximation $\bar{B}(x)$ to the dark-grey solid line function $B(x)$. The top left plot corresponds to the case when $\sigma_{\max}^{2(1)} \leq \sigma_{\min}^{2(K)}$ and the bottom left plot is the resultant $\bar{C}(x)$, for which $\bar{C}^{-1}(u)$ can be seen to be continuous in the interval $[0, 1]$. The top right plot corresponds to the case when $\sigma_{\max}^{2(1)} > \sigma_{\min}^{2(K)}$ and the bottom right plot is the resultant $\bar{C}(x)$, for which $\bar{C}^{-1}(u)$ can be seen to exhibit a discontinuity in the interval $[0, 1]$.

However, in the event $\sigma_{\max}^{2(1)} > \sigma_{\min}^{2(\bar{K})}$, then $\bar{C}^{-1}(u)$ is not continuous in the interval $[0, 1]$ (see RHS Figure 1), having a single discontinuity at

$$u = \begin{cases} \sum_{j=1}^{2^{q-2}} w_{i,r,1}^{(j)} & \text{if } \sigma_{i,r,1}^{2(\bar{K})} < \sigma_{i,r,2}^{2(1)}, \\ \sum_{j=1}^{2^{q-2}} w_{i,r,2}^{(j)} & \text{if } \sigma_{i,r,2}^{2(\bar{K})} < \sigma_{i,r,1}^{2(1)}, \end{cases} \tag{40}$$

leading to the same problem associated with resampling according to a discrete cumulative distribution function. To overcome this issue, we use the following slight correction to $\bar{C}(x)$, rather than $\bar{C}(x)$ itself, to approximate the distribution function associated with (35),

$$\bar{\mathbb{P}}(\sigma_i^2 \le x | Y_{1:i}, R_i = r) := \Upsilon(x) := \begin{cases} \bar{C}(x) & \text{if } \sigma_{\max}^{2(1)} \le \sigma_{\min}^{2(\bar{K})}, \\ C_1(x) & \text{if } \sigma_{i,r,1}^{2(\bar{K})} < \sigma_{i,r,2}^{2(1)}, \\ C_2(x) & \text{if } \sigma_{i,r,2}^{2(\bar{K})} < \sigma_{i,r,1}^{2(1)}, \end{cases} \tag{41}$$

where

$$C_1(x) := \bar{C}(x) + (\gamma_{\bar{K},1} + \gamma_{0,2}) H\Big(\frac{x - \sigma_{\min}^{2(\bar{K})}}{\sigma_{\max}^{2(1)} - \sigma_{\min}^{2(\bar{K})}}\Big) - \gamma_{\bar{K},1} \mathbb{I}(x \ge \sigma_{i,r,1}^{2(\bar{K})}) - \gamma_{0,2} \mathbb{I}(x \ge \sigma_{i,r,2}^{2(1)}) \tag{42}$$

and

$$C_2(x) := \bar{C}(x) + (\gamma_{\bar{K},2} + \gamma_{0,1}) H\Big(\frac{x - \sigma_{\min}^{2(\bar{K})}}{\sigma_{\max}^{2(1)} - \sigma_{\min}^{2(\bar{K})}}\Big) - \gamma_{\bar{K},2} \mathbb{I}(x \ge \sigma_{i,r,2}^{2(\bar{K})}) - \gamma_{0,1} \mathbb{I}(x \ge \sigma_{i,r,1}^{2(1)}). \tag{43}$$

Resampling K values of σ_i^2 from (41) can be done in $O(K)$ operations. See Algorithm 1 in the Appendix for pseudo code summarising the steps taken to proceed from time step i to $i+1$. We should also note that the procedure to combine $\bar{A}(x)$ and $\bar{B}(x)$ to yield $\Upsilon(x)$ can be reiterated to accommodate more than two regimes. For instance, when we have a third regime bringing with it another sub-distribution function $\bar{E}(x)$, one just reiterates the procedure with $\Upsilon(x) \mapsto \bar{A}(x)$ and $\bar{E}(x) \mapsto \bar{B}(x)$. In this way, a continuous approximation that enables smooth resampling of the $\sigma_i^2 | Y_{1:i}, R_i$ when there are more than two regimes is feasible.

4. Extension to the case of missing observations

Now, assume that we do not observe the full MS-GARCH series $Y_{1:N}$ but have only $n < N$ observations taken on an increasing set of times $1 = t_1 < \ldots < t_n = N$, such that $t_i \in \{2, \ldots, N-1\}$ for $i = 2, \ldots, n-1$. The SMC procedure from the previous section can be adapted for parameter estimation of the partially observed MS-GARCH series $Y_{t_1:t_n} := \{Y_{t_1}, \ldots, Y_{t_n}\}$.

Figure 2 displays SMC approximated likelihood profile plots, using the standard bootstrap resampler, for a simulated MS-GARCH(1,1) series with $N = 2000$ and $n = 1600$, that is only 80% of the series is (randomly) observed. As can be seen, the approximated likelihood surface produced using SMC with the standard bootstrap resampling mechanism is profoundly discontinuous (even with the randomness across parameter changes fixed). On the other hand, Figure 3 displays the approximated likelihood profile plots of the same MS-GARCH(1,1) simulated data set using the modified resampling procedure we will detail in this section. While, for reasons to be clear shortly, the procedure we describe is not guaranteed to be continuous in the transition probability parameters $P_{1,2}$ and $P_{2,1}$, there is undeniably a vast improvement in smoothness in Figure 3 over Figure 2, in fact, to the extent of being sufficiently smooth for use with numerical optimisers. Indeed, as will be seen in the simulation studies of Section 5.2, this procedure we propose proves itself to be a computationally feasible and accurate method for parameter estimation of MS-GARCH(1,1) series in the presence of missing observations.

We now proceed to detail the modified SMC procedure. Start by assuming that at time step i one has

$$\tilde{\mathbb{P}}(\sigma_{t_i+1}^2 | Y_{t_1:t_i}, R_{t_i+1} = r) := \sum_{k=1}^{2^{q-1}} v_{i,r}^{(k)} \delta_{\{\sigma_{t_i+1,r}^{2(k)}\}}(\sigma_{t_i+1}^2), \tag{44}$$

for which $v_{i,r}^{(k)}, \sigma_{t_i+1,r}^{2(k)}, r = 1, 2, k = 1, \ldots, 2^{q-1}$ along with $\hat{\mathbb{P}}(R_{t_i+1} | Y_{t_1:t_i})$ are determined recursively from the previous time step $i-1$ and their derivation revealed as we proceed to obtain $v_{i+1,r}^{(k)}, \sigma_{t_{i+1}+1,r}^{2(k)}$, $r = 1, 2$, $k = 1, \ldots, 2^{q-1}$ and $\hat{\mathbb{P}}(R_{t_{i+1}+1} | Y_{t_1:t_{i+1}})$, required to repeat the procedure at the next time step. Proceeding, for each $r = 1, 2$, draw 2^{q-2} values of $\sigma_{t_i+1}^2$ from a continuous approximation (details given shortly) to the discrete distribution (44). Denote these draws by $\hat{\sigma}_{t_i+1,r}^{2(k)}$, $k = 1, \ldots, 2^{q-2}, r = 1, 2$ and construct

$$\hat{\mathbb{P}}(\sigma_{t_i+1}^2 | Y_{t_1:t_i}, R_{t_i+1} = r) := \frac{1}{2^{q-2}} \sum_{k=1}^{2^{q-2}} \delta_{\{\hat{\sigma}_{t_i+1,r}^{2(k)}\}}(\sigma_{t_i+1}^2). \tag{45}$$

