Estimation of observation driven volatility models with intractable likelihoods - a sequential Monte Carlo approach

Damien Chee Ho Wee

School of Mathematics and Statistics

University of New South Wales

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

October 2017

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Wee

First name: Damien Other name/s: Chee Ho

Abbreviation for degree as given in the University calendar: PhD

School: Mathematics and Statistics Faculty: Science

Title: Estimation of observation driven volatility models with intractable likelihoods - a sequential Monte Carlo approach

Abstract 350 words maximum: (PLEASE TYPE)

This thesis focusses on application as well as modifications of sequential Monte Carlo (SMC) utilising the smooth resampling procedure of Pitt and Malik [2011] (smooth bootstrap) as a statistically and computationally efficient method for parameter estimation of discrete and continuous time stochastic processes that have intractable likelihoods; arising in the modelling of volatility, primarily in financial markets but also in other fields.

For the models and applications we consider, the likelihoods are intractable arising either from observations of a discrete time process being missing or temporally aggregated, or from discrete observation of a continuous time process. The methods are developed for the discrete time GARCH(1,1) model (Bollerslev [1986]) for conditional heteroscedasticity when there are missing observations or when observation is only through temporal aggregates of the underlying process. The methods to be presented can be generalised to many variants of the GARCH model including the EGARCH(1,1) (Nelson [1991]) and GJR-GARCH(1,1) (Glosten et al. [1992]) for instance. More challenging are the continuous time GARCH model (COGARCH) of Kluppelberg et al. [2004] and the Markov switching (MS-)GARCH model (cf. Bauwens et al. [2010]). The COGARCH process evolves in continuous time but is observed at, typically, irregular time intervals and as a result many events in the underlying process are not directly observed, while the MS-GARCH model allows model parameters to evolve over time according to an unobserved regime process. The same fundamental methods are developed for these two more complicated variants also.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

…… …… ……

Signature Witness Signature Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

ORIGINALITY STATEMENT

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Signed …….....

Date …….....

Abstract

This thesis focusses on application as well as modifications of sequential Monte Carlo (SMC) utilising the smooth resampling procedure of Pitt and Malik [2011] (smooth bootstrap) as a statistically and computationally efficient method for parameter estimation of discrete and continuous time stochastic processes that have intractable likelihoods; arising in the modelling of volatility, primarily in financial markets but also in other fields. Intractability of the likelihoods of the models we consider arise due to either missing observations, temporally aggregated observations, or from discrete observation of a continuous time process.

The first class of models we consider is the famous GARCH(1,1) model of Bollerslev [1986]. Estimation of this class of model is straightforward given the full observation series, however due to the inherent observation driven volatility mechanism, the marginal likelihood of a subset of the full observation series, that is when some observations are missing, is in most cases of interest intractable. Also intractable is the likelihood from observations through temporally aggregated elements. We illustrate application of the smooth bootstrap to provide parameter estimates in both these cases via simulated maximum likelihood. While for the case of partial observation of the GARCH series, application of the smooth bootstrap filter is straightforward, for the case of observation through temporally aggregated elements, identification of a specific state space representation is required for ease of implementation.

The second class of model considered is the COGARCH(1,1) model of Kluppelberg et al. [2004], a continuous time counterpart to the discrete time GARCH(1,1) model, invented for use in modelling increasingly available high frequency financial data sets as well as utilised in the well established continuous time option pricing framework. The likelihood of discrete observations of the continuous time COGARCH process is an intractable high dimensional integral. We identify a specific state space representation for this model allowing for ease of implementation of the smooth bootstrap to provide parameter estimates through simulated maximum likelihood.

The third class of models we consider are the relatively new Markov switching (MS-)GARCH(1,1) models (cf. Bauwens et al. [2010]), a concept to allow for structural changes in the otherwise fixed GARCH parameters. These models assume the world can exist in one of a finite set of regimes whereby regime evolution is through a Markov chain. The data generating process has the same GARCH model structure across all regimes but that each regime puts in effect its own set of GARCH parameters. Even provided the full observation series, 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. We extend the smooth resampling procedure of Pitt and Malik [2011], which is limited to hidden Markov models where the state equation can be summarised by a single state variable during the resampling, to provide a means for parameter estimation for these MS-GARCH(1,1) models, in which one needs two variables, volatility (which has a continuous support) and regime (which has a finite support), to evolve the hidden state process.

Acknowledgements

I am truly grateful to:

- my two supervisors Prof. William Dunsmuir and Dr. Feng Chen, for your patient guidance, encouragement and generosity with your time and expertise throughout my PhD studies.
- my parents, for the never-ending love, care and support you have bestowed upon me my entire life.
- my close friends Hugh Macfarlane, Lisa Teh, Tejay Lovelock and Richard McCarthy, for always keeping my spirits high.
- God, for the wonders of life.

Contents

Chapter 1

Introduction

This thesis focusses on the development of statistically and computationally efficient methods for accurate estimation of intractable likelihoods for discrete and continuous time stochastic processes that arise in the modelling of volatility, primarily in financial markets but also in other fields. With these approximated likelihoods, the otherwise infeasible task of obtaining parameters estimates through maximum likelihood estimation is then possible.

For the models and applications we consider, the likelihoods are intractable arising either from observations of a discrete time process being missing or temporally aggregated, or from discrete observation of a continuous time process. The methods are developed for the discrete time GARCH(1,1) model (Bollerslev [1986]) for conditional heteroscedasticity, when there are missing observations or when observation is only through temporal aggregates of the underlying process. The methods to be presented can be generalised to many variants of the GARCH model including the EGARCH(1,1) (Nelson [1991]) and GJR-GARCH(1,1) (Glosten et al. [1992]) for instance. More challenging are the continuous time GARCH model (COGARCH) of Kluppelberg et al. [2004] and the Markov switching (MS-)GARCH model (cf. Bauwens et al. [2010]). The COGARCH process evolves in continuous time but is observed at, typically, irregular time intervals and as a result many events in the underlying process are not directly observed, while the MS-GARCH model

allows model parameters to evolve over time according to an unobserved regime process. The same fundamental methods are developed for these two more complicated variants also.

Before reviewing the basic details of each model considered in the thesis, we outline the main technical aspects of the approach used to obtain a simulated estimate of the likelihood which is continuous in the model parameters and hence capable of being optimised efficiently and for which a suitable Hessian can be estimated to be used for calculation of standard errors of parameter estimates.

Here we introduce the key methods of this thesis initially through consideration of a simple non-trivial example, that of the GARCH(1,1) discrete time process for stochastic volatility. Since its introduction by Bollerslev [1986] 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. Cripps and Dunsmuir [2003]). GARCH models have a benefit over competing stochastic volatility models in that using a single source of randomness (the latter requires two sources) the model is able to capture a number of stylised features associated with financial returns data such leptokurtosis, clustering of large moves and small moves (volatility persistence) and mean reverting volatility (cf. Engle and Patton [2001]). Formally, the GARCH(1,1) model is the following system of recurrence equations

$$y_i = \sigma_i z_i \tag{1.1}$$

$$\sigma_i^2 = \omega + \alpha y_{i-1}^2 + \beta \sigma_{i-1}^2 \tag{1.2}$$

for $i = 1, \ldots, N$ whereby

- $(y_i)_{i=1,\ldots,N}$ is termed the observation series,
- $(z_i)_{i=1,\ldots,N}$ is an independently and identically distributed (i.i.d) noise series with density h and finite second moment,

- $\omega > 0, \alpha \geq 0$ and $\beta \geq 0$ are model parameters.

It was proven in Nelson [1990] that (1.1)-(1.2) has a unique strictly stationary and ergodic solution if and only if

$$\mathbb{E}(\log(\alpha z_1^2 + \beta)) < 0. \tag{1.3}$$

As is seen from the volatility equation (1.2) the variance of observation y_i is dependent on the full history of observations prior $(y_j)_{j=1,\dots,i-1}$, that is $\sigma_i^2 = \upsilon(y_{i-1}, \dots, y_1)$ for some function υ. Thus, provided the full observation series $(y_i)_{i=1,\dots,N}$ one can determine the full volatility series $(\sigma_i)_{i=1,\dots,N}$, then assuming the density h of the noise sequence and parameters specifying it are known, the likelihood of $(y_i)_{i=1,\dots,N}$, conditional on initial values y_0 and σ_0^2 can be easily computed as

$$p_y(y_1, \dots, y_N; \theta) = \prod_{i=1}^{N} \frac{h(y_i/\sigma_i)}{\sigma_i}, \tag{1.4}$$

where $\theta = (\omega, \alpha, \beta)$. Utilising (1.4), maximum likelihood estimation of the parameters θ is straightforward.

Now, if an observation y_j is omitted, this disrupts knowledge of future volatilities $(\sigma_i)_{i=j+1,\dots,N}$. Assume that we observe a GARCH(1,1) series on a set of $n < N$ time points $t_1 = 1 < t_2 < \ldots < t_n = N$, where $t_j \in \{2, \dots, N-1\}$ for $j = 2, \dots, n-1$. Denote by $\mathbb{M}$ the complement of $\mathbb{O} := \{t_1, \dots, t_n\}$ in the set $\mathbb{T} := \{1, 2, \dots, N\}$, $y_{\mathbb{O}}$ the collection of observed data and $y_{\mathbb{M}}$ the collection of missing observations. Define $Y_i = y_{t_i}$ for $i = 1, \dots, n$ and let $Y = (Y_1, \dots, Y_n)$ be the vector of observations. The marginal likelihood of observations Y is then given by

$$p_Y(Y_1, \dots, Y_n; \theta) = \int_{\mathbb{R}^{N-n}} p_y(y_1, \dots, y_N; \theta) dy_{\mathbb{M}} \tag{1.5}$$

that is, the full information density defined in (1.4) integrated over the missing observations. The likelihood defined in (1.5) requires evaluation of integrals of dimension $N - n$, which for even a modest amount of missing observations is intractable, making maximum likelihood infeasible.

Somewhat suprisingly, there is not a lot of literature on the topic of how to estimate GARCH models when one has gaps in the observation series. This is particularly surprising given how important GARCH models are in practice and that it is not unusual for real world data sets to contain missing observations. For instance, most economic or financial time series can be considered to have missing measurements on holidays or weekends, "*although 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*" -Bondon and Bahamonde [2012].

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. Figure 1.1 displays a time series of log yield changes, across trading days of the Korean securities exchange (KRX), of a Korean treasury bond (4.25 coupon, maturity 9th October 2014) for the period 10th September 2013 to 29th August 2014.

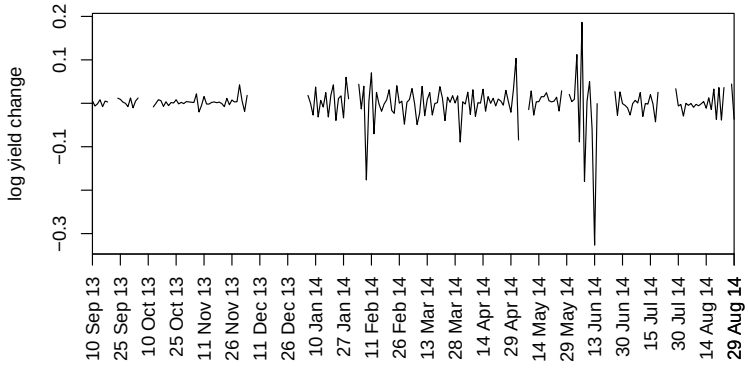


Figure 1.1: Time series of log yield changes of a Korean treasury bond (4.25 coupon, maturity 9th October 2014) for the period 10th September 2013 to 29th August 2014. This data was obtained from Bloomberg.

While daily closing prices of stock price data on trading days are rarely missing, as we move to a higher frequency of observation the data acquisition process is more susceptible to transmission failures. Figure 1.2 displays the time series of high frequency one-minute log price changes of the stock Boeing Co. (BA), for each trading minute of the New York Stock Exchange (NYSE) on the 2nd of January 2015.