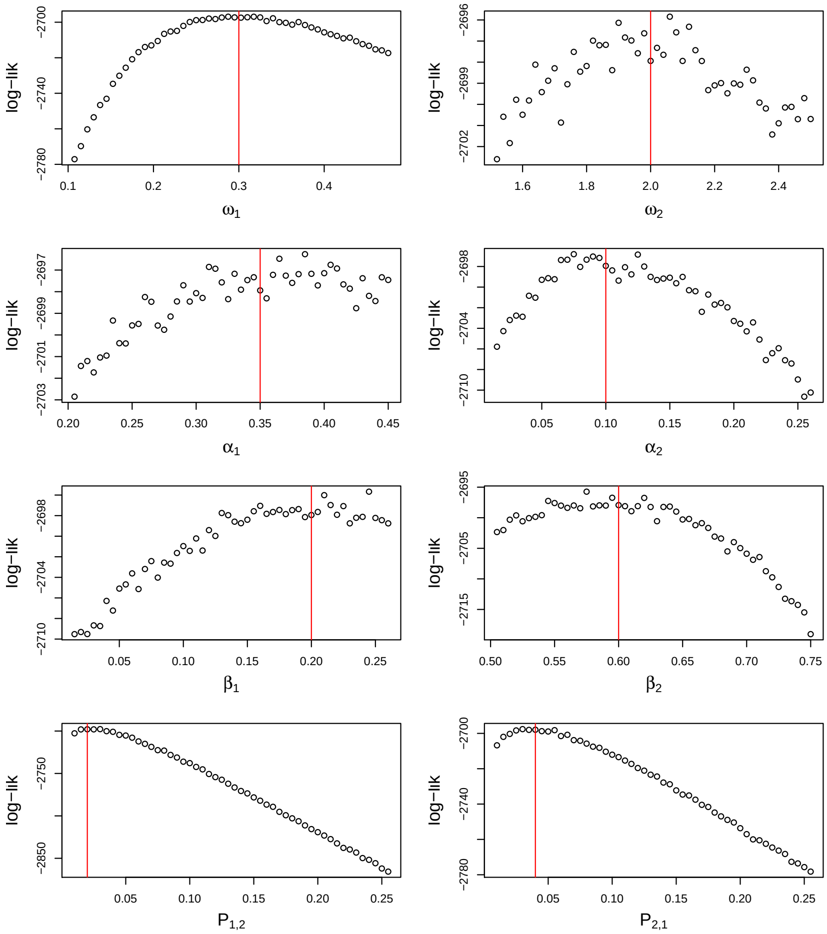

Figure 2: SMC approximated likelihood profile plots, using bootstrap resampling, for a partially observed MS-GARCH(1,1) series ($N = 2000$ with 20% of observations missing). The true parameters are indicated by the vertical lines.

In the case $t_{i+1} = t_i + 1$, set $\hat{\sigma}^{2(k)}_{t_{i+1},r} = \hat{\sigma}^{2(k)}_{t_i+1,r}$ and $R^{(k,r)}_{t_{i+1}} = r$ for $k = 1, \dots, 2^{q-2}, r = 1, 2$. However, in the case $n_{i+1} := t_{i+1} - t_i > 1$, we require simulation to move from $(\sigma^2_{t_i+1}, R_{t_i+1})$ to $(\sigma^2_{t_{i+1}}, R_{t_{i+1}})$. To this end, for each $r = 1, 2$, simulate 2^{q-2} regime trajectories $R_{t_i+2}, \dots, R_{t_{i+1}}$ given $R_{t_i+1} = r$, denote these $R^{(k,r)}_{t_i+2:t_{i+1}} := \{R^{(k,r)}_{t_i+2}, \dots, R^{(k,r)}_{t_{i+1}}\}$, $k = 1, \dots, 2^{q-2}$, along with 2^{q-2} sets of $n_{i+1} - 1$ innovations $z^{(k,r)}_{t_i+1:t_{i+1}-1} := \{z^{(k,r)}_{t_i+1}, \dots, z^{(k,r)}_{t_{i+1}-1}\}$. Given $\hat{\sigma}^{2(k)}_{t_i+1,r}$, $R^{(k,r)}_{t_i+2:t_{i+1}}$, $R^{(k,r)}_{t_i+1} = r$ and $z^{(k,r)}_{t_i+1:t_{i+1}-1}$ obtain $\hat{\sigma}^{2(k)}_{t_{i+1},r}$ recursively from

$$\hat{\sigma}^{2(k)}_{t_i+1+j,r} = h\Big(z^{(k,r)}_{t_i+j}\sqrt{\hat{\sigma}^{2(k)}_{t_i+j,r}}, \hat{\sigma}^{2(k)}_{t_i+j,r}, R^{(k,r)}_{t_i+j}, R^{(k,r)}_{t_i+1+j}\Big), \tag{46}$$

for $j = 1, \dots, n_{i+1} - 1$. We then construct

$$\hat{\mathbb{P}}(\sigma^2_{t_{i+1}}, R_{t_{i+1}} | Y_{t_1:t_i}) := \frac{1}{2^{q-2}} \sum_{r=1}^{2} \hat{\mathbb{P}}(R_{t_i+1} = r | Y_{t_1:t_i}) \sum_{k=1}^{2^{q-2}} \delta_{\{\hat{\sigma}^{2(k)}_{t_{i+1},r}, R^{(k,r)}_{t_{i+1}}\}}(\sigma^2_{t_{i+1}}, R_{t_{i+1}}), \tag{47}$$

yielding the approximate likelihood component

$$\hat{p}(Y_{t_{i+1}} | Y_{t_1:t_i}) := \frac{1}{2^{q-2}} \sum_{r=1}^{2} \hat{\mathbb{P}}(R_{t_i+1} = r | Y_{t_1:t_i}) \sum_{k=1}^{2^{q-2}} p(Y_{t_{i+1}} | \sigma^2_{t_{i+1}} = \hat{\sigma}^{2(k)}_{t_{i+1},r}, R_{t_{i+1}} = R^{(k,r)}_{t_{i+1}}). \tag{48}$$

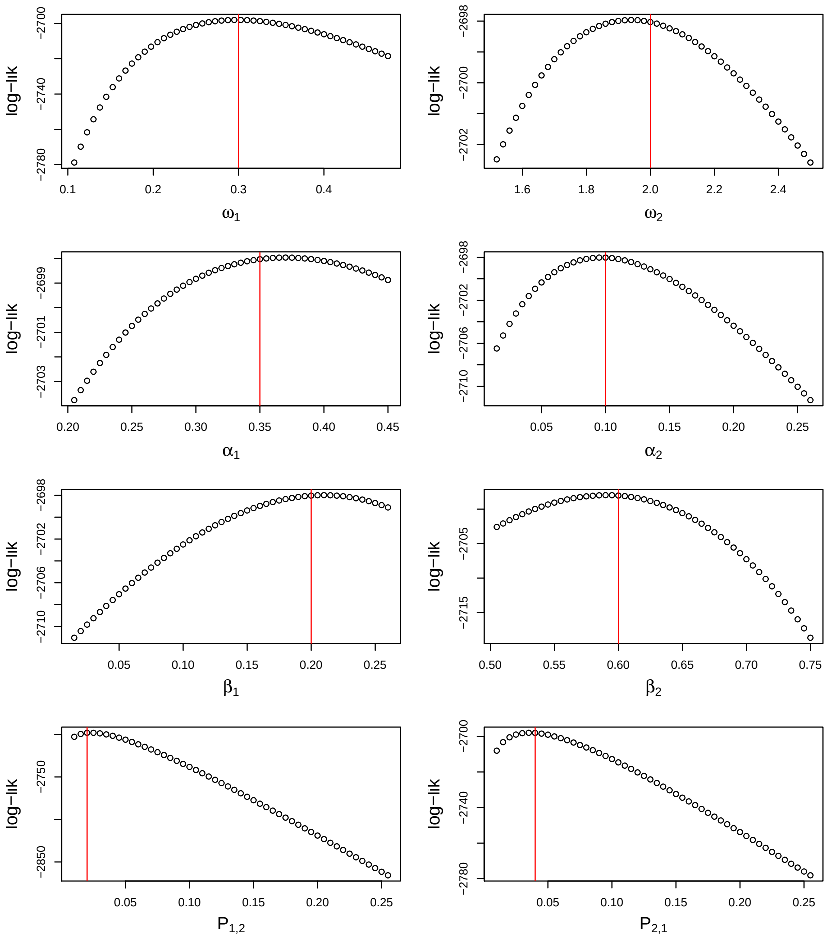

Figure 3: SMC approximated likelihood profile plots, using the resampling mechanism presented in Section 4, for a partially observed MS-GARCH(1,1) series (same data set as Figure 2). The true parameters are indicated by the vertical lines.

Then from (19) and

$$\mathbb{P}(\sigma^2_{t_{i+1}+1}|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1}) = \int_{X_{t_{i+1}}} \mathbb{P}(\sigma^2_{t_{i+1}+1}, R_{t_{i+1}+1}|X_{t_{i+1}}, Y_{t_{i+1}}) \frac{p(Y_{t_{i+1}}|X_{t_{i+1}})\mathbb{P}(X_{t_{i+1}}|Y_{t_1:t_i})}{\mathbb{P}(R_{t_{i+1}+1}|Y_{t_1:t_{i+1}})p(Y_{t_{i+1}}|Y_{t_1:t_i})}, \tag{49}$$

where $X_{t_{i+1}} := \{\sigma^2_{t_{i+1}}, R_{t_{i+1}}\}$, we arrived at

$$\tilde{\mathbb{P}}(\sigma^2_{t_{i+1}+1}|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1} = r) = \sum_{k=1}^{2^{q-1}} v^{(k)}_{i+1,r} \delta_{\{\sigma^{2(k)}_{t_{i+1}+1,r}\}}(\sigma^2_{t_{i+1}+1}) \tag{50}$$

and

$$\hat{\mathbb{P}}(R_{t_{i+1}+1} = r|Y_{t_1:t_{i+1}}) := \frac{1}{2^{q-2}} \sum_{k=1}^{2^{q-1}} \frac{\Lambda^{(k)}_{i+1,r}}{\hat{p}(Y_{t_{i+1}}|Y_{t_1:t_i})}, \tag{51}$$

where denoting $\varphi(k) := k - 2^{q-2}$, we have defined:

$$v^{(k)}_{i+1,r} := \frac{\Lambda^{(k)}_{i+1,r}}{2^{q-2}\hat{p}(Y_{t_{i+1}}|Y_{t_1:t_i})\hat{\mathbb{P}}(R_{t_{i+1}+1} = r|Y_{t_1:t_{i+1}})}, \tag{52}$$

$$\Lambda_{i+1,r}^{(k)} := \begin{cases} p(Y_{t_{i+1}}|\sigma_{t_{i+1}}^2 = \hat{\sigma}_{t_{i+1},1}^{2(k)}, R_{t_{i+1}} = R_{t_{i+1}}^{(k,1)}) P_{R_{t_{i+1}}^{(k,1)},r} \hat{\mathbb{P}}(R_{t_i+1} = 1|Y_{t_1:t_i}), & \text{for } k = 1,\dots,2^{q-2}, \\ p(Y_{t_{i+1}}|\sigma_{t_{i+1}}^2 = \hat{\sigma}_{t_{i+1},2}^{2(\varphi(k))}, R_{t_{i+1}} = R_{t_{i+1}}^{(\varphi(k),2)}) P_{R_{t_{i+1}}^{(\varphi(k),2)},r} \hat{\mathbb{P}}(R_{t_i+1} = 2|Y_{t_1:t_i}), & \text{for } k = 2^{q-2}+1,\dots,2^{q-1}, \end{cases} \tag{53}$$

and

$$\sigma_{t_{i+1}+1,r}^{2(k)} := \begin{cases} h(Y_{t_{i+1}}, \hat{\sigma}_{t_{i+1},1}^{2(k)}, R_{t_{i+1}}^{(k,1)}, r), & \text{for } k = 1,\dots,2^{q-2}, \\ h(Y_{t_{i+1}}, \hat{\sigma}_{t_{i+1},2}^{2(\varphi(k))}, R_{t_{i+1}}^{(\varphi(k),2)}, r), & \text{for } k = 2^{q-2}+1,\dots,2^{q-1}. \end{cases} \tag{54}$$

The required quantities $\tilde{\mathbb{P}}(\sigma_{t_{i+1}+1}^2|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1})$ and $\hat{\mathbb{P}}(R_{t_{i+1}+1}|Y_{t_1:t_{i+1}})$ to repeat the process for the next time step have been provided.

Now, the reason why this procedure is not guaranteed to be continuous in the transition probability parameters $P_{1,2}$ and $P_{2,1}$ is that, unlike $z_{t_i+1:t_{i+1}-1}^{(k,r)}$ which are invariant across any parameter changes, the noise components $R_{t_i+2:t_{i+1}}^{(k,r)}$ are driven by $P_{1,2}$ and $P_{2,1}$. The weights $v_{i+1,r}^{(k)}$, see (52)-(53), in addition to $\sigma_{t_{i+1}}^2$ have components that differ by R_{t_i+1} and $R_{t_{i+1}}$. While by design, R_{t_i+1} has been made invariant across parameter changes, the component $R_{t_{i+1}}$ can still vary as either $P_{1,2}$ or $P_{2,1}$ changes. Despite this, sufficiently smooth likelihood surfaces for numerical optimisation (as confirmed in the simulation studies of Section 5.2) are still able to be obtained by approximating the discrete distribution (50) by a continuous version obtained through partitioning by R_{t_i+1} into two step-wise functions and proceeding in the manner that (41) was obtained.

5. Simulation studies

Three sets of simulation studies are performed. The first two compare performance of the modified SMC procedure described in Section 3 against the generalised collapsing procedure (GCP) of [3]. While the third set of simulation studies is used to assess the extended capabilities of Section 4's SMC procedure in dealing with varying degrees of missingness in MS-GARCH series.

5.1. Full Y observation series

5.1.1. Simulation study A

We simulate 1000 data sets of 1500 observations of a two regime MS-GARCH(1,1) series with standard normal innovations and true parameters $\omega_1 = 0.3$, $\alpha_1 = 0.35$, $\beta_1 = 0.2$, $\omega_2 = 2$, $\alpha_2 = 0.1$, $\beta_2 = 0.6$, $P_{1,2} = 0.02$, $P_{2,1} = 0.04$, $\mu_1 = 0.06$ and $\mu_2 = -0.09$. This parameter set, inspired by empirical studies, was devised by [5]. Parameter estimates obtained utilising the SMC procedure and GCP approximation on these 1000 data sets are summarised in Table 1.

The SMC and GCP methods are seen to be on par in terms of estimation performance, having very similar root mean square error (RMSE) for all parameters. Both methods perform reasonably well, with average bias, as a percentage of true parameter value, being small for most of the parameters. Estimation performance of both methods appears stabilised by $q = 8$, with no substantial improvement in either average bias or standard deviation (SD) of the parameter estimates across the 1000 data sets by increasing to $q = 10$ or $q = 12$.

Numerical Hessians from the SMC approximated log-likelihood and GCP log-likelihood at their respective optimal parameter values are computed. The inverse of the negative of these numerical Hessians are used to obtain approximate standard errors. The summary statistic *Mean SE* is the average of these approximated standard errors across the 1000 data sets. Both methods appear to produce, for all parameters, Mean SE that are fairly close to observed SD.

Computational times using an Intel Xeon Platinum 8274 3.2 GHz processor for both SMC and GCP optimisations, both utilising integrated C++ and R implementations, are presented in Table 2. Both methods display optimisation times roughly linear in 2^q.

Table 1

Summary of SMC and GCP estimates for 1000† simulated data sets of 1500 observations from a MS-GARCH model with true parameters $\omega_1 = 0.3$, $\alpha_1 = 0.35$, $\beta_1 = 0.2$, $\omega_2 = 2$, $\alpha_2 = 0.1$, $\beta_2 = 0.6$, $P_{1,2} = 0.02$, $P_{2,1} = 0.04$, $\mu_1 = 0.06$ and $\mu_2 = -0.09$. **Mean**: the average of the individual estimates, **Bias**: the Mean minus the true parameter value, **SD**: the standard deviation of the individual estimates, **Mean SE**: the average of the standard errors obtained from the Hessian of the (approximated) log-likelihood, **RMSE**: the root mean squared error of the individual estimates compared to the true parameter value.

SMC	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
$q = 8$										
Mean	0.2991	0.3499	0.2009	2.4157	0.0991	0.5335	0.0205	0.0422	0.0612	-0.0938
Bias	-0.0009	-0.0001	0.0009	0.4157	-0.0009	-0.0665	0.0005	0.0022	0.0012	-0.0038
SD	0.0548	0.0719	0.0965	1.1937	0.0629	0.2125	0.0062	0.0133	0.0230	0.1155
RMSE	0.0548	0.0719	0.0965	1.2640	0.0629	0.2227	0.0062	0.0135	0.0230	0.1155
Mean SE	0.0549	0.0681	0.0940	1.2018	0.0637	0.2132	0.0062	0.0124	0.0237	0.1176
$q = 10$										
Mean	0.3001	0.3499	0.1996	2.4160	0.1003	0.5323	0.0205	0.0420	0.0612	-0.0937
Bias	0.0001	-0.0001	-0.0004	0.4160	0.0003	-0.0677	0.0005	0.0020	0.0012	-0.0037
SD	0.0551	0.0717	0.0973	1.2119	0.0632	0.2137	0.0062	0.0128	0.0229	0.1151
RMSE	0.0551	0.0717	0.0973	1.2813	0.0632	0.2242	0.0062	0.0130	0.0229	0.1152
Mean SE	0.0526	0.0682	0.0895	1.1504	0.0631	0.2033	0.0059	0.0119	0.0237	0.1160
$q = 12$										
Mean	0.3000	0.3501	0.1997	2.4010	0.1001	0.5348	0.0205	0.0419	0.0612	-0.0942
Bias	0.0000	0.0001	-0.0003	0.4010	0.0001	-0.0652	0.0005	0.0019	0.0012	-0.0042
SD	0.0546	0.0718	0.0969	1.2280	0.0631	0.2165	0.0060	0.0126	0.0229	0.1162
RMSE	0.0546	0.0718	0.0969	1.2918	0.0631	0.2261	0.0061	0.0128	0.0230	0.1163
Mean SE	0.0524	0.0685	0.0891	1.1251	0.0599	0.1995	0.0059	0.0119	0.0237	0.1149