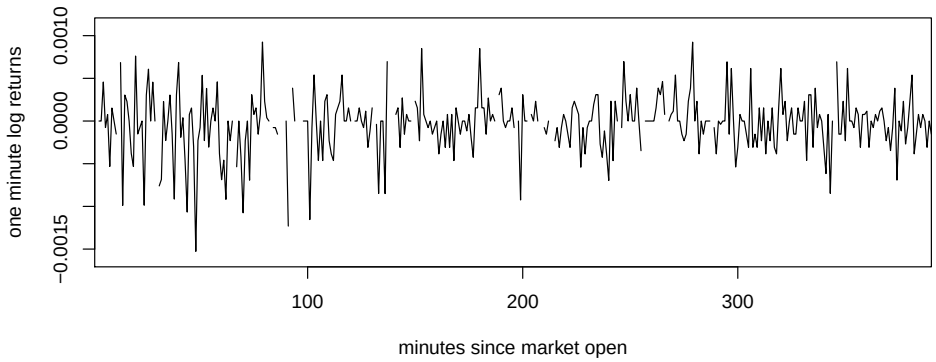


Figure 1.2: Time series of one-minute log price changes of the stock Boeing Co. (BA), for each trading minute of the New York Stock Exchange (NYSE) on the 2nd of January 2015. This data was obtained from https://www.finam.ru/.

Outside of financial applications, the GARCH model has been applied to model wind speed. Mechanical failures of measurement equipment can indeed lead to gaps in high frequency wind data. The top plot in Figure 1.3 is a time series of 30 minute north-south wind speed measurements taken at Fort Denison, Sydney from 00:00 hours 1st December 1999 to 00:00 hours 13th April 2000. The bottom plot in Figure 1.3 displays the residuals after fitting an Autoregressive model of order 3 to this wind data.

Furthermore, consider regression models where one wishes to account for correlation in residuals by modelling them as a GARCH series. It can happen that some recordings of the response or predictors are missing leading to unobserved residuals. For instance, in finance, in line with arbitrage pricing theory (APT) (see Ross [1976]), it is commonplace to study linearly the relationship between

an asset's return and a number of macro-economic/financial factors or theoretical indices. The return for a domestic asset could have a foreign asset as a factor and due to time-zone and different public holiday schedules the response and predictor series do not line up perfectly. Even domestic factors can be missing; for instance, the New York Stock Exchange (NYSE) and NASDAQ are open on Columbus day and Veteran's day while the US bond market is not, thus factors derived from the US bond market may not be available for those days. Furthermore, a complete record of factors such as historical accounting ratios may be difficult to find.

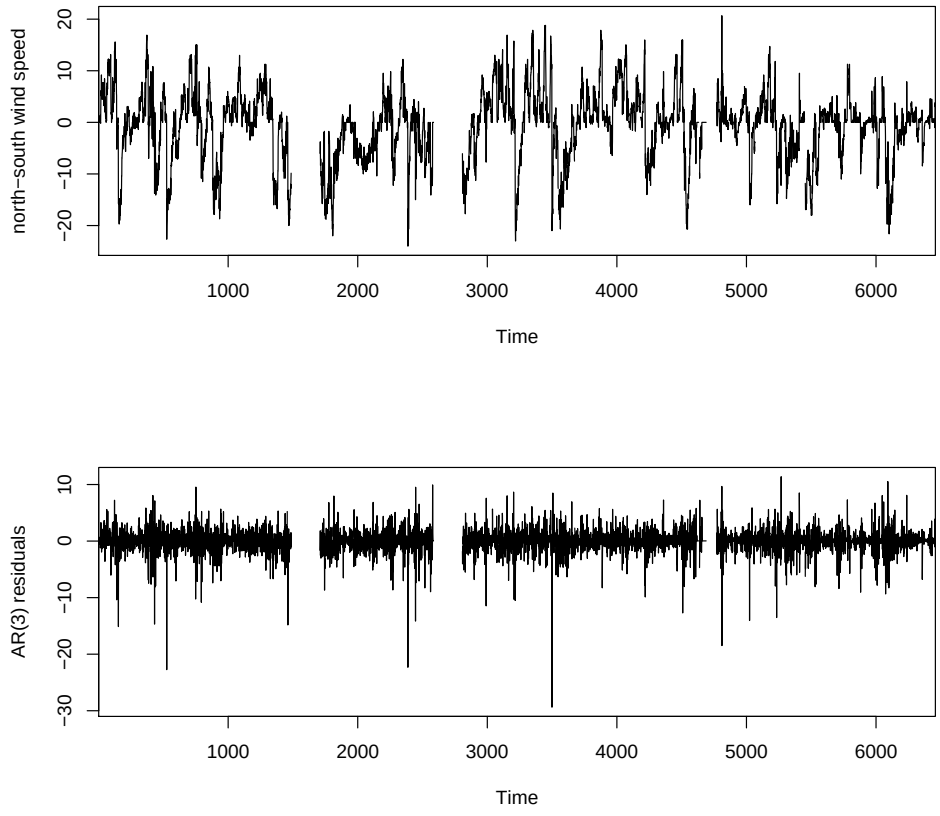


Figure 1.3: (Top:) Time series of 30 minute north-south wind speed measurements taken at Fort Denison, Sydney from 00:00 hours 1st December 1999 to 00:00 hours 13th April 2000. (Bottom:) Residuals after fitting an Autoregressive model of order 3 to the wind data.

Various approaches to approximate the integrand (1.5) could be considered such as series expansion (e.g. the Laplace approximation), Monte Carlo evaluation of the likelihood (e.g. importance sampling), or, a combination of both – see Davis and Rodriguez-Yam [2005] for example for a discussion of these concepts.

However, the integrand in (1.5) cannot be guaranteed to be log-concave for all sample paths and configurations of sampling times, thus inhibiting the use of Laplace approximations. While importance sampling, without a good approximating density, is prone to uneven coverage of samples with some samples having very low probability and only a few samples having substantial probability mass (Poole and Mackworth [2010]). For instance, if one were to simply simulate $z_{i,k}$, $i \in \mathbb{M}$, $k = 1, \ldots, K$ i.i.d random variables with density h and define

$$\tilde{y}_i^{(k)} = \begin{cases} y_i & \text{if } i \in \mathbb{O} \\ z_{i,k}\sqrt{\upsilon(\tilde{y}_{i-1}^{(k)}, \ldots, \tilde{y}_1^{(k)})} & \text{if } i \in \mathbb{M}, \end{cases} \tag{1.6}$$

indeed

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \frac{p_y(\tilde{y}_1^{(k)}, \ldots, \tilde{y}_N^{(k)}; \theta)}{\prod_{i\in M} h(z_{i,k})/\sqrt{\upsilon(\tilde{y}_{i-1}^{(k)}, \ldots, \tilde{y}_1^{(k)})}} = p_Y(Y_1, \ldots, Y_n; \theta), \tag{1.7}$$

however K will have to be quite large to yield a reasonable estimate of $p_Y(Y_1, \ldots, Y_n ; \theta)$, as a reckless amount of simulation effort will be wasted on many k for which $p_y(\tilde{y}_1^{(k)}, \ldots, \tilde{y}_N^{(k)}; \theta) \approx 0$, for the bulk of probability mass lies within only a small region of the vast simulation space. Furthermore, due to the path dependency, one component $z_{i,k}$ of $(z_{i,j})_{j=1,\ldots K}$ is enough to be the difference in whether $\tilde{y}_1^{(k)}, \ldots, \tilde{y}_N^{(k)}$ falls in a region of meaningful probability mass or not.

A means to overcome the disproportionate allocation of probability mass among simulated samples for hidden state space models is that of sequential Monte Carlo (SMC), a.k.a particle filters, a Monte Carlo technique first introduced in Gordon

et al. [1993] for which theoretical convergence results were established by Del Moral [1996]. SMC utilises a bootstrap resampling mechanism that leads to more simulated samples in regions of substantial probability mass.

We assume throughout that the process of interest is (y_t) where t denotes discrete time for the GARCH type models or denotes continuous time for the COGARCH model. Observations are recorded at times $t_1 < t_2 < \ldots < t_n$ which are not necessarily equally spaced. The observations could represent the values of y itself or aggregations of it over the intervals between successive t_i. Regardless, the available observations are denoted Y_i, for $i = 1, \ldots, n$. For all models considered herein a discrete time Markov process $(X_i)_{i=1,\ldots,n}$ can be defined to give an unobserved state process of the general form

$$X_i = g(X_{i-1}, Y_{i-1}, W_i), \tag{1.8}$$

in terms of which the observations $(Y_i)_{i=1,\ldots,n}$ are defined as

$$Y_i = f(X_i, Z_i), \tag{1.9}$$

where f and g are suitably defined functions depending on the particular model being considered, $(Z_i)_{i=1,\ldots,n}$ is a serially independent noise series and $(W_i)_{i=1,\ldots,n}$ is another serially independent noise series that is independent of $(Z_i)_{i=1,\ldots,n}$. Note that, in this specification, the Y_i's are serially correlated, but conditionally independent given the hidden state X_i. The hidden state X_i can be seen to be Markovian as $X_i = g\Big(X_{i-1}, f(X_{i-1}, Z_{i-1}), W_i\Big)$.

Utilising SMC one is able to obtain, through simulation, a consistent and unbiased estimate of the likelihood $p_Y(Y_1, \ldots, Y_n; \theta)$ (cf. Del Moral [1996]). However, while the SMC's bootstrap resampling leads to more samples in regions of substantial probability mass; it comes with the drawback that, even fixing randomness, the estimated likelihood surface is unlikely smooth in θ and thus generally not

amenable to numerical optimisers.

The bootstrap resampling mechanism is equivalent to resampling according to a step-wise cumulative distribution function (CDF). In the case when the support of X_i is some interval of $\mathbb{R}$, Pitt and Malik [2011] found an elegant solution to the discontinuity problem by replacing the step-wise CDF with a piecewise linear approximation.

We show in Chapter 3, how to obtain parameter estimates for a GARCH(1,1) series with missing observations by framing the problem in the form (1.8)-(1.9) and utilising SMC with the continuous resampling procedure of Pitt and Malik [2011], henceforth referred to as smooth SMC.

While a particular model may have several state space representations, as we will detail in Section 2.1, the SMC algorithm requires evaluation of $p(Y_i|X_i)$ and thus certain state space representations have more convenient implementations while others may be computationally challenging.

Consider the situation when one has as observations, at time points $t_1 < t_2 < \ldots < t_n = N$, $t_i \in \{1, \ldots, N-1\}$, $i = 1, \ldots, n-1$, the aggregated GARCH series between times t_i and $t_{i-1}+1$, that is $\breve{y}_i = \sum_{k=1}^{n_i} y_{t_{i-1}+k}$ where $n_i = t_i - t_{i-1}$, $t_0 = 0$. This situation can be put in the form of (1.8)-(1.9) with a natural choice of X_i to be $\sigma^2_{t_{i-1}+1}$, however $p(\sum_{k=1}^{n_i} y_{t_{i-1}+k}|\sigma^2_{t_{i-1}+1})$ cannot be evaluated analytically and this hinders computational feasibility of the SMC procedure. In Chapter 4 we identify an alternative specification of the process X_i which allows for a computationally feasible and easy implementation of the smooth SMC procedure providing a method to obtain parameter estimates from an aggregated GARCH series.

In Chapter 5 we identify a computationally feasible state space representation, for smooth SMC, of a discretely observed COGARCH(1,1) process, whereby the COGARCH(1,1) model introduced in Kluppelberg et al. [2004] is the following

system of stochastic differential equations

$$\begin{aligned} \mathrm{d}G_t &= \sigma_t \, \mathrm{d}L_t, \\ \mathrm{d}\sigma_{t+}^2 &= (\beta - \eta\sigma_t^2) \, \mathrm{d}t + \varphi\sigma_t^2 \, \mathrm{d}[L, L]_t^{(d)}, \end{aligned} \tag{1.10}$$

where L_t is a Lévy process, $\beta > 0$, $\eta > 0$ and $\varphi > 0$ are model parameters, and $[L, L]^{(d)}$ denotes the discrete quadratic variation of the Lévy process (cf. Protter [2005], p.66). Here and hereafter, $t\pm$ indicates the right/left-hand limit at t.