GCP	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
$q = 8$										
Mean	0.3005	0.3502	0.1986	2.3734	0.1000	0.5387	0.0205	0.0419	0.0612	-0.0937
Bias	0.0005	0.0002	-0.0014	0.3734	0.0000	-0.0613	0.0005	0.0019	0.0012	-0.0037
SD	0.0547	0.0719	0.0971	1.1937	0.0632	0.2139	0.0060	0.0127	0.0228	0.1151
RMSE	0.0547	0.0719	0.0972	1.2507	0.0632	0.2226	0.0060	0.0128	0.0229	0.1152
Mean SE	0.0532	0.0686	0.0892	1.1174	0.0637	0.2044	0.0059	0.0124	0.0237	0.1160
$q = 10$										
Mean	0.3004	0.3502	0.1987	2.3730	0.0997	0.5390	0.0205	0.0419	0.0613	-0.0936
Bias	0.0004	0.0002	-0.0013	0.3730	-0.0003	-0.0610	0.0005	0.0019	0.0013	-0.0036
SD	0.0547	0.0719	0.0971	1.1928	0.0632	0.2141	0.0060	0.0127	0.0229	0.1151
RMSE	0.0547	0.0719	0.0971	1.2498	0.0632	0.2226	0.0060	0.0128	0.0229	0.1152
Mean SE	0.0525	0.0685	0.0889	1.1246	0.0609	0.2059	0.0059	0.0123	0.0237	0.1158
$q = 12$										
Mean	0.3004	0.3501	0.1988	2.3763	0.0997	0.5385	0.0205	0.0419	0.0613	-0.0936
Bias	0.0004	0.0001	-0.0012	0.3763	-0.0003	-0.0615	0.0005	0.0019	0.0013	-0.0036
SD	0.0546	0.0720	0.0971	1.1951	0.0632	0.2146	0.0060	0.0127	0.0229	0.1151
RMSE	0.0547	0.0720	0.0972	1.2530	0.0632	0.2232	0.0060	0.0128	0.0229	0.1152
Mean SE	0.0525	0.0705	0.0896	1.1980	0.0636	0.2114	0.0060	0.0119	0.0237	0.1156

†Results from two data sets for which both the SMC and GCP methods produced abnormally large estimates of $P_{2,1}$ were removed.

Table 2

Summary, across the 1000 simulated data sets, of computational times for the SMC and GCP optimisations utilising an Intel Xeon Platinum 8274 3.2 GHz processor. Likelihoods calculated in C++ with parameter optimisation performed using R's implementation of the Broyden - Fletcher - Goldfarb - Shanno (BFGS) algorithm with a relative tolerance set to 10^{-6}.

	SMC			GCP		
q	Avg. time (secs) to complete optimisation	Avg. number of function evaluations	Avg. number of gradient evaluations	Avg. time (secs) to complete optimisation	Avg. number of function evaluations	Avg. number of gradient evaluations
8	7.69	71.70	23.23	3.98	58.47	21.91
10	29.47	68.09	22.86	15.34	58.34	21.88
12	116.22	63.23	22.33	62.43	58.36	21.87

5.1.2. Simulation study B

The second parameter set we consider is $\omega_1 = \omega_2 = 0.1$, $\alpha_1 = \alpha_2 = 0.2$, $\beta_1 = 0.4$, $\beta_2 = 0.85$, $P_{1,2} = 0.01$, $P_{2,1} = 0.05$ and $\mu_1 = \mu_2 = 0$. In this way, the second regime is unstable (in that it has a persistence in excess of 1) and both GARCH volatility regimes are identical except for β which switches the system between stable and unstable persistences. Note that this parameter specification satisfies (6). What is challenging is the fact that only one parameter differentiates between the two GARCH volatility regimes. However, despite this difficulty, both the SMC and GCP methods, as seen in Table 3, still perform quite well estimating the parameters with fairly low average bias.

Table 3

Summary of SMC and GCP estimates for 1000 simulated data sets of 1500 observations from a MS-GARCH model with standard normal innovations and true parameters $\omega_1 = \omega_2 = 0.1$, $\alpha_1 = \alpha_2 = 0.2$, $\beta_1 = 0.4$, $\beta_2 = 0.85$, $P_{1,2} = 0.01$, $P_{2,1} = 0.05$ and $\mu_1 = \mu_2 = 0$. Refer to Table 1 for the definitions of the summary statistics Bias, SD etc.

SMC	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
$q = 8$										
Mean	0.0966	0.1995	0.4191	0.1362	0.1684	0.8665	0.0088	0.0503	0.0004	0.0056
Bias	-0.0034	-0.0005	0.0191	0.0362	-0.0316	0.0165	-0.0012	0.0003	0.0004	0.0056
SD	0.0164	0.0462	0.0741	0.0915	0.0768	0.0843	0.0043	0.0221	0.0137	0.1261
RMSE	0.0168	0.0462	0.0765	0.0984	0.0830	0.0859	0.0045	0.0221	0.0137	0.1262
Mean SE	0.0165	0.0424	0.0726	0.0696	0.0707	0.0776	0.0043	0.0212	0.0140	0.1117
$q = 10$										
Mean	0.0977	0.2003	0.4123	0.1364	0.1724	0.8637	0.0095	0.0520	0.0002	0.0049
Bias	-0.0023	0.0003	0.0123	0.0364	-0.0276	0.0137	-0.0005	0.0020	0.0002	0.0049
SD	0.0171	0.0467	0.0783	0.1031	0.0836	0.0873	0.0052	0.0238	0.0140	0.1281
RMSE	0.0173	0.0467	0.0792	0.1094	0.0880	0.0884	0.0052	0.0238	0.0140	0.1282
Mean SE	0.0168	0.0441	0.0746	0.0812	0.0806	0.0895	0.0049	0.0224	0.0141	0.1094

GCP	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
$q = 8$										
Mean	0.0992	0.1996	0.4047	0.1398	0.1714	0.8650	0.0105	0.0539	0.0001	0.0043
Bias	-0.0008	-0.0004	0.0047	0.0398	-0.0286	0.0150	0.0005	0.0039	0.0001	0.0043
SD	0.0194	0.0496	0.0896	0.1249	0.0914	0.0972	0.0077	0.0277	0.0140	0.1247
RMSE	0.0194	0.0496	0.0897	0.1311	0.0958	0.0983	0.0077	0.0279	0.0140	0.1247
Mean SE	0.0190	0.0467	0.0812	0.1205	0.0975	0.1086	0.0066	0.0252	0.0141	0.1128
$q = 10$										
Mean	0.0996	0.2004	0.4023	0.1369	0.1725	0.8647	0.0104	0.0535	-0.0001	0.0031
Bias	-0.0004	0.0004	0.0023	0.0369	-0.0275	0.0147	0.0004	0.0035	-0.0001	0.0031
SD	0.0200	0.0492	0.0917	0.1196	0.0920	0.0933	0.0083	0.0259	0.0146	0.1277
RMSE	0.0200	0.0492	0.0917	0.1252	0.0961	0.0945	0.0083	0.0261	0.0146	0.1277
Mean SE	0.0181	0.0469	0.0792	0.1143	0.1008	0.1059	0.0066	0.0249	0.0143	0.1127

5.2. Partial Y observation series

In this section we test the performance of the SMC estimation method against varying degrees of missingness. In order to study the impact of missingness, rather than total number of observations, on the performance of the method, we keep the effective sample size at 1500 observations. This allows us to study the impact of missingness in isolation from any impact on performance attributable to smaller sample sizes. Denote p_m as the missing percentage. For each missing scenario we simulate 1000 different data sets of a length $N = \lfloor \frac{1500}{1-p_m} \rfloor$ two regime MS-GARCH(1,1) series with standard normal innovations with true parameters $\omega_1 = 0.3$, $\alpha_1 = 0.35$, $\beta_1 = 0.2$, $\omega_2 = 2$, $\alpha_2 = 0.1$, $\beta_2 = 0.6$, $P_{1,2} = 0.02$, $P_{2,1} = 0.04$, $\mu_1 = 0.06$ and $\mu_2 = -0.09$. Then for each data set $\lfloor p_m \times N \rfloor$ different points between 1 and N (non-inclusive) were randomly (with equal chance of selection) deleted. Note that the configuration of missing observations is different between the 1000 data sets.