The smooth SMC procedure is limited to models where the state equation can be summarised by a single state variable during the resampling. In Chapter 6 we extend the smooth SMC procedure of Pitt and Malik [2011] for parameter estimation of MS-GARCH(1,1) models which are defined as the following system

$$Y_i = \sigma_i z_i + \mu(R_i) \tag{1.11}$$

$$\sigma_i^2 = \omega(R_i) + \alpha(R_i)(Y_{i-1} - \mu(R_{i-1}))^2 + \beta(R_i)\sigma_{i-1}^2 \tag{1.12}$$

for $i = 1, \ldots, N$ whereby

- $(R_i)_{i=1,\ldots,N}$ is an unobserved discrete time ergodic homogeneous Markov chain on a finite space $\mathcal{R} := \{1, \ldots, J\}$,
- $\omega(R_i), \alpha(R_i), \beta(R_i)$ and $\mu(R_i)$ are functions taking values on the respective finite sets $\omega(\mathcal{R}) := \{\omega_1, \ldots, \omega_J\} \in (0, \infty)^J, \alpha(\mathcal{R}) := \{\alpha_1, \ldots, \alpha_J\} \in [0, \infty)^J, \beta(\mathcal{R}) := \{\beta_1, \ldots, \beta_J\} \in [0, \infty)^J$ and $\mu(\mathcal{R}) := \{\mu_1, \ldots, \mu_J\} \in \mathbb{R}^J$,
- $(z_i)_{i=1,\ldots,N}$ as previously defined.

Thus, to summarise, the remainder of this thesis is outlined as follows. In Chapter 2 we introduce the sequential Monte Carlo (SMC) methodology and illustrate its use, in Chapter 3, for parameter estimation of GARCH(1,1) models in the presence of missing observations utilising the continuous resampling procedure of Pitt and Malik [2011]. In Chapter 4 we deal with the task of parameter estimation for GARCH(1,1) models when observations are not merely missing, but have been temporally aggregated. An alternative state space representation is utilised to again perform estimation through a SMC framework. Furthermore, we make the connection that an AR(1)-GARCH(1,1) can be seen as a general form of a temporally aggregated GARCH(1,1) series and thus also illustrate estimation of AR(1)-GARCH(1,1) in the presence of missing data through a SMC framework. In Chapter 5 we venture from the equally spaced discrete time GARCH(1,1) to the continuous time GARCH (COGARCH) model of Kluppelberg et al. [2004]. Through a two-step estimation method we again utilise SMC to provide a means for parameter estimation of a compound Poisson driven COGARCH(1,1). In Chapter 6 we extend the continuous resampling procedure of Pitt and Malik [2011], which is limited to models where the state equation can be summarised by a single state variable during the resampling, to provide a means for parameter estimation for Markov switching GARCH(1,1) models in which one needs two variables, volatility (which has a continuous support) and regime (which has a finite support), to evolve the state equation.

In response to the need for higher dimensional resampling, to tackle models for which two or more continuous variables are required to evolve the state equation, a prototype copula approach is proposed in Appendix A.3. This method however, is still in development and to be further investigated. Proof of concept is demonstrated for parameter estimation of ARMA(1,1)-GARCH(1,1) models in the presence of missing data.

Chapter 2

The sequential Monte Carlo likelihood approximation

Organisation of this chapter is as follows. Section 2.1 introduces the SMC method for likelihood approximation for a general hidden state space model of the form (1.8)-(1.9). Section 2.2 provides illustration of how the bootstrap resampling mechanism employed in SMC results in a likelihood surface prone to being discontinuous in model parameters. A discontinuous likelihood surface is problematic for use in numerical optimisers, inhibiting the ability to obtain parameter estimates through maximum likelihood. An adjustment proposed by Pitt and Malik [2011] to yield a continuous likelihood surface, possible when the support of the hidden state process is some interval of the real line, is detailed in Section 2.3.

2.1 Sequential Monte Carlo for likelihood approximation

Suppose we wish to estimate a set of parameters θ that govern the system (1.8)-(1.9), the likelihood is given as follows,

$$\text{lik}(\theta) = \prod_{i=1}^{n} p(Y_i|Y_{1:i-1}) = \prod_{i=1}^{n} \int p(Y_i|X_i)\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1}), \tag{2.1}$$

with $p(Y_i|Y_{1:i-1})$ and $\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1})$ interpreted as $p(Y_1)$ and $\mathbb{P}(\,\mathrm{d}X_1)$ respectively when $i = 1$. In most cases of interest the conditional distributions $\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1})$ are intractable, which hinders direct evaluation of (2.1). In these cases, one can resort to sequential Monte Carlo, which involves first simulating an i.i.d. random sample $X_i^{(1:K)} = \left\{X_i^{(1)}, \ldots, X_i^{(K)}\right\}$, called *particles*, from an approximation to $\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1})$, denoted by $\tilde{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1})$, to be described below. Then secondly, replacing $\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1})$ by the empirical distribution of the particles $\hat{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1}) = \frac{1}{K}\sum_{k=1}^{K} \delta_{X_i^{(k)}}(\,\mathrm{d}X_i)$, to approximate the integrals with respect to the conditional distributions as follows

$$\int p(Y_i|X_i)\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1}) \approx \int p(Y_i|X_i)\hat{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1}) = \frac{1}{K}\sum_{k=1}^{K} p(Y_i|X_i^{(k)}). \tag{2.2}$$

The SMC approximation of the likelihood is then given by

$$\widehat{\text{lik}}(\theta) = \prod_{i=1}^{n} \Big(\frac{1}{K}\sum_{k=1}^{K} p(Y_i|X_i^{(k)})\Big). \tag{2.3}$$

At this point, we emphasise that two approximations to the conditional distribution $\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1})$ are used: the "tilde" version $\tilde{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1})$ is used to generate particles, while the "hat" version $\hat{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1})$ is used both to approximate the integrals in (2.2), and to produce the "tilde" version in the next time interval where it is needed, as we shall explain next.

To present the method to generate particles from $\tilde{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1})$, we note that

$$\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1}) = \int \mathbb{P}(\,\mathrm{d}X_i|X_{i-1}, Y_{i-1})\mathbb{P}(\,\mathrm{d}X_{i-1}|Y_{1:i-1}), \tag{2.4}$$

which motivates the "tilde" version approximation to $\mathbb{P}(\,\mathrm{d}X_i|Y_{1:i-1})$,

$$\tilde{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1}) = \int \mathbb{P}(\,\mathrm{d}X_i|X_{i-1}, Y_{i-1})\hat{\mathbb{P}}(\,\mathrm{d}X_{i-1}|Y_{1:i-1}), \tag{2.5}$$

where the posterior $\hat{\mathbb{P}}(\,\mathrm{d}X_{i-1}|Y_{1:i-1})$, for $i > 1$, is determined by the particles from the previous time step $i-1$ via

$$\begin{aligned}\mathbb{P}(\,\mathrm{d}X_{i-1}|Y_{1:i-1}) &= \frac{p(Y_{i-1}|X_{i-1})\mathbb{P}(\,\mathrm{d}X_{i-1}|Y_{1:i-2})}{\int p(Y_{i-1}|X_{i-1})\mathbb{P}(\,\mathrm{d}X_{i-1}|Y_{1:i-2})} \\ &\approx \frac{\sum_{k=1}^{K} p(Y_{i-1}|X_{i-1}^{(k)})\delta_{X_{i-1}^{(k)}}(\,\mathrm{d}X_{i-1})}{\sum_{k=1}^{K} p(Y_{i-1}|X_{i-1}^{(k)})} =: \hat{\mathbb{P}}(\,\mathrm{d}X_{i-1}|Y_{1:i-1}).\end{aligned} \tag{2.6}$$

Define $w_i^k := p(Y_i|X_i^{(k)})/\sum_{j=1}^{K} p(Y_i|X_i^{(j)})$ for $k = 1, \ldots, K, i = 1, \ldots, n$, then combining (2.5) and (2.6) gives

$$\tilde{\mathbb{P}}(\,\mathrm{d}X_i|Y_{1:i-1}) = \sum_{k=1}^{K} \mathbb{P}\Big(\,\mathrm{d}X_i|X_{i-1}^{(k)}, Y_{i-1}\Big) w_{i-1}^k. \tag{2.7}$$

To simulate a particle according to (2.7), one would first sample an index ν from the set $\{1, \ldots, K\}$ according to the probabilities $\{w_{i-1}^1, \ldots, w_{i-1}^K\}$ then simulate a random variable W_i to draw a $X_i = g(X_{i-1}^{(\nu)}, Y_{i-1}, W_i)$. Equivalently, this process can be thought of as *bootstrap* resampling of the particles from the previous time step $\{X_{i-1}^{(1)}, \ldots, X_{i-1}^{(K)}\}$ and then evolving this bootstrapped sample, rather than the original sample, to generate the sample of $X_i's$ for the next time step. This bootstrap resampling provides a mechanism to propagate particles in which the observation was more likely to result, thus concentrating "*computation effort on more "promising" regions of the state space*" -Kantas et al. [2009]. The bootstrap resampling SMC procedure we have described was developed by Gordon

et al. [1993]. It was established in Del Moral [1996] that even for a finite particle size K the estimator (2.3) is a consistent and unbiased estimator for the true likelihood function. Computation underflow by multiplication of many small terms can hinder evaluation of (2.3). An alternative is to calculate $\log\left(\widehat{\text{lik}}(\theta)\right)$ which is an asymptotically (as $K \to \infty$) unbiased estimator of the true log-likelihood, although for finite K the bias is[1] $\mathcal{O}(K^{-1})$ (see Del Moral [1996]).

2.2 Problem of discontinuous SMC likelihoods

While bootstrap resampling leads to more samples in regions of substantial probability mass; it comes with the drawback that (2.3) is generally not amenable to numerical optimisers as even fixing randomness and evaluating (2.3) at different parameter values the constructed likelihood surface is unlikely to be smooth. Using bootstrap resampling, discontinuities in the likelihood surface arise because at each time step, the particles $X_i^{(k)}$ are effectively resampled from a discontinuous empirical cumulative distribution function (ECDF)

$$\hat{F}_K(x) = \hat{\mathbb{P}}(X_i \leq x | Y_{1:i}) = \sum_{k=1}^{K} \hat{\mathbb{P}}(X_i^{(k)} | Y_{1:i}) \mathbb{I}(X_i^{(k)} \leq x) \tag{2.8}$$

where $\mathbb{I}(.)$ is the indicator function. Figure 2.1 illustrates how a small change in parameters can be exacerbated by sampling from a step-wise discontinuous distribution. While both the vertical and horizontal shifts in the ECDF are continuous in θ, by inverting the ECDF and sampling X_i from a fixed set of uniform random variables, one cannot guarantee that the whole set of sampled particles $X_i^{(1:K)}$ is continuous in θ. While it is possible that the initial magnitude of the difference in the sampled particles across the parameter change is small, as the procedure is sequential in nature, the difference will be compounded in all later iterations.

[1] For functions A and B that map positive integers to some subset of $\mathbb{R}$, $A(j)$ is $\mathcal{O}(B(j))$ means there exists constants $C > 0$ and $j_0 \in \mathbb{Z}_+$ such that $|A(j)| \leq C|B(j)|$ for every integer $j > j_0$.

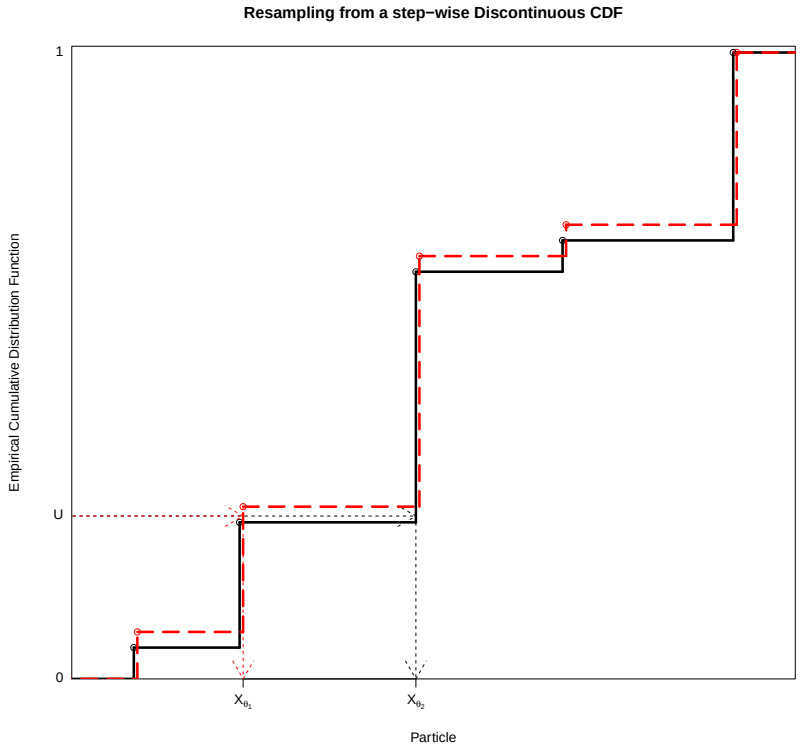


Figure 2.1: The black (solid line) step function is the ECDF under some parameter θ_1 and the red (dashed line) step function is the ECDF under some parameter θ_2 very close to θ_1. A small shift in parameter θ causes a small shift in the ECDF. Fixing the same uniform U and inverting the two discontinuous step-wise ECDFs, as illustrated above, can result in drastically different particles X_{θ_1} and X_{θ_2} sampled.

2.3 Solution for continuous SMC likelihoods when $X \subset \mathbb{R}$

In the case when the support of X_i is some interval of $\mathbb{R}$, Pitt and Malik [2011] provide a simple solution to remedy the discontinuity problem by constructing a continuous approximation $\tilde{F}_K(x)$ of $\hat{F}_K(x)$ and then resampling particles $X_i^{(k)}$ by inverting uniforms based on $\tilde{F}_K(x)$. Assume the particles $X_i^{(k)}$, $k = 1, \ldots, K$ are sorted in ascending order, then

$$\tilde{F}_N(x) = \pi_0 \mathbb{I}(x \geq X_i^{(1)}) + \sum_{k=1}^{K-1} \pi_k H\left(\frac{x - X_i^{(k)}}{X_i^{(k+1)} - X_i^{(k)}}\right) + \pi_K \mathbb{I}(x \geq X_i^{(K)}) \qquad (2.9)$$

where $\pi_0 = \hat{\mathbb{P}}(X_i^{(1)}|Y_{1:i})/2$, $\pi_K = \hat{\mathbb{P}}(X_i^{(K)}|Y_{1:i})/2$ and $\pi_k = \Big(\hat{\mathbb{P}}(X_i^{(k+1)}|Y_{1:i}) + \hat{\mathbb{P}}(X_i^{(k)}|Y_{1:i})\Big)/2$ for $k = 1, \dots K-1$ and $H(z) := \max\Big(0, \min(z,1)\Big)$. Figure 2.2 illustrates the idea behind the procedure (see also Figure 1 in Pitt and Malik [2011]). It was shown in Pitt and Malik [2011] that the distance $\|\hat{F}_K(x) - \tilde{F}_K(x)\|_\infty$ is of order $\frac{1}{K}$.

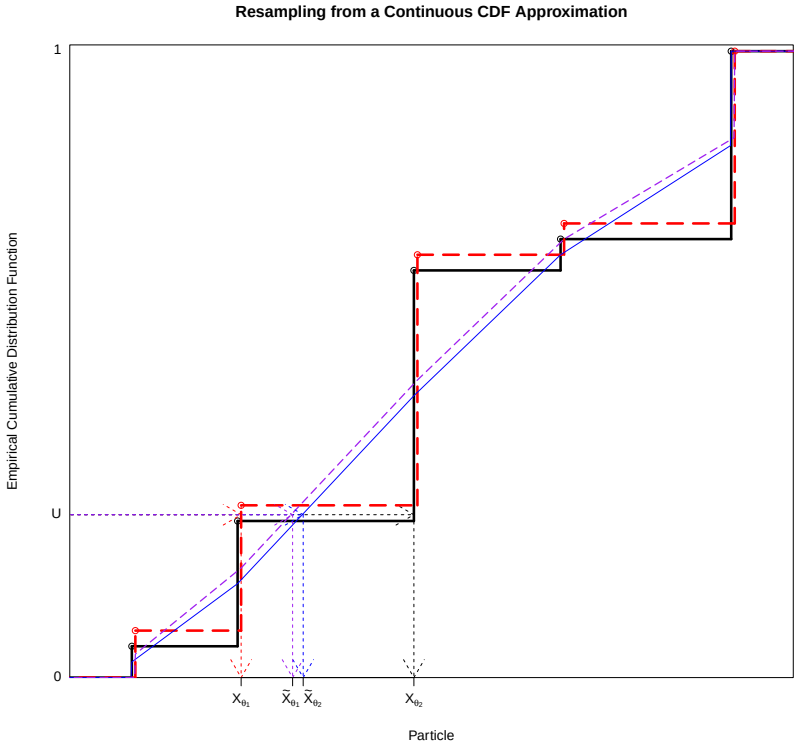


Figure 2.2: Previously, inverting the same uniform across the ECDFs under θ_1 (black solid line) and θ_2 (red long-dashed line) could result in drastically different particles X_{θ_1} and X_{θ_2} sampled. The continuous approximation of the ECDF under θ_1 is the blue (thin solid line) function and respectively under θ_2 is the purple (short-dashed line) function. By instead inverting these continuous approximations, large deviations between sampled particles under small parameter changes are eliminated.

Chapter 3

Application: Likelihood inference for partially observed GARCH(1,1) series

This chapter illustrates application of the SMC methodology for parameter estimation of GARCH(1,1) models in the presence of missing observations. Throughout, the missingness is assumed to occur independently of the process being observed. This may be deterministic sampling (such as regular omission of data on stock indices on weekends or scheduled holidays), random (occuring for instance due to intermittent failure of recording equipment) or a combination. However, whether an observation is missed is not dependent on the values of the process of interest either at the missing times or at any other observed times.

Organisation of this chapter is as follows. Section 3.1 details how a partially observed GARCH(1,1) time series is of the form (1.8) – (1.9) allowing use of the SMC method. In Section 3.2.1, through simulation studies, the performance of this estimation method at varying degrees of missingness is assessed and also compared in Section 3.2.2 against the performance of several common imputation methods. Furthermore, in Section 3.2.3, for the special case of a partially observed

GARCH(1,1) series at equally spaced time points the performance of this method is compared against Drost and Nijman [1993]'s "Weak" GARCH formula. Utilisation of the method on several real world applications such as modelling both daily and high frequency intraday stock price returns as well as high frequency wind strength measurements are illustrated in Section 3.3. The chapter concludes with a discussion in Section 3.4.

3.1 SMC for partially observed GARCH(1, 1)

Throughout we let $n_i = t_i - t_{i-1}$, $i = 1, \ldots, n$ with $t_0 = 0$. To obtain, via SMC, an approximation of the marginal likelihood of a partially observed GARCH(1,1) time series (1.5), it is straightforward to put the problem in the form of (1.8)−(1.9) by setting, for $i = 1, \ldots, n$: $Y_i = y_{t_i}$, $Z_i = z_{t_i}$, $X_i = \sigma^2_{t_i}$,

$$W_i = \begin{cases} \{z_{t_i-1}, \ldots, z_{t_{i-1}+1}\} & \text{when } n_i > 1, \\ \emptyset & \text{when } n_i = 1, \end{cases} \tag{3.1}$$

with

$$X_i = g(X_{i-1}, Y_{i-1}, W_i) = \begin{cases} \omega + \alpha Y^2_{i-1} + \beta X_{i-1}, & \text{when } n_i = 1 \\ \omega + \omega\Big(\sum_{j=1}^{n_i-1} \prod_{k=1}^{j} (\alpha z^2_{t_i-k} + \beta)\Big) & \\ \quad + (\alpha Y^2_{i-1} + \beta X_{i-1}) \prod_{j=1}^{n_i-1} (\alpha z^2_{t_i-j} + \beta) & \text{when } n_i > 1. \end{cases} \tag{3.2}$$

Then $Y_i = f(X_i, Z_i) = \sqrt{X_i} Z_i$ and

$$p(Y_i|X_i) = \frac{1}{\sqrt{X_i}} h\Big(\frac{Y_i}{\sqrt{X_i}}\Big). \tag{3.3}$$

Assuming one can simulate the noise z_i as well as evaluate its density h, the SMC likelihood approximation detailed in Section 2.1 can be performed. As the support of X_i is an interval of $\mathbb{R}$, a continuous log-likelihood surface can be constructed, using the method detailed in section 2.3, and subsequently maximised to obtain parameter estimates and their standard errors using the Hessian.

3.2 Simulation studies

3.2.1 Impact of proportion of missing data on SMC performance

In this section we test the performance of the SMC estimation method at varying degrees of missingness. In order to study the impact of different amounts of missingness, rather than variations in total number of observations, on the performance of the method we keep the total sample size at 2000 observations. For example, 20% missing with 2000 effective sample size results from a series of 2500 consecutive values of which 20% are missing. This allows us to study the impact of missingness in isolation from any impact on performance attributable to smaller sample sizes. Denote mp as the missing percentage. For each missing scenario we simulate 1000 different data sets of a length $N = \lfloor \frac{2000}{1-mp} \rfloor$ GARCH(1,1) series with standard normal innovations with true parameters $\omega = 0.1, \alpha = 0.08$ and $\beta = 0.9$, then for each data set $\lfloor mp \times N \rfloor$ 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.

The SMC algorithm for likelihood approximation was implemented in C++ and the optimisation was performed in R using the *constrOptim* function from the *stats* library. Parameters were optimised over the set

$$\Omega = \{\omega > 0; \alpha > 0; \beta > 0; \alpha + \beta < 1\}. \tag{3.4}$$

The inverse of the negative of the numerical Hessian $H(\theta)$, of the SMC obtained log-likelihood, at the optimum $\hat{\theta}$ is used as the estimated covariance matrix of $\hat{\theta}$ from which standard errors of individual parameter estimates are obtained throughout.

Table 3.1 displays the results from 1000 simulations of the GARCH model with missing percentages of 5%, 10%, 20%, 35% and 50%. This range was chosen to reflect the extent of 'missingness' encountered in the real examples considered in Section 3.3. In this table and in similar tables throughout the thesis we display the following summary statistics, all taken over the 1000 replications that are simulated:

i) **Mean** the average of the individual estimates.

ii) **Bias** the Mean minus the true parameter value.

iii) **SD** the standard deviation of the individual estimates.

iv) **Mean SE** the average of the standard errors obtained from the Hessian, as described above, for each individual replication.

v) **RMSE** the root mean squared error of the individual estimates compared to the true parameter values.

Results obtained utilising 250 and 1000 particles are also shown to assess the impact of increasing particle simulation effort on the statistical performance of the resulting parameter estimates. Comparing these, it is clear that using four times as many particles scarcely impacts the statistical performance of SMC as an approximation to the log-likelihood and the resulting parameters and their reported standard errors, thus we will concentrate on summarising the impact on estimation associated with increasing the proportion of missing data.

Some points about Table 3.1 are:

- The average estimated standard error closely approximates the observed standard deviation suggesting that the use of the Hessian for obtaining standard errors is reliable.

- Across all three parameters the RMSE is primarily due to variance with squared bias being small in comparison.

- The bias as a percentage of true value is largest for estimation of ω and is essentially negligible for estimation of α and β.

- The RMSE is stable with increasing proportion of missingness for α and β. However, for ω, RMSE decreases as the percentage of missingness increases due, roughly equally, to bias and standard deviation both decreasing by around 25% for the 50% missing scenario relative to the no missing scenario. This may be somewhat counterintuitive. How can gaps in the data improve estimation of the long-run variance for the GARCH process? We speculate that the gaps in the data produce near independent complete subsequences of realisations and that this in turn increases statistical efficiency. This phenomena has been observed in other time series situations as demonstrated in Dunsmuir [1981].

In summary, SMC has been demonstrated to provide accurate estimates of the GARCH(1,1) parameters with performance that certainly does not degrade as the percentage of missing data increases (holding the overall sample size constant) and there is actually improvement in the estimation of ω, the parameter that directly measures the overall level of variability in the series, as missingness increases. Increasing the number of particles beyond 250 is not warranted in this setting.

Additional graphical interpretation of the results of Table 3.1 as the percentage of missingness increases can be obtained from Figure 3.1. We discuss these in more detail in the next subsection where we compare the SMC based likelihood method with other available methods based on single and multiple imputation.