Missing percentages of 10%, 20%, 35% and 50% were analysed. Table 4 provides summary statistics for the parameters estimated from 1000 simulated data sets, at each missing percentage level, using the SMC method with values of $q = 8, 10$ and 12. In general, estimation performance in terms of RMSE is seen to worsen as missingness levels increase, driven by both increases in Bias and SD. However, we see that increasing q, for a given level of missingness, in most cases improves estimation performance, driven mainly by decreases in Bias but also sometimes with decreases in SD also. Even at the extreme case of 50% missing, in comparison with Table 1, estimation performance has not drastically deteriorated. Mean SE appears, in general, to be reasonably close to the observed SD for all parameters. Computational times using an Intel Xeon Platinum 8274 3.2 GHz processor for the SMC optimisations are presented in Table 5. The average time to complete an optimisation can be seen to be roughly linear in N as well as 2^q.

Table 4
SMC estimation with $q = 8, 10$ and 12 for partially observed MS-GARCH(1,1) series: Summary statistics from 1000 replications for each level of missingness. Refer to Table 1 for the definitions of the summary statistics Bias, SD etc.

Missing	q=8	ω_1	α_1	β_1	ω_2	α_2	β_2	$P^{\dagger}_{1,2}$	$P^{\dagger}_{2,1}$	μ_1	$\mu^{\ddagger}_2$
10%	Mean	0.2796	0.3335	0.2432	2.3346	0.1010	0.5509	0.0209	0.0439	0.0606	-0.0965
	Bias	-0.0204	-0.0165	0.0432	0.3346	0.0010	-0.0491	0.0009	0.0039	0.0006	-0.0065
	SD	0.0586	0.0859	0.1070	1.2098	0.0727	0.2001	0.0077	0.0170	0.0236	0.1269
	RMSE	0.0620	0.0875	0.1154	1.2552	0.0727	0.2060	0.0078	0.0174	0.0236	0.1271
	Mean SE	0.0534	0.0723	0.0953	1.1830	0.0628	0.2159	0.0057	0.0130	0.0241	0.1150
20%	Mean	0.2766	0.3264	0.2518	2.3624	0.1166	0.5356	0.0207	0.0433	0.0614‡	-0.0964
	Bias	-0.0234	-0.0236	0.0518	0.3624	0.0166	-0.0644	0.0007	0.0033	0.0014‡	-0.0064
	SD	0.0632	0.0938	0.1188	1.2080	0.0879	0.1947	0.0079	0.0180	0.0253‡	0.1279
	RMSE	0.0674	0.0967	0.1296	1.2612	0.0895	0.2051	0.0079	0.0183	0.0253‡	0.1281
	Mean SE	0.0558	0.0777	0.1004	1.2250	0.0717	0.2183	0.0058	0.0118	0.0247‡	0.1110
35%	Mean	0.2714	0.3234	0.2689	2.4420	0.1303	0.5133	0.0201	0.0424	0.0591	-0.0968
	Bias	-0.0286	-0.0266	0.0689	0.4420	0.0303	-0.0867	0.0001	0.0024	-0.0009	-0.0068
	SD	0.0675	0.1042	0.1277	1.2893	0.1000	0.2042	0.0078	0.0163	0.0273	0.1264
	RMSE	0.0733	0.1075	0.1451	1.3630	0.1045	0.2218	0.0078	0.0165	0.0273	0.1266
	Mean SE	0.0574	0.0801	0.1042	1.0840	0.0694	0.1899	0.0050	0.0111	0.0249	0.1094
50%	Mean	0.2683	0.3124	0.2814	2.3712	0.1388	0.5083	0.0194	0.0404	0.0601	-0.1007
	Bias	-0.0317	-0.0376	0.0814	0.3712	0.0388	-0.0917	-0.0006	0.0004	0.0001	-0.0107
	SD	0.0686	0.1195	0.1267	1.1804	0.0888	0.1853	0.0071	0.0148	0.0280	0.1308
	RMSE	0.0756	0.1253	0.1506	1.2374	0.0969	0.2067	0.0071	0.0148	0.0280	0.1312
	Mean SE	0.0606	0.0878	0.1107	1.0571	0.0739	0.1877	0.0047	0.0094	0.0254	0.1113

Missing	q=10	ω_1	α_1	β_1	ω_2	α_2	β_2	$P^{\dagger}_{1,2}$	$P^{\dagger}_{2,1}$	μ_1	$\mu^{\ddagger}_2$
10%	Mean	0.2856	0.3443	0.2281	2.3124	0.0967	0.5544	0.0204	0.0426	0.0604	-0.0953
	Bias	-0.0144	-0.0057	0.0281	0.3124	-0.0033	-0.0456	0.0004	0.0026	0.0004	-0.0053
	SD	0.0541	0.0885	0.0957	1.1561	0.0678	0.1955	0.0070	0.0143	0.0243	0.1261
	RMSE	0.0560	0.0887	0.0997	1.1976	0.0679	0.2007	0.0070	0.0145	0.0243	0.1262
	Mean SE	0.0537	0.0705	0.0942	1.2158	0.0604	0.2143	0.0058	0.0116	0.0238	0.1117
20%	Mean	0.2812	0.3371	0.2380	2.3612	0.1039	0.5443	0.0206	0.0431	0.0612	-0.0962
	Bias	-0.0188	-0.0129	0.0380	0.3612	0.0039	-0.0557	0.0006	0.0031	0.0012	-0.0062
	SD	0.0584	0.0925	0.1094	1.2263	0.0749	0.2021	0.0071	0.0146	0.0257	0.1265
	RMSE	0.0614	0.0934	0.1158	1.2784	0.0750	0.2096	0.0071	0.0149	0.0257	0.1267
	Mean SE	0.0561	0.0754	0.1016	1.2376	0.0653	0.2223	0.0060	0.0115	0.0244	0.1151
35%	Mean	0.2743	0.3290	0.2539	2.4317	0.1142	0.5262	0.0209	0.0434	0.0599	-0.1019
	Bias	-0.0257	-0.0210	0.0539	0.4317	0.0142	-0.0738	0.0009	0.0034	-0.0001	-0.0119
	SD	0.0633	0.1047	0.1191	1.2898	0.0896	0.2025	0.0073	0.0147	0.0267	0.1277
	RMSE	0.0683	0.1068	0.1307	1.3601	0.0907	0.2155	0.0074	0.0151	0.0267	0.1283
	Mean SE	0.0584	0.0811	0.1053	1.3793	0.0712	0.2440	0.0056	0.0116	0.0247	0.1198
50%	Mean	0.2722	0.3144	0.2678	2.4439	0.1182	0.5178	0.0213	0.0441	0.0603	-0.1030
	Bias	-0.0278	-0.0356	0.0678	0.4439	0.0182	-0.0822	0.0013	0.0041	0.0003	-0.0130
	SD	0.0690	0.1175	0.1323	1.2709	0.0909	0.2027	0.0076	0.0143	0.0288	0.1360
	RMSE	0.0744	0.1228	0.1487	1.3462	0.0927	0.2187	0.0077	0.0149	0.0288	0.1366
	Mean SE	0.0650	0.0890	0.1183	1.4293	0.0779	0.2512	0.0054	0.0110	0.0255	0.1205

Missing	q=12	ω_1	α_1	β_1	ω_2	α_2	β_2	$P^{\dagger}_{1,2}$	$P^{\dagger}_{2,1}$	μ_1	$\mu^{\ddagger}_2$
10%	Mean	0.2888	0.3493	0.2202	2.3462	0.0988	0.5516	0.0203	0.0427	0.0600	-0.0982
	Bias	-0.0112	-0.0007	0.0202	0.3462	-0.0012	-0.0484	0.0003	0.0027	0.0000	-0.0082
	SD	0.0592	0.0894	0.1042	1.2243	0.0726	0.2036	0.0072	0.0154	0.0257	0.1325
	RMSE	0.0603	0.0894	0.1061	1.2723	0.0726	0.2093	0.0072	0.0156	0.0257	0.1328
	Mean SE	0.0537	0.0706	0.0932	1.3718	0.0699	0.2564	0.0057	0.0130	0.0236	0.1164
20%	Mean	0.2880	0.3479	0.2212	2.3608	0.1022	0.5429	0.0204	0.0427	0.0611	-0.0961
	Bias	-0.0120	-0.0021	0.0212	0.3608	0.0022	-0.0571	0.0004	0.0027	0.0011	-0.0061
	SD	0.0601	0.0918	0.1093	1.1777	0.0861	0.2017	0.0069	0.0146	0.0265	0.1326
	RMSE	0.0613	0.0918	0.1113	1.2317	0.0861	0.2096	0.0069	0.0148	0.0265	0.1327
	Mean SE	0.0569	0.0757	0.0999	1.2642	0.0643	0.2268	0.0056	0.0117	0.0240	0.1160
35%	Mean	0.2811	0.3452	0.2342	2.4070	0.0991	0.5421	0.0206	0.0432	0.0594	-0.1024
	Bias	-0.0189	-0.0048	0.0342	0.4070	-0.0009	-0.0579	0.0006	0.0032	-0.0006	-0.0124
	SD	0.0654	0.1047	0.1189	1.2913	0.0791	0.2121	0.0070	0.0149	0.0273	0.1369
	RMSE	0.0681	0.1048	0.1237	1.3539	0.0791	0.2199	0.0070	0.0152	0.0273	0.1375
	Mean SE	0.0581	0.0777	0.1049	1.3311	0.0688	0.2363	0.0059	0.0119	0.0245	0.1195
50%	Mean	0.2770	0.3328	0.2490	2.4531	0.1048	0.5335	0.0209	0.0441	0.0601	-0.1063
	Bias	-0.0230	-0.0172	0.0490	0.4531	0.0048	-0.0665	0.0009	0.0041	0.0001	-0.0163
	SD	0.0718	0.1224	0.1311	1.3382	0.0955	0.2097	0.0073	0.0151	0.0293	0.1408
	RMSE	0.0754	0.1236	0.1400	1.4128	0.0956	0.2200	0.0074	0.0156	0.0293	0.1417
	Mean SE	0.0631	0.0854	0.1150	1.5177	0.0763	0.2724	0.0057	0.0113	0.0252	0.1180

‡Winsorised top/bottom 0.5% †Winsorised top 1.5%

Table 5
Summary, across the 1000 simulated data sets at each missing percentage level, of computational times for the SMC optimisations utilising an Intel Xeon Platinum 8274 3.2 GHz processor. Likelihoods calculated in C++ with parameter optimisation performed using R's implementation of the Broyden - Fletcher - Goldfarb - Shanno (BFGS) algorithm with a relative tolerance set to 10^{-6}.

		Avg. time (secs) to complete optimisation			Avg. number of function evaluations			Avg. number of gradient evaluations		
Missing	N	$q = 8$	$q = 10$	$q = 12$	$q = 8$	$q = 10$	$q = 12$	$q = 8$	$q = 10$	$q = 12$
10%	1667	9.52	32.89	114.22	143.89	116.75	82.23	31.78	28.42	24.15
20%	1875	12.41	46.17	157.68	164.11	137.55	96.64	33.57	31.32	26.34
35%	2308	17.25	64.37	231.46	177.05	146.28	109.27	33.97	31.96	28.22
50%	3000	23.25	88.32	316.56	176.78	151.14	112.98	33.03	31.70	28.40

6. Real data analysis

6.1. Full Y observation series

Daily percentage log returns on the S&P 500 price index from May 20, 1999 to April 25, 2011 (3000 observations) were obtained from *Yahoo! Finance*. This time series is displayed as the top plot in Figure 4. Returns were calculated on successive trading days, so no returns are considered missing. A zero mean, standard normal innovation, two regime MS-GARCH(1,1) model was fit to this series using both the SMC and GCP methods. The fits, along with estimated 95% confidence intervals, for both methods are provided in Table 6. The fits from both methods can be seen to be very similar and within each other's confidence intervals, thus we proceed with commentary based on the SMC estimates. Regime 1 has a stationary variance of $\frac{0.0123}{1-0.019-0.9541} = 0.457$ and stationary probability $\mathbb{P}(R_i = 1) = \frac{0.0011}{0.0011+0.0015} = 42.31\%$, while regime 2 has a stationary variance of 2.526 and stationary probability of 57.69%. Regime 2 is the more volatile regime. Both persistence parameters (α and β) are statistically significant for regime 2, while regime 1's confidence interval for α contains zero, possibly signalling a non-stochastic time dependent volatility process. The bottom plot of Figure 4 displays the time series of the SMC estimated probability of being in regime 1 given the current and past observations. Based on this plot, one could conjecture four regime changes. The process starts out in regime 2 (the more volatile regime) and this prevails till roughly late August 2003. Regime 1 then remains in effect until about late July 2007. Thereafter there is a switch back to regime 2, which remains prevalent until around perhaps October 2010. While indeed between May 2009 to May 2010 there is some ambiguity in which regime is in effect, there is a more definitive indication of a transition out of regime 2 by October 2010. Thereafter regime 1 continues until the end of the study period. The "dot-com bubble" which began in April 1997 and ended in June 2003 [cf. 23] is consistent with the prevalence of regime 2 during this time period. The second period that regime 2 is prevalent coincides with the "subprime mortgage crisis" of 2007-2009.

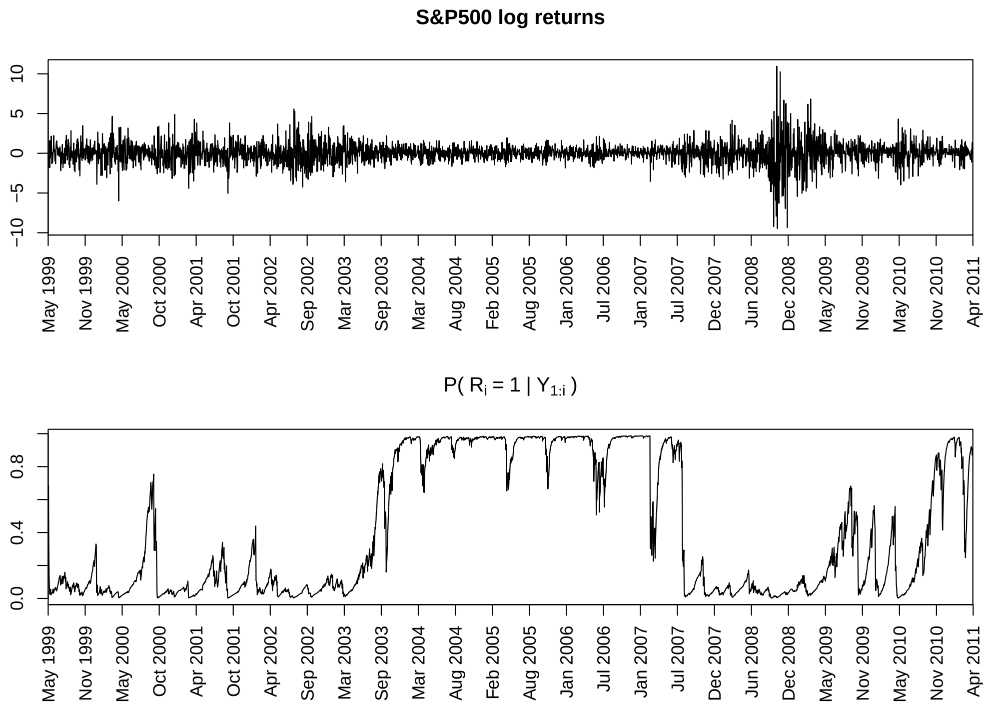

Figure 4: (Top:) Daily percentage log returns on the S&P 500 price index from May 20, 1999 to April 25, 2011. (Bottom:) SMC estimated probability of being in regime 1 given the current and past observations.

Table 6
The fits, along with estimated 95% confidence intervals, for the S&P 500 index returns data, using both the SMC and GCP methods. By **regime exit probability**, in reference to regime 1 for instance, this refers to the parameter $P_{1,2}$.