A summary, across the 1000 simulations for each level of missingness, of computational times using an Intel Xeon 3.06 GHz processor is provided in Table 3.2. The time to complete a SMC likelihood evaluation can be seen to increase, roughly linearly, with N as well as the number of particles K.

Table 3.1: SMC estimation of the GARCH(1,1) model with missing data: Summary statistics for 1000 replications for each level of missingness: Mean is the sample mean of the estimates, Bias is Mean minus the true parameter value, SD denotes standard deviation of the estimates across replications, Mean SE is the average reported standard error using the Hessian and RMSE is the root mean squared error of the estimates across replications.

		250 Particles					1000 Particles				
% Missing	0%	5%	10%	20%	35%	50%	5%	10%	20%	35%	50%
$\omega = 0.1$											
Mean	0.1176	0.1179	0.1178	0.1168	0.1142	0.1116	0.1177	0.1176	0.1164	0.1135	0.1105
Bias	0.0176	0.0179	0.0178	0.0168	0.0142	0.0116	0.0177	0.0176	0.0164	0.0135	0.0105
SD	0.0424	0.0409	0.0423	0.0403	0.0381	0.0345	0.0409	0.0419	0.0400	0.0376	0.0341
Mean SE	0.0383	0.0378	0.0373	0.0361	0.0340	0.0319	0.0378	0.0373	0.0361	0.0340	0.0320
RMSE	0.0459	0.0446	0.0459	0.0437	0.0406	0.0364	0.0446	0.0454	0.0433	0.0400	0.0357
$\alpha = 0.08$											
Mean	0.0801	0.0799	0.0801	0.0802	0.0806	0.0801	0.0798	0.0801	0.0802	0.0804	0.0799
Bias	0.0001	-0.0001	0.0001	0.0002	0.0006	0.0001	-0.0002	0.0001	0.0002	0.0004	-0.0001
SD	0.0131	0.0131	0.0131	0.0132	0.0135	0.0127	0.0131	0.0131	0.0132	0.0135	0.0127
Mean SE	0.0135	0.0134	0.0134	0.0134	0.0133	0.0133	0.0134	0.0134	0.0134	0.0133	0.0133
RMSE	0.0131	0.0131	0.0131	0.0132	0.0135	0.0127	0.0131	0.0131	0.0132	0.0135	0.0127
$\beta = 0.9$											
Mean	0.8956	0.8956	0.8954	0.8956	0.8959	0.8970	0.8957	0.8955	0.8957	0.8963	0.8975
Bias	-0.0044	-0.0044	-0.0046	-0.0044	-0.0041	-0.0030	-0.0043	-0.0045	-0.0043	-0.0037	-0.0025
SD	0.0178	0.0176	0.0180	0.0181	0.0183	0.0171	0.0177	0.0180	0.0181	0.0183	0.0171
Mean SE	0.0179	0.0178	0.0178	0.0177	0.0175	0.0174	0.0178	0.0178	0.0178	0.0175	0.0174
RMSE	0.0183	0.0182	0.0186	0.0187	0.0188	0.0173	0.0182	0.0185	0.0186	0.0186	0.0172

Table 3.2: Summary, across the 1000 simulations for each level of missingness, of computational times for the SMC optimisation using an Intel Xeon 3.06 GHz.

		250 particles			1000 particles		
% Missing	N	Time (secs) to complete optimisation	Avg. number of function calls	Avg. time to evaluate function (secs)	Time (secs) to complete optimisation	Avg. number of function calls	Avg. time to evaluate function (secs)
5%	2105	16.91	157.45	0.107	71.67	157.77	0.454
10%	2222	17.95	160.04	0.112	75.39	158.63	0.475
20%	2500	20.72	169.24	0.122	87.22	168.98	0.516
35%	3076	25.38	175.62	0.145	106.48	175.39	0.607
50%	4000	30.64	168.36	0.182	126.54	166.31	0.761

3.2.2 Comparison of SMC with common imputation methods

Additionally, we perform three common imputation procedures and compare their performance with the SMC estimates. Imputation procedures yield artifically complete data sets for which standard estimation methods are applied. The first procedure is a form of single imputation (SI), whereby all missing values are replaced by the mean, which in this case is zero. The second procedure, multiple imputation (MI) (cf. Rubin [1987]) simulates m replications of the complete data set, the estimate of the parameter of interest is taken as the average of the m complete data estimates. The MI procedure simulates the missing innovations z_i, where $i \in \mathbb{M}$ to "complete" the data set. However, unlike the SMC procedure which concentrates simulation effort on choices of the z_i which are more likely given the observed data, MI does not probabilistically match the simulated $z_i, i \in \mathbb{M}$ to the observed $y_i, i \in \mathbb{O}$. As a result, simulation using the MI method is inefficient. MI was performed using the code available in the SAS online manual[1] using $m = 250$ replications. Finally, the third procedure is to ignore missing values (IG), skipping over their occurrence, by concatenating the segments of non-missing observations

[1] https://support.sas.com/rnd/app/ets/examples/garchimpute/index.htm

to form a shorter series with no gaps.

Estimates, under the same amounts of missingness as considered earlier, are displayed in Table 3.3 for SI and MI and Table 3.4 for IG. Figure 3.1 provides a visual comparison of the bias, standard deviation and RMSE performance of the four estimators through the different degrees of missingness. Figure 3.2 displays kernel density plots of the estimates obtained from the SMC algorithm and the three imputation procedures for the 5% and 50% missing scenarios. Taken together these tables and figures demonstrate:

- For the imputation methods (SI, MI and IG) the bias, across all three parameters, increases in absolute value as the percentage of missing data increases. The effect of missingness on the standard deviation of estimates from the imputation methods is less clear to decipher, however looking at the totality of bias and standard deviation through RMSE we see estimation performance in this regard worsens, in particular for α and β, as missingness is increased.

- In constrast the SMC method (using either 250 or 1000 particles) give very stable values for bias, standard deviation and hence overall RMSE as the missingness increases.

- MI is the closest competitor to SMC but is considerably more biased in its estimates of α and β for moderate to high levels of missingness.

- The kernel density plots in Figure 3.2 show that the distributions for the imputation methods are skewed for large amounts of missingness whereas those for SMC are close to symmetric and closer to normally distributed. For small percentages of missing data (5%) the densities are similar for all methods with the substantial positive skewness for the estimates of ω.

In summary, for small amounts of missing data, the imputation and SMC methods give similarly accurate estimation results. However, as the percentage of

missingness increases, SMC clearly provides an overall superior set of estimates in terms of bias and standard deviation and in such situations the increased computational cost of SMC is worthwhile.

Table 3.3: Summary of estimation performance, across 1000 replications for each level of missingness, utilising the single and multiple imputation methods.

	single imputation					multiple imputation				
% Missing	5%	10%	20%	35%	50%	5%	10%	20%	35%	50%
$\omega = 0.1$										
Mean	0.1142	0.1203	0.1101	0.0898	0.0671	0.1175	0.1170	0.1139	0.1097	0.0867
Bias	0.0142	0.0203	0.0101	-0.0102	-0.0329	0.0175	0.0170	0.0139	0.0097	-0.0133
SD	0.0411	0.0573	0.0879	0.0621	0.0740	0.0402	0.0406	0.0380	0.0671	0.0285
Mean SE	0.0370	0.0387	0.0294	0.0227	0.0124	0.0397	0.0404	0.0377	0.0614	0.0262
RMSE	0.0435	0.0608	0.0885	0.0629	0.0810	0.0438	0.0440	0.0405	0.0678	0.0315
$\alpha = 0.08$										
Mean	0.0764	0.0751	0.0670	0.0554	0.0421	0.0783	0.0765	0.0725	0.0670	0.0553
Bias	-0.0036	-0.0049	-0.0130	-0.0246	-0.0379	-0.0017	-0.0035	-0.0075	-0.0130	-0.0247
SD	0.0129	0.0118	0.0128	0.0103	0.0107	0.0129	0.0126	0.0121	0.0143	0.0102
Mean SE	0.0125	0.0121	0.0100	0.0075	0.0050	0.0133	0.0131	0.0125	0.0125	0.0090
RMSE	0.0134	0.0127	0.0182	0.0267	0.0394	0.0130	0.0131	0.0143	0.0194	0.0267
$\beta = 0.9$										
Mean	0.8987	0.8967	0.9037	0.9153	0.9295	0.8973	0.8992	0.9035	0.9097	0.9238
Bias	-0.0013	-0.0033	0.0037	0.0153	0.0295	-0.0027	-0.0008	0.0035	0.0097	0.0238
SD	0.0180	0.0215	0.0350	0.0271	0.0391	0.0173	0.0169	0.0165	0.0150	0.0146
Mean SE	0.0172	0.0180	0.0150	0.0129	0.0087	0.0181	0.0181	0.0173	0.0164	0.0131
RMSE	0.0181	0.0217	0.0352	0.0312	0.0490	0.0175	0.0169	0.0169	0.0178	0.0280

Table 3.4: Summary of estimation performance, across 1000 replications for each level of missingness, utilising the IG method.

	$\omega = 0.1$					$\alpha = 0.08$				
% Missing	5%	10%	20%	35%	50%	5%	10%	20%	35%	50%
Mean	0.1235	0.1342	0.1454	0.1744	0.2196	0.0814	0.0831	0.0865	0.0924	0.0992
Bias	0.0235	0.0342	0.0454	0.0744	0.1196	0.0014	0.0031	0.0065	0.0124	0.0192
SD	0.0435	0.1380	0.0498	0.0568	0.0671	0.0134	0.0137	0.0143	0.0159	0.0163
Mean SE	0.0397	0.0413	0.0448	0.0511	0.0615	0.0136	0.0139	0.0144	0.0151	0.0163
RMSE	0.0494	0.1422	0.0674	0.0936	0.1371	0.0135	0.0141	0.0157	0.0202	0.0252

	$\beta = 0.9$				
% Missing	5%	10%	20%	35%	50%
Mean	0.8931	0.8892	0.8835	0.8717	0.8559
Bias	-0.0069	-0.0108	-0.0165	-0.0283	-0.0441
SD	0.0182	0.0328	0.0204	0.0228	0.0246
Mean SE	0.0182	0.0187	0.0197	0.0213	0.0240
RMSE	0.0195	0.0346	0.0262	0.0363	0.0505

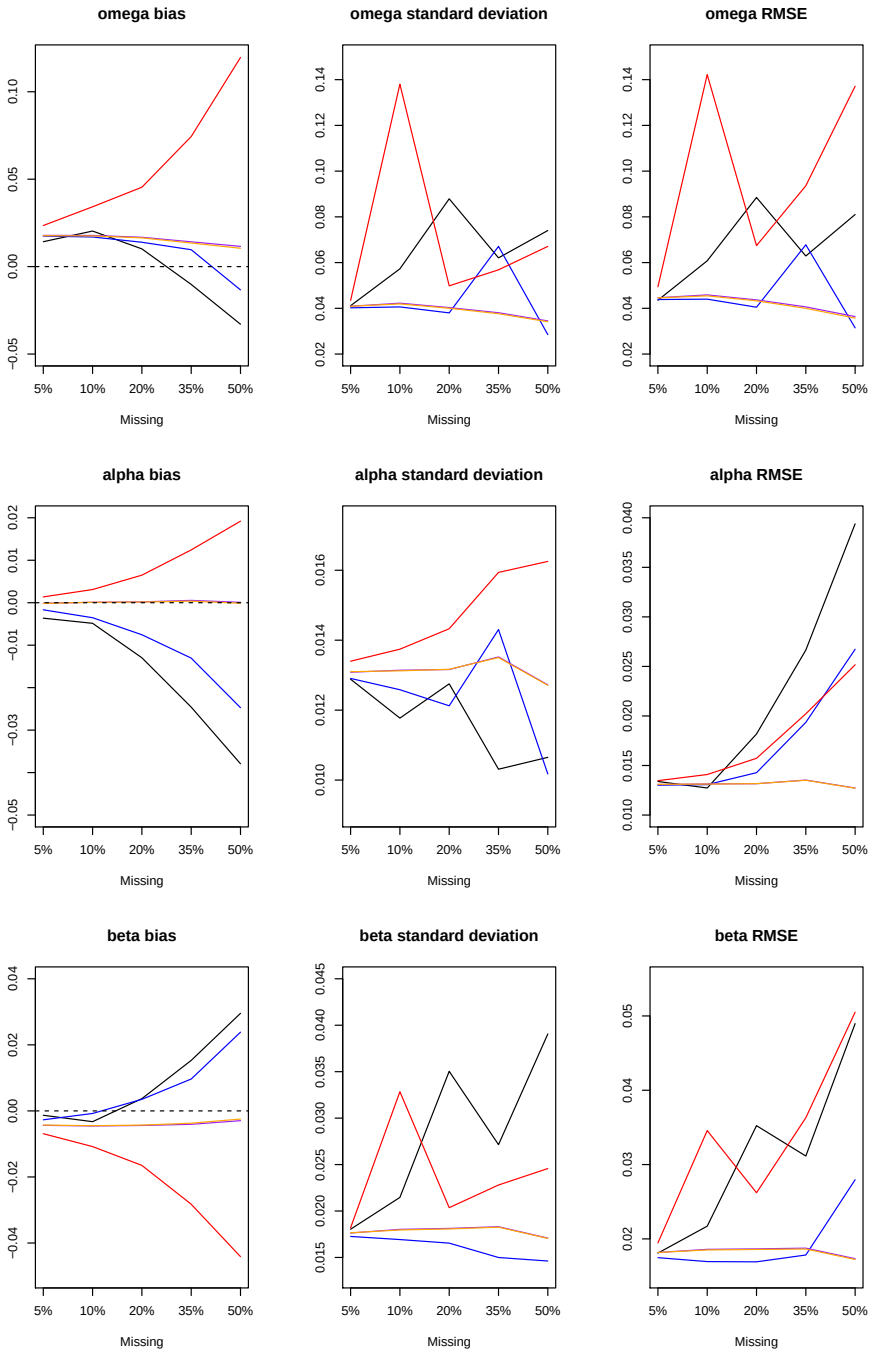


Figure 3.1: Visual summary of estimation performance utilising the single imputation (black), multiple imputation (blue), IG (red), SMC at 250 particles (purple) and SMC at 1000 particles (orange), at different levels of missingness.

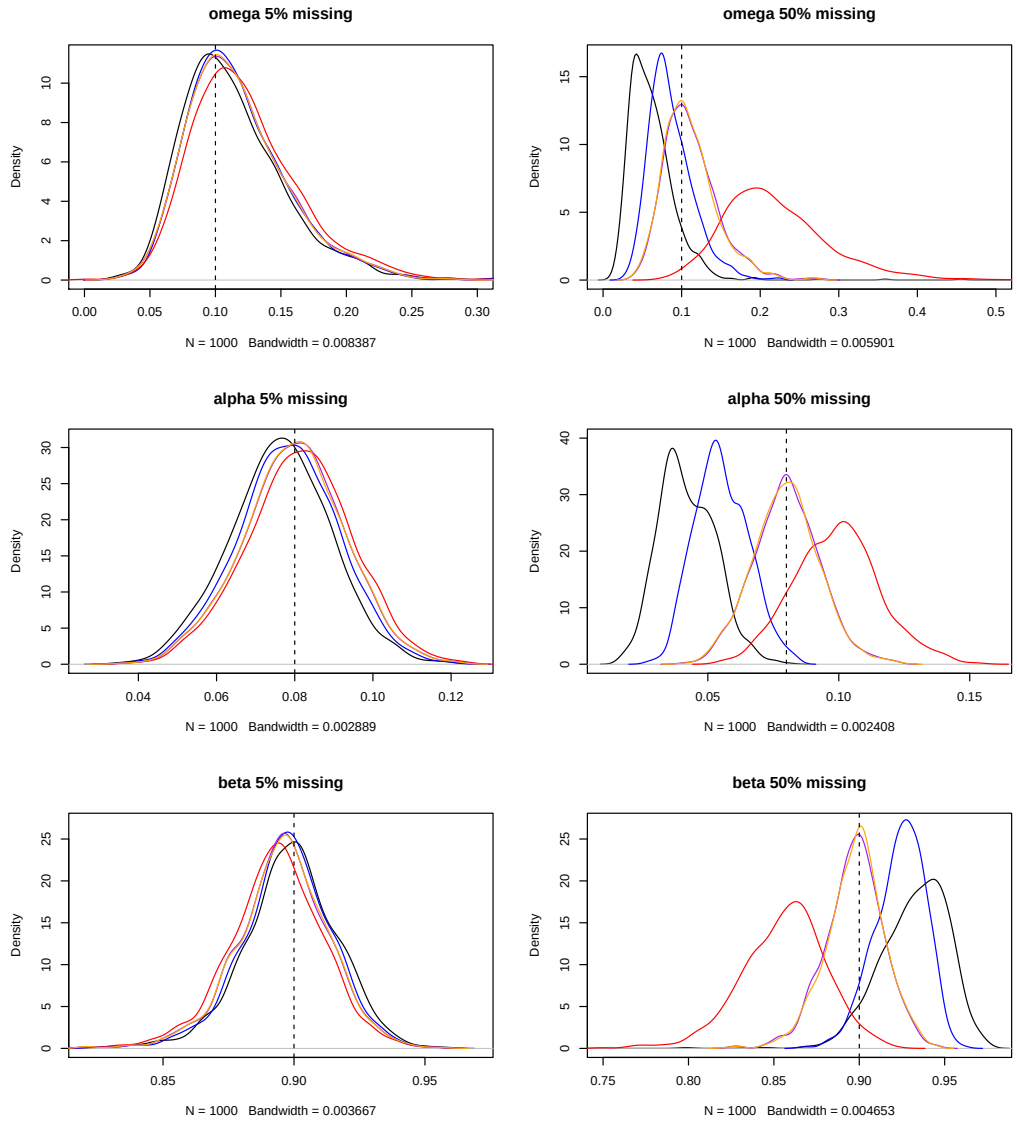


Figure 3.2: Kernel density plots of the estimates obtained utilising the single imputation (black), multiple imputation (blue), IG (red), SMC at 250 particles (purple) and SMC at 1000 particles (orange) for the 5% (left) and 50% (right) missing scenarios.

3.2.3 Comparison with "Weak" GARCH approximation

An important result established by Drost and Nijman [1993] is that if one observed every $m-th$ observation of a GARCH series $(Y_i)_{i=1,\ldots,N}$, the observed series $(Y_{im})_{i=1,\ldots,\lfloor \frac{N}{m} \rfloor}$ is itself not a GARCH series. Drost and Nijman [1993] introduce the concept of *Weak* GARCH, whereby a sequence $(\epsilon_i)_{i=1,\ldots,N}$ is said to be a *Weak*

GARCH process if

$$B[\epsilon_i|\epsilon_{i-1},\epsilon_{i-2},\ldots]=0 \tag{3.5}$$

$$B[\epsilon_i^2|\epsilon_{i-1},\epsilon_{i-2},\ldots]=:h_i=\omega+\alpha\epsilon_{i-1}^2+\beta h_{i-1} \tag{3.6}$$

where $B[x|\epsilon_{i-1},\epsilon_{i-2},\ldots]$ denotes the best linear predictor of x in terms of $1,\epsilon_{i-1},\epsilon_{i-2},\ldots,\epsilon_{i-1}^2,\epsilon_{i-2}^2,\ldots,$ that is,

$$\mathbb{E}\Big(x-B[x|\epsilon_{i-1},\epsilon_{i-2},\ldots]\Big)\epsilon_{i-j}^r=0 \text{ for } j\geq 1 \text{ and } r=0,1,2. \tag{3.7}$$

It has been shown that if $(\epsilon_i)_{i=1,\ldots,N}$ is a weak GARCH series with parameters ω,α and β then $(\epsilon_{im})_{i=1,\ldots,\lfloor\frac{N}{m}\rfloor}$ is also a weak GARCH with parameters

$$\omega_{(m)}=\omega\frac{1-(\alpha+\beta)^m}{1-\alpha-\beta} \tag{3.8}$$

$$\alpha_{(m)}=(\alpha+\beta)^m-\beta_{(m)} \tag{3.9}$$

with $\beta_{(m)}\in(0,1)$ as the solution to

$$\frac{\beta_{(m)}}{1+\beta_{(m)}^2}=\frac{\beta(\alpha+\beta)^{m-1}}{1+\alpha^2\frac{1-(\alpha+\beta)^{2m-2}}{1-(\alpha+\beta)^2}+\beta^2(\alpha+\beta)^{2m-2}}. \tag{3.10}$$

From the above results, one may then infer via (3.8)-(3.10) a weak GARCH at one frequency by fitting a weak GARCH at another frequency. The series $(Y_{im})_{i=1,\ldots,\lfloor\frac{N}{m}\rfloor}$ is a weak GARCH series and since a GARCH series is also a weak GARCH series one is able to utilise (3.8)-(3.10) to approximate the parameters of the higher frequency standard GARCH series $(Y_i)_{i=1,\ldots,N}$ from the lower observational frequency weak GARCH series $(Y_{im})_{i=1,\ldots,\lfloor\frac{N}{m}\rfloor}$.

Below, we compare the performance of this approximation against that of the SMC estimates using 250 particles. We generate 1000 data sets of a GARCH(1,1) series with standard normal innovations with parameters $\omega=0.1,\alpha=0.08$ and

$\beta = 0.9$, taking $N = 10000$ and $m = 5$ which leads to 80% of observations missing. Following Drost and Nijman [1993], the weak GARCH parameters $\omega_{(m)}, \alpha_{(m)}$ and $\beta_{(m)}$ are estimated using Quasi-Maximum Likelihood (QML) (cf. Lee and Hansen [1994]). The obtained weak GARCH estimates are then converted to estimates of ω, α and β via (3.8)-(3.10), this process hereafter referred to as the Drost approximation.

It is seen in Table 3.5 that the SMC estimates, driven by lower variances, are slightly better than the Drost approximation estimates in terms of RMSE. Using an Intel Xeon 3.06 GHz processor the Drost optimisation takes roughly 2 seconds, while the SMC optimisation takes roughly 93 seconds ($\approx$ 260 function evaluations at $\approx$ 0.36 seconds per function evaluation). The Drost approximation is very fast for a small loss in efficiency, however the method has limited applicability for missing data in practice as it is restricted to only equally spaced missing data.

To our knowledge, the appropriate covariance estimator for the QML applied to weak GARCH estimation has not been developed and thus we do not report Mean SE for this method in Table 3.5.

Table 3.5: Comparison of performance of the Drost and SMC estimates across 1000 simulated data sets.

	Drost approxmiation			SMC estimates		
	ω	α	β	ω	α	β
Mean	0.1121	0.0798	0.8974	0.1090	0.0808	0.8973
Bias	0.0121	-0.0002	-0.0026	0.0090	0.0008	-0.0027
SD	0.0356	0.0163	0.0220	0.0336	0.0148	0.0200
RMSE	0.0376	0.0163	0.0221	0.0348	0.0148	0.0202
Mean SE	NA	NA	NA	0.0296	0.0143	0.0187

3.3 Real data analysis

In this section we apply the SMC estimation method in four situations which demonstrate various ways in which missingness can arise in volatility modelling. Section 3.3.1 considers daily adjusted closing values of various stock indices over a roughly 10 year period. Section 3.3.2 involves application to 30 minute wind data from three weather stations in the Sydney region. Section 3.3.3 is a playful example which revisits the hypothesis that bright sunny days tend to increase returns on the stock market and Section 3.3.4 concerns high frequency capital asset pricing and is a situation in which the missing data occurs not only in the individual stock under consideration but also in the portfolio against which it is regressed.

3.3.1 Stock data

We consider the time series of daily adjusted close of the Australian S&P/ASX 200 Index (AXJO), the Mexican IPC Index (MXX) and the Japanese Nikkei 225 Index (N225) for the period 28th of February 1994 to the 31st of March 2004 obtained from *Yahoo! Finance*. These time series span 3685 calendar days for which 2633 are weekdays. Of those weekdays, 82, 109 and 147 are, respectively for Australia, Mexico and Japan, public holidays. Denote

- $P_i^{(AXJO)}$, $P_i^{(MXX)}$, $P_i^{(N225)}$, $i = 1, \ldots, 3685$, the adjusted close of the three indices for each of the 3685 calenders days. There are respectively 30.77%, 31.51% and 32.54% of observations missing in these three series attributed to weekends and public holidays.