	Regime	ω	α	β	regime exit probability
SMC	1	0.0123 (0.0016, 0.0230)	0.0190 (-0.0021, 0.0401)	0.9541 (0.9170, 0.9911)	0.0015 (-0.0008, 0.0037)
	2	0.0538 (0.0250, 0.0827)	0.0941 (0.0664, 0.1218)	0.8846 (0.8535, 0.9156)	0.0011 (-0.0009, 0.0032)
GCP	1	0.0126 (0.0029, 0.0223)	0.0195 (-0.0028, 0.0418)	0.9533 (0.9173, 0.9893)	0.0014 (-0.0008, 0.0036)
	2	0.0531 (0.0256, 0.0805)	0.0938 (0.0667, 0.1210)	0.8845 (0.8536, 0.9154)	0.0011 (-0.0006, 0.0028)

6.2. Partial Y observation series

Henry Hub is a natural gas pipeline that is the pricing point for natural gas futures on the New York Mercantile Exchange (NYMEX). A time series for the Henry Hub natural gas spot price on trading days of the NYMEX from March 17, 2003 to March 24, 2008 was obtained from the Federal Reserve Bank of St. Louis's FRED database. Log returns calculated from successive NYMEX trading days are displayed as the top plot in Figure 5. Due to hurricane Rita, Henry Hub was forced to shut down from September 23, 2005 to October 6, 2005 resulting in an 11 day stretch of unavailable NYMEX trading day natural gas spot price returns. Thus we have $N = 1257$ and $n = 1246$. A zero mean, standard normal innovation, two regime MS-GARCH(1,1) model was fit to this returns series multiplied by 100 using the SMC method of Section 4. The fits obtained, along with estimated 95% confidence intervals, are provided in Table 7. Regime 1 has a stationary variance of 5.22 and stationary probability of 69.5% while regime 2 has a stationary variance of 184.82 and stationary probability of 30.5%. Regime 2 is the much more volatile regime. Both persistence parameters are statistically significant for regime 2, while regime 1's confidence interval for α contains zero. The bottom plot of Figure 5 displays the time series of the SMC estimated probability of being in regime 2 given the current (if available) and past observations. These probabilities over the gap of missing observations are indicated by the grey line segment in that plot. This plot appears to indicate 11 regime switches during the study period. The natural gas market tends to have two annual phases, an injection season from April to October and a withdrawal season from November to March [cf. 8]. Injections of gas into storage are made during periods of low demand and withdrawn from storage during periods of peak demand such as the cold winter months. During the injection season, demand for gas comes primarily from the energy consuming sector. As alternatives to gas for electricity generation are available, the demand for gas from these energy consumers is relatively price elastic. In contrast, during the withdrawal season, demand is primarily driven by heating needs from residential and commercial consumers. Presented with limited short term alternatives, demand for gas for heating purposes during this period is fairly price inelastic. While demand during the winter months is expected to be elevated, the exact level is uncertain, depending on realised weather conditions. The "*fixed supply and uncertain demand create substantial volatility in the months leading up to the end of the storage filling season, that is, September and October, and during the early winter months when cold spells are likely to be still ahead. During the rest of the year average volatility is substantially lower as lower gas demand during the spring and summer months alleviates pressure on supply*" [1]. These cycles of lower and higher periods of volatility appear to be identified in the conditional regime occurrence probabilities in Figure 5. The comparatively longer duration in the higher volatility regime beginning in the later half of 2005 may perhaps be due to disruptions to production and distribution channels brought about by the onset and aftermath of hurricanes Katrina and Rita.

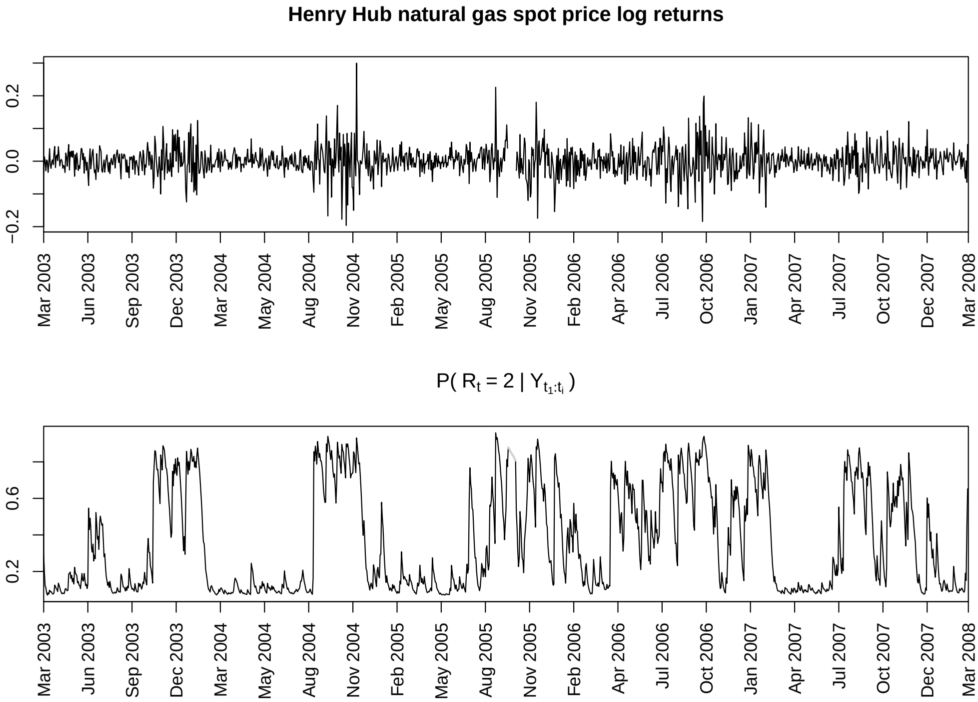

Figure 5: (Top:) Daily NYMEX trading day Henry Hub natural gas spot price log returns from March 17, 2003 to March 24, 2008. Due to hurricane Rita, Henry Hub was forced to shut down from September 23, 2005 to October 6, 2005 resulting in an 11 day stretch of unavailable NYMEX trading day natural gas spot price returns. (Bottom:) SMC estimated probability of being in regime 2 given the current (if available) and past observations. These probabilities over the gap of missing observations are indicated by the grey line segment.

Table 7
The fits, along with estimated 95% confidence intervals, for the Henry Hub returns data using the SMC method.

Regime	ω	α	β	regime exit probability
1	0.5288 (0.3512, 0.7065)	0.00013 (-0.00002, 0.00028)	0.8986 (0.8702, 0.9270)	0.0130 (0.0100, 0.0160)
2	0.8243 (0.6133, 1.0353)	0.02294 (0.00038, 0.04550)	0.9726 (0.9518, 0.9935)	0.0297 (0.0219, 0.0374)

7. Concluding remarks

In summary, this paper has shown it is possible to devise a SMC algorithm for MS-GARCH(1,1) models that produces approximated likelihood surfaces that can be numerically maximised, in turn providing a computationally feasible and reliable frequentist method for obtaining parameter estimates, flexible enough to be used even when one does not observe the full series. If provided full observation of the MS-GARCH(1,1) series, the GCP method of [3] is able to be utilised. However, unlike SMC, the GCP procedure is unequipped for dealing with partially observed MS-GARCH(1,1) data sets.

Extension of the GCP method to estimate MS-GARCH(p,q) for either $p > 1$ or $q > 1$ has so far not been developed and the feasibility of this requires further research. For the SMC framework proposed here to be extended to higher order MS-GARCH models would require development of a method for resampling in more than one continuous dimension conducive to smooth SMC likelihoods.

This paper, similarly to [3], has focused on the two regime MS-GARCH(1,1). However, as discussed at the end of Section 3.1.2, a continuous approximation that enables smooth resampling of the $\sigma_i^2 | Y_{1:i}, R_i$ when there are more than two regimes is feasible, and as such we do not perceive any difficulties, besides perhaps computational resources, inhibiting extension of the SMC method to more than two regimes.

The fair split $K_1 = K_2$ we have employed effectively importance samples the regimes from $\mathbb{Q}(R_i = 1) = \mathbb{Q}(R_i = 2) = 1/2$ rather than $\hat{\mathbb{P}}(R_i | Y_{1:i})$. Future research to seek alternative importance sampling distributions for the regime besides $\mathbb{Q}$, that could perhaps be a function of the time index i, but for reasons outlined in Procedure 2 (Section 3) need be independent of the model parameters θ, could be explored to improve simulation efficiency of the algorithm. One such idea is to proceed as we have with the fair split importance sampler but with a modest value of q to estimate a parameter set θ_0 along with $\hat{\mathbb{P}}(R_i | Y_{1:i}; \theta_0)$. Then one proceeds to importance sample the regimes from $\hat{\mathbb{P}}(R_i | Y_{1:i}; \theta_0)$ in estimation of θ. However, the additional computational cost of estimating the auxiliary model may well offset any simulation efficiency achieved by replacing $\mathbb{Q}$ with $\hat{\mathbb{P}}(R_i | Y_{1:i}; \theta_0)$.

Acknowledgements

The authors would like to express their gratitude and thank two anonymous referees and an Associate Editor for their help in improving the manuscript through their constructive comments. This research was supported by the Australian Government through the Australian Postgraduate Award. This research includes computations using the Linux computational cluster Katana supported by the Faculty of Science, UNSW Australia. The research also benefited from the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS enabled capability supported by the Australian Government.