- $\tilde{P}_i^{(AXJO)}$, $\tilde{P}_i^{(MXX)}$ and $\tilde{P}_i^{(N225)}$, $i = 1, \ldots, 2633$, the adjusted close of the three indices for each of the 2633 weekdays. That is, $\tilde{P}_i$, operates on a time scale that skips over weekends and sequentially numbers weekdays only. There are respectively 3.11%, 4.14% and 5.58% of observations missing in these three series attributed to public holidays.

- $\bar{P}_i^{(AXJO)}, i = 1, \ldots, 2551$, $\bar{P}_i^{(MXX)}, i = 1, \ldots, 2524$ and $\bar{P}_i^{(N225)}, i = 1, \ldots, 2486$ the adjusted close of the three indices on a time scale that skips over all missing values and sequentially numbers days on which the respective market is operating. There are no observations missing in these series.

A common approach by practitioners is to ignore missing values and perform their analysis on the series $(\bar{P}_i)$, however with the SMC methodology, a tool to perform analysis on the series (P_i) and $(\tilde{P}_i)$ is now available.

For each of the three indices, we calculate three log return series $(R_i), (\tilde{R}_i), (\bar{R}_i)$, whereby, for instance $R_i := 100 \times \ln(P_i) - \ln(P_{i-1})$. The percentage of missing data in $(R_i^{(AXJO)})$, $(R_i^{(MXX)})$, $(R_i^{(N225)})$, $(\tilde{R}_i^{(AXJO)})$, $(\tilde{R}_i^{(MXX)})$ and $(\tilde{R}_i^{(N225)})$ is respectively 45.47%, 47.2%, 47.86%, 5.36%, 7.83% and 9.95%. We fit to the nine log returns series $\mu + \epsilon_i$, where (ϵ_i) is a GARCH(1,1) series with standard normal innovations. The parameter estimates along with confidence intervals are displayed in Table 3.6.

From the table, we see in all three indices, that estimates of α and β are not statistically different under the three time treatments. However, in terms of ω, consistent across the three indices, we see that the ω estimates from the $\tilde{R}_i$ and $\bar{R}_i$ series lie above the upper 95% level of the ω estimate from R_i.

To recap: R_i calculates returns over calendar days. $\tilde{R}_i$ calculates returns over calendar days from Mon-Tue, Tue-Wed, Wed-Thu and Thu-Fri, but instead of treating the time from Fri-Mon as 3 days, this convention compresses this physical time to 1 day. $\bar{R}_i$ has the same time conventions as $\tilde{R}_i$, with the addition of further physical time compression when trading holidays occur.

A reasonable conjecture would be to expect estimates of μ and the long-run average variance $\frac{\omega}{1-\alpha-\beta}$ to be higher in the more (physical) time compressed series. However, from the table, there does not seem to be any statistical difference in these quantities between the three time treatments. Although, looking at the point estimates of $\frac{\omega}{1-\alpha-\beta}$, these appear, for all three indices, in line with the conjecture.

Table 3.6: Estimated GARCH parameters with 95% confidence intervals for the returns data.

	ω	α	β	μ	$\omega/(1-\alpha-\beta)$
$R_i^{(AXJO)}$	0.018 (0.007, 0.028)	0.098 (0.068, 0.129)	0.874 (0.833, 0.916)	0.030 (0.000, 0.060)	0.644 (0.489, 0.799)
$\tilde{R}_i^{(AXJO)}$	0.028 (0.013, 0.043)	0.098 (0.070, 0.126)	0.860 (0.817, 0.903)	0.034 (0.006, 0.063)	0.675 (0.537, 0.812)
$\bar{R}_i^{(AXJO)}$	0.031 (0.014, 0.048)	0.093 (0.065, 0.120)	0.861 (0.815, 0.906)	0.035 (0.006, 0.063)	0.676 (0.551, 0.801)
$R_i^{(MXX)}$	0.031 (0.010, 0.053)	0.102 (0.067, 0.138)	0.890 (0.851, 0.928)	0.151 (0.091, 0.212)	3.998 (1.329, 6.666)
$\tilde{R}_i^{(MXX)}$	0.064 (0.029, 0.099)	0.131 (0.094, 0.168)	0.854 (0.813, 0.896)	0.126 (0.071, 0.180)	4.307 (1.659, 6.956)
$\bar{R}_i^{(MXX)}$	0.053 (0.020, 0.087)	0.115 (0.081, 0.148)	0.874 (0.836, 0.912)	0.122 (0.067, 0.176)	4.808 (1.165, 8.451)
$R_i^{(N225)}$	0.020 (0.007, 0.034)	0.065 (0.043, 0.088)	0.925 (0.899, 0.950)	-0.013 (-0.067, 0.041)	2.056 (1.322, 2.791)
$\tilde{R}_i^{(N225)}$	0.047 (0.023, 0.071)	0.081 (0.057, 0.104)	0.898 (0.870, 0.926)	-0.016 (-0.067, 0.034)	2.191 (1.582, 2.801)
$\bar{R}_i^{(N225)}$	0.057 (0.031, 0.083)	0.077 (0.054, 0.099)	0.898 (0.869, 0.926)	0.004 (-0.047, 0.055)	2.236 (1.702, 2.771)

It is not within the scope of this thesis to advocate whether a practitioner should ignore or incorporate data gaps due to public holidays and or weekends, it is merely an illustration of a tool at the practitioners disposal should they wish to incorporate these data gaps and compare the impact on volatility levels and dynamics of the different ways in which gaps in the price series are treated. Bondon and Bahamonde [2012] in their analysis of fitting an ARCH model in the presence of missing observations used the Chilean IPSA stock index where weekend gaps were ignored but public holidays were treated as missing observations.

If one ignores the data gaps, then one is making the assumption that no additional volatility occurs on these days, whereas if one accounts for the gaps then one is making the assumption that these days carry the same amount of volatility as a trading day. Intuitively it may be realistic to have a mindset in between, as "*investors do receive and process information during periods when markets are closed, and the information processed over a weekend will affect the price at and after the opening on Monday*" -Fortune [1999], however in the absence of volatility spurred from market transactions it may be unrealistic to assume non-trading days carry the same amount of volatility as trading days.

In Section 3.3.3 we present an example of daily stock data regression with GARCH residuals where we ignore gaps due to public holidays and weekends, however missing residuals are encountered due to missingness in the predictor series on trading days.

3.3.1.1 Diagnostics

As the SMC procedure does not yield point estimates of (σ_i^2), standardised residuals cannot be calculated and investigated. As an alternative, a diagnostic, proposed in Pitt et al. [2014], to assess the fitted model is to estimate the distribution functions $F(Y_i|Y_{1:i-1})$ for $i = 1, \ldots, n$ via

$$F(Y_i|Y_{1:i-1}) = \int F(Y_i|X_i)\mathbb{P}(dX_i|Y_{1:i-1}) \tag{3.11}$$

$$\approx \frac{1}{K}\sum_{k=1}^{K} F(Y_i|X_i^{(k)}) =: \hat{u}_i. \tag{3.12}$$

Then if the parameters and model are true, the estimated distribution functions should be independently uniformly distributed through time, so $\hat{u}_i \sim$ i.i.d $U(0,1)$ for $i = 1, \ldots, n$ as $K \to \infty$ (see Rosenblatt [1952]).

For the six missing data time series in Section 3.3.1, Figure 3.3 displays QQ-plots of the estimated $\hat{u}_i$ against the $U(0,1)$ distribution. All QQ-plots appear to be fairly linear, supporting $\hat{u}_i$ to be uniformly distributed for all six series.

As a visual for the volatility evolution one can plot quantiles from the ECDF of $\sigma_{t_i}^2|Y_{1:i}$ obtained at each time step $i = 1, \ldots, n$ of the SMC procedure evaluated at the fitted parameter estimates. Figure 3.4 displays the time series of the median as well as 5% and 95% quantiles from the continuous ECDF of $\sigma_{t_i}^2|Y_{1:i}$ for the series $R_i^{(AXJO)}$.

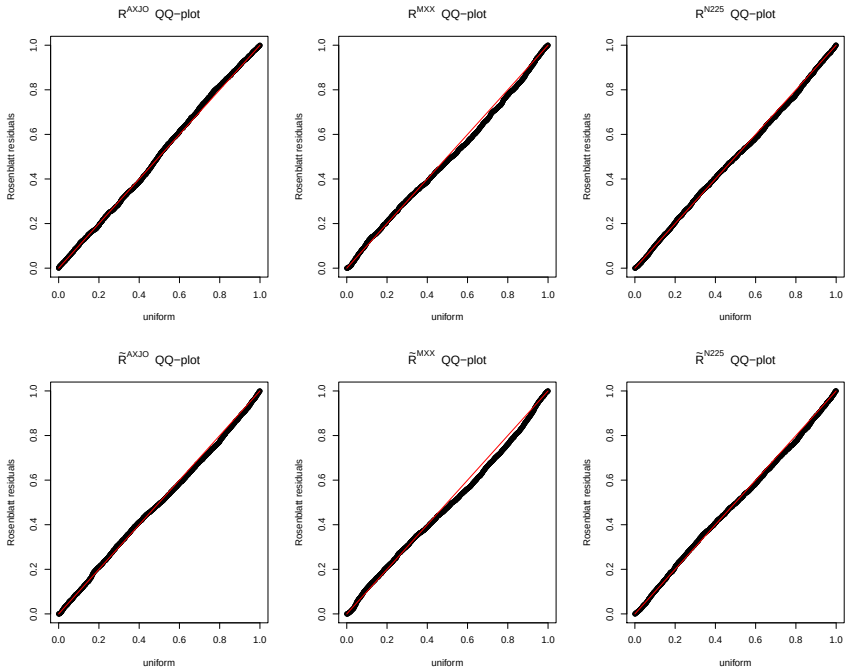


Figure 3.3: QQ-plots of the estimated $\hat{u}_i$ against the $U(0,1)$ distribution for each of the six return series.

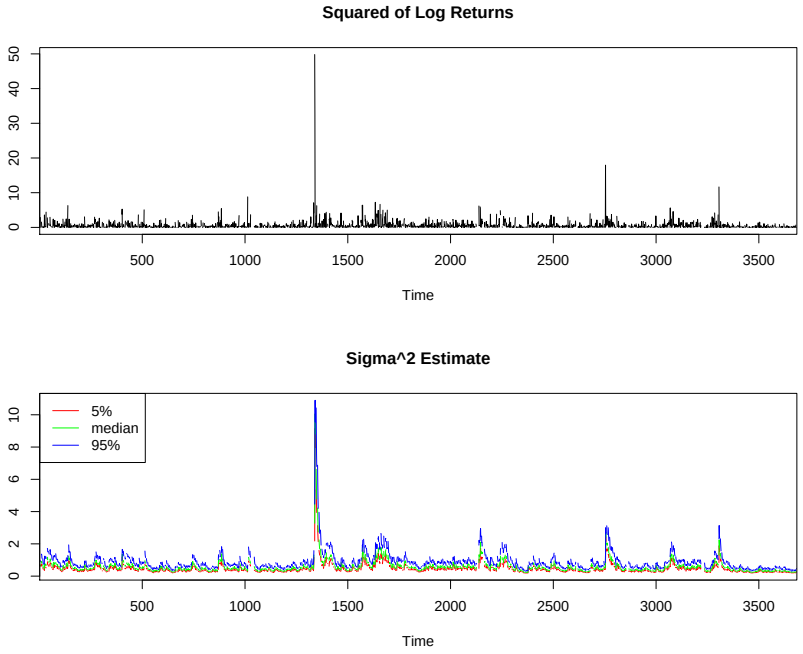


Figure 3.4: Top: squared log returns of the S&P/ASX 200 index (R_i^2) for the period 28th of February 1994 to the 31st of March 2004 . Bottom: the 5%, 50% and 95% quantiles from inverting the continuous ECDF of $\sigma_{t_i}^2 | Y_{1:i}$ obtained at each time step of the SMC algorithm.

3.3.2 Wind data

Cripps and Dunsmuir [2003] develop bivariate GARCH models for minute-by-minute North-South and East-West components of winds in Sydney Harbour in the context of understanding daily sailing conditions, particularly formation of the sea-breeze effect, in preparation for the Sydney Olympic games. In their analysis there were no missing observations on the days for which they performed the analysis. Modelling volatility in wind speeds is also of interest in understanding wind turbine power generation performance for intraday and interday management of wind farms (cf. Kay et al. [2009], Trombe et al. [2012] or Jeon and Taylor [2016]). Here we consider analysis of some 30 minute wind data over an approximately four and half month period for three locations in the Sydney region: North Head, Fort Denison and Penrith. Data were obtained from the Australian Bureau of Meteorology[1] and a summary of the data used here is in Table 3.7. This data was also used in the original Sydney Olympics modelling project as part of a wider study of synoptic conditions which favour sea-breeze formation.

We use the north-south component of wind speed at these three locations to illustrate our method for fitting GARCH models with missing data. The half hourly wind speed measurements were first modelled using AIC optimally selected autoregressive models of degree p (AR(p)) for $1 \leq p \leq 20$ using the `arima` command in R which allows for missing data in maximum likelihood estimation. Residuals from these optimal lag autoregressive fits (see Figure 3.5) were then modelled using the GARCH(1,1) model with standard normal innovations giving the results in Table 3.8. GARCH effects are observed in these series with parameter values that vary by location.

The analysis presented here treats the residual process from an autoregressive fit as the series of interest for GARCH modelling. However, the method of fitting the autoregression ignores GARCH effects in the innovations for that process. Ad-

[1] http://www.bom.gov.au/climate/data/stations/about-weather-station-data.shtml

ditionally, the autoregressive residuals will have volatility induced by the increased variability of prediction over gaps in the series which could confound the estimation of any underlying volatility in the innovation of the autoregression. Hence the analysis presented here is a two-step method. Ideally a method which combines the fitting of the autoregression and the GARCH model for innovations should be used. Although, as no such one-step method is known, the pragmatic two-step approach was employed.

In Chapter 4 however, methodology to perform simultaneous estimation of AR(1) models with GARCH(1,1) residuals will be introduced, for which we then use to reanalyse these series. This approach though, does not appear to be extendable to AR(p) models with GARCH(1,1) residuals for $p > 1$.

Table 3.7: Description of half hourly wind data at three locations in the Sydney Region. Gap lengths summarise the lengths of gaps in the observed record.

Site 66197 North Head Manly, 00:00 hrs Dec 1 1999 to 10:00 hrs Apr 20 2000,
Site 67113 Penrith Lakes, 00:00 hrs Dec 1 1999 to 10:00 hrs Apr 19, 2000,
Site 66022 Fort Denison, 00:00 hrs Dec 1 1999 to 12:30 hrs Apr 13 2000.

	North Head	Penrith	Fort Denison
N (length)	6798	6741	6458
Missing (%)	5%	5%	9%
Gap lengths	$1:9,222$	$1:8,13,18,213$	$1:6,30,82,191,222$

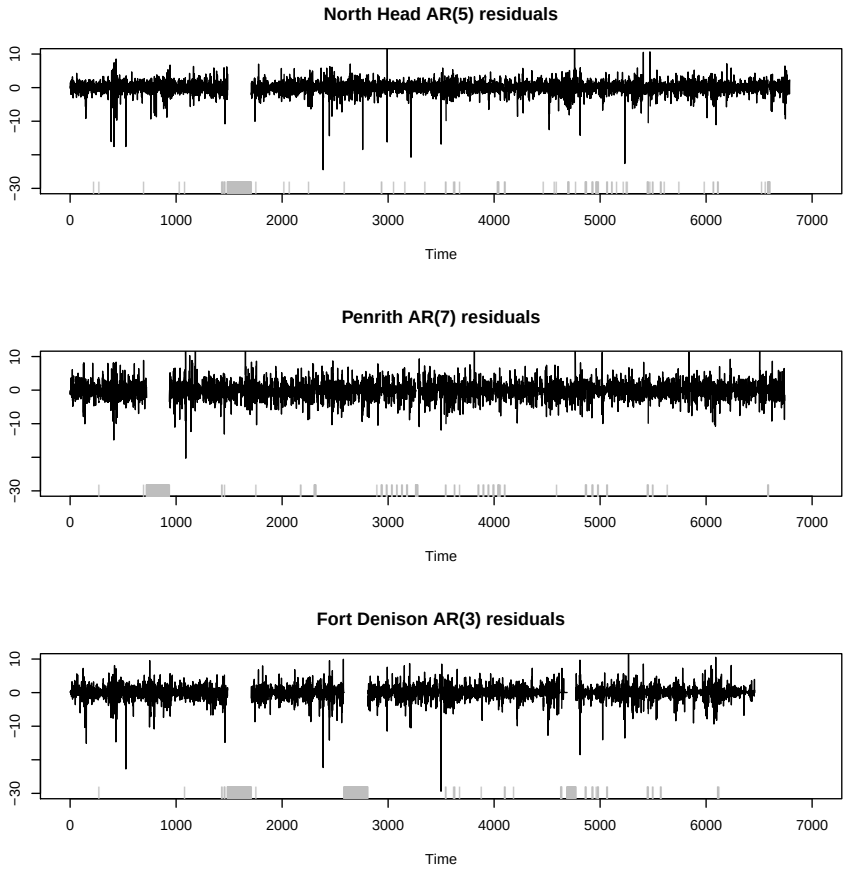


Figure 3.5: Residuals from autoregressive model fits at three locations in the Sydney region – North Head is located on land on the north of the entrance to Sydney Harbour, Fort Denison is located in the middle of the Sydney Main Harbour and Penrith is located inland west of Sydney at Lake Penrith and is not susceptible to sea-breeze effects. Plots show residuals from the best fitting AR(p) model in each case for the North-South wind speed component. The grey "rug" plot in each panel shows where data are missing.

Table 3.8: Estimated GARCH parameters with 95% confidence intervals for the wind data.

	ω	α	β
North Head	0.8930 (0.7291, 1.0568)	0.2610 (0.2180, 0.3040)	0.5316 (0.4664, 0.5968)
Penrith	0.9286 (0.6850, 1.1721)	0.1509 (0.1241, 0.1777)	0.6852 (0.6243, 0.7460)
Fort Denison	0.3017 (0.2246, 0.3788)	0.1825 (0.1470, 0.2179)	0.7779 (0.7391, 0.8168)

3.3.3 Stock data regression against incomplete weather data

To illustrate how missing data can arise in the GARCH setting when covariates are missing we use an example from research on the possible link between sunshine and stock returns. A number of studies such as Persinger [1975], Cunningham [1979] and Howarth and Hoffman [1984] have found that the number of hours of sunshine exposure is inversely correlated with a negative mood or pessimistic outlook. In the relatively new field of *behavioural finance* (cf. Shiller [2003]), it is considered that psychological factors can impact an individual's financial decision making.

These observations have led to empirical studies such as Saunders [1993], Hirshleifer and Shumway [2003] and Worthington [2009] that seek to test whether weather (in particular sunshine) has a significant impact on stock market returns. Saunders [1993] using daily returns from the Dow-Jones Industrial Average, New York Stock Exchange and American Stock Exchange found that on days when cloud cover (which is inversely related to sunshine) was 100 percent, returns were below average, while days below 20 percent cloud cover had above average returns. Hirshleifer and Shumway [2003] conducted an international study of the relationship between cloud cover and returns in 26 markets. Like Saunders [1993] a negative relationship was found to exist, but only in the three markets of Milan, Rio de Janeiro, and Vienna were these results statistically significant. Worthington [2009] for the period 1958 to 2005 performed a regression of the daily returns of the Australian All Ordinaries price index against eight daily weather variables, one of which being hours of sunshine. Heteroskedasticity in the residuals were detected and adjusting the standard errors by means of White [1980] and Newey and West [1987], Worthington [2009] found no statistical evidence that the weather variables influenced the Australian market returns.

Daily global solar exposure (DGSE) is the total amount of solar energy falling on a horizontal surface for a day. Typical values for DGSE range from above 0 to 35 MJ/m2 (megajoules per square metre). The values are usually highest in clear

sun conditions and lowest during very cloudy days. Thus one can use DGSE as a proxy for sunshine.

At this date, one can obtain freely from the Australian Bureau of Meteorology (ABOM) historical DGSE recorded from a number of weather stations. According to the ABOM website, missing observations can occur due to problems with the satellite or processing of the images used to estimate the solar exposure.

We obtain from *Yahoo! Finance* historical closing prices of the ASX200 index from the 4th of January 1993 to the 31st of December 2016. For the sake of this analysis, we do not treat weekends or public holidays as missing. We obtain from the ABOM DGSE recorded from the Observatory Hill weather station (the closest station to Sydney's central business district) for the same period as the ASX200 returns. For the trading days in that period, we have 6082 daily ASX log returns (R_i) and 5799 DGSE readings $(DSGE_i)$, corresponding to 4.65% of missing DGSE readings.

Figure 3.6 displays the average return for days with DGSE readings falling within specified MJ/m2 bands. Interestingly, days with DGSE $<$ 32 MJ/m2 have an average return over the period of 0.016% whereas the average return for days with DGSE $\geq$ 32 MJ/m2 is 0.19%, about 11.7 times larger than the latter. Whether this impact on returns is statistically significant is the next question. We fit the model

$$R_i = \mu + \mu_D \times \mathbb{I}(DSGE_i \geq 32 \text{ MJ/m2}) + \epsilon_i \tag{3.13}$$

and test if μ_D is statistically significant, however unlike Worthington [2009] we directly model heteroskedasticity taking (ϵ_i) as a GARCH(1,1) sequence with standard normal innovations. It is not our aim to perform as comprehensive an investigation as Worthington [2009] on the effects of weather on the Australian market, this example is just to provide illustration of a regression analysis with GARCH errors when some observations of the covariate series are missing leading to miss-

ingness in the residual series.

Table 3.9 provides the estimated μ, μ_D and GARCH parameters along with 95% confidence intervals. As zero is contained within the confidence band for μ_D, like Worthington [2009], we too find no statistical evidence of sunshine effecting Australian index returns.

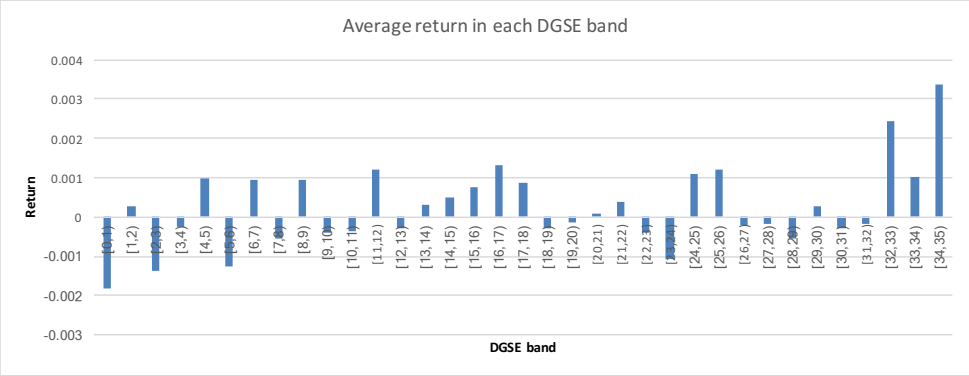


Figure 3.6: Average return for days with DGSE readings falling within the specified MJ/m2 bands for the trading days between 4th of January 1993 to the 31st of December 2016 for which DGSE was available.

Table 3.9: Estimated μ, μ_D and GARCH parameters along with 95% confidence intervals.

	Estimate	95% confidence intervals
$\omega \times 10^6$	1.2050	(0.7394, 1.6705)
α	0.0784	(0.0348, 0.1219)
β	0.9082	(0.902, 0.9144)
μ	0.0004	(0.0001, 0.0008)
μ_D	0.0011	(-0.0003, 0.0025)

3.3.4 High frequency CAPM

The capital asset pricing model (CAPM), developed by Sharpe [1964] and Lintner [1965], is the following linear model of the relationship between an asset's return R_i and the return of a market portfolio $R_{M,i}$

$$R_i - R_{f,i} = \alpha_{CAPM} + \beta_{CAPM}(R_{M,i} - R_{f,i}) + \epsilon_i \tag{3.14}$$

where $R_{f,i}$ is the risk free interest rate, (