Appendix

Algorithm 1 SMC procedure: from time step i to $i+1$

Assume at time i we have $\hat{\sigma}^{2(j)}_{i,r}$ *for* $j = 1, \dots, K/2, r = 1, 2$ *and* $\hat{\mathbb{P}}(R_i = r|Y_{1:i}), r = 1, 2$ *with* $K = 2^{q-1}$.
Set $\hat{p}(Y_{i+1}|Y_i) \leftarrow 0$
for j in $1 : K/2$ **do**
  for r in $1 : 2$ **do**
    for s in $1 : 2$ **do**
      $\sigma^{2(j+(r-1)K/2)}_{i+1,s} \leftarrow h(Y_i, \hat{\sigma}^{2(j)}_{i,r}, r, s)$
      $w^{(j+(r-1)K/2)}_{i+1,s} \leftarrow P_{r,s}\hat{\mathbb{P}}(R_i = r|Y_{1:i})p(Y_{i+1}|\sigma^2_{i+1} = \sigma^{2(j+(r-1)K/2)}_{i+1,s}, R_{i+1} = s)(2/K)$
      $pastState^{(j+(r-1)K/2)}_s = r$
      $\hat{p}(Y_{i+1}|Y_i) \leftarrow \hat{p}(Y_{i+1}|Y_i) + w^{(j+(r-1)K/2)}_{i+1,s}$
Simulate $U_1 < \dots < U_{K/2}$ *as i.i.d standard uniforms sorted in ascending order*
for r in $1 : 2$ **do**
  Set $\hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1}) \leftarrow 0$
  for j in $1 : K$ **do**
    $\grave{w}^{(j)}_{i+1,r} \leftarrow w^{(j)}_{i+1,r}/\hat{p}(Y_{i+1}|Y_i)$
    $\hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1}) \leftarrow \hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1}) + \grave{w}^{(j)}_{i+1,r}$
  Re-index the 3-tuple $\{\sigma^{2(j)}_{i+1,r}, \grave{w}^{(j)}_{i+1,r}, pastState^{(j)}_r\}_{j=1,\dots,K}$ *ascending in order of the first element* $\sigma^{2(j)}_{i+1,r}$.
  Set $cdf \leftarrow cdf_1 \leftarrow cdf_2 \leftarrow res \leftarrow p_1 \leftarrow p_2 \leftarrow z_1 \leftarrow z_2 \leftarrow idx_1 \leftarrow idx_2 \leftarrow 0$
  for j in $1 : K$ **do**
    $\grave{w}^{(j)}_{i+1,r} \leftarrow \grave{w}^{(j)}_{i+1,r}/\hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1})$
    for s in $1 : 2$ **do**
      $\bar{s} \leftarrow mod(s, 2) + 1$
      if $pastState^{(j)}_r = s$ **then**
        $cdf_s \leftarrow cdf_s + \frac{1}{2}(\grave{w}^{(j)}_{i+1,r} + p_s)$
        $p_s \leftarrow \grave{w}^{(j)}_{i+1,r}$
        $z_s \leftarrow \sigma^{2(j)}_{i+1,r}$
        $idx_s \leftarrow idx_s + 1$
        if $idx_{\bar{s}} = K/2$ **then**
          $cdf_{\bar{s}} \leftarrow cdf_{\bar{s}} + \frac{1}{2}p_{\bar{s}}$
          $p_{\bar{s}} \leftarrow 0$
        else if $idx_{\bar{s}} > 0$ **then**
          $\grave{k} \leftarrow \min\{k \in \mathbb{Z}_+ : \sigma^{2(k)}_{i+1,r} > z_{\bar{s}}$ and $pastState^{(k)}_r = \bar{s}\}$
          $res \leftarrow \frac{1}{2}(p_{\bar{s}} + \grave{w}^{(\grave{k})}_{i+1,r}) \times (\sigma^{2(j)}_{i+1,r} - z_{\bar{s}})/(\sigma^{2(\grave{k})}_{i+1,r} - z_{\bar{s}})$
    $\gamma_{j-1} \leftarrow cdf_1 + cdf_2 + res - cdf$
    $cdf \leftarrow cdf + \gamma_{j-1}$
    $res \leftarrow 0$
  $\gamma_K \leftarrow 1 - cdf$
  $\tau \leftarrow \gamma_0$
  $\nu \leftarrow 1$
  for j in $1 : K/2$ **do**
    if $U_j < \gamma_0$ **then**
      $\hat{\sigma}^{2(j)}_{i+1,r} = \sigma^{2(1)}_{i+1,r}$
    else if $U_j > 1 - \gamma_K$ **then**
      $\hat{\sigma}^{2(j)}_{i+1,r} = \sigma^{2(K)}_{i+1,r}$
    else
      while $U_j > \tau$ **do**
        $\tau \leftarrow \tau + \gamma_\nu$
        $\nu \leftarrow \nu + 1$
      $\hat{\sigma}^{2(j)}_{i+1,r} \leftarrow \sigma^{2(\nu-1)}_{i+1,r} + \frac{U_j - (\tau - \gamma_{\nu-1})}{\gamma_{\nu-1}}(\sigma^{2(\nu)}_{i+1,r} - \sigma^{2(\nu-1)}_{i+1,r})$

References

[1] Alterman, S. (2012). Natural gas price volatility in the UK and North America. *Oxford Institute for Energy Studies*, NG, 60.

[2] Augustyniak, M. (2014). Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics and Data Analysis*, 76:61–75.

[3] Augustyniak, M., Boudreault, M., and Morales, M. (2017). Maximum Likelihood Estimation of the Markov-Switching GARCH Model Based on a General Collapsing Procedure. *Methodology and Computing in Applied Probability*.

[4] Bauwens, L., Dufays, A., and Rombouts, J. (2014). Marginal likelihood for Markov-switching and change-point GARCH. *Journal of Econometrics*, 178:508–522.

[5] Bauwens, L., Preminger, A., and Rombouts, J. (2010). Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13:218–244.

[6] Bildirici, M. and Ersin, O. (2014). Modeling Markov Switching ARMA-GARCH Neural Networks Models and an Application to Forecasting Stock Returns. *The Scientific World Journal*.

[7] Billio, M., Casarin, R., and Osuntuyi, A. (2016). Efficient Gibbs sampling for Markov switching GARCH models. *Computational Statistics and Data Analysis*, 100:37–57.

[8] Bluegold Research (2017). The Logic Of This Injection Season: What You Need To Know. `https://seekingalpha.com/article/4060690-logic-injection-season-need-know`. [Online; accessed 19-August-2018].

[9] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

[10] Bondon, P. and Bahamonde, N. (2012). Least squares estimation of ARCH models with missing observations. *Journal of Time Series Analysis*, 33(6):880–891.

[11] Cripps, E. and Dunsmuir, W. (2003). Modeling the Variability of Sydney Harbor Wind Measurements. *Journal of Applied Meteorology*, 42(8):1131–1138.

[12] Del Moral, P. (1996). Non Linear Filtering: Interacting Particle Solution. *Markov Processes and Related Fields*, 2(4):555–580.

[13] Dieobold, F. (1986). Modeling The Persistence Of Conditional Variances: A Comment. *Econometric Reviews*, 5(1):51–56.

[14] Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business and Economic Statistics*, 15(1):26–34.

[15] Elliott, R., Lau, J., Miao, H., and Siu, T. (2012). Viterbi-based estimation for Markov switching GARCH models. *Applied Mathematical Finance*, 19(3):219–231.

[16] Francq, C., Roussignol, M., and Zakoian, J. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22:197–220.

[17] Francq, C. and Zakoian, J. (2008). Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference. *Computational Statistics and Data Analysis*, 52(6):3027–3046.

[18] Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F*, 140:107–113.

[19] Gray, S. (1996). Modeling the conditional distribution of interest-rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62.

[20] Haas, M., Mittnik, S., and Paolella, M. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

[21] Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

[22] Hamilton, J. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593.

[23] Ironman (2010). Here's Why The Dot Com Bubble Began And Why It Popped. https://www.businessinsider.com.au/heres-why-the-dot-com-bubble-began-and-why-it-popped-2010-12?r=US&IR=T. [Online; accessed 19-August-2018].

[24] Karelmo, K. (2010). Interest Rate Term Structure Modeling in the Presence of Missing Data. Working paper.

[25] Klaassen, F. (2002). Improving GARCH Volatility Forecasts with Regime-Switching GARCH. *Empirical Economics*, 27:363–394.

[26] Lamoureux, C. and Lastrapes, W. (1990). Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. *Journal of Finance*, 45:221–229.

[27] Mikosch, T. and Starica, C. (2004). Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *Review of Economics and Statistics*, 86:378–390.

[28] Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334.

[29] Perez-Quiros, G. and Timmermann, A. (2001). Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities. *Journal of Econometrics*, 103(1-2):259–306.

[30] Pitt, M. and Malik, S. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, 165:190–209.

[31] Schwert, G. (1989). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, 44(5):1115–1153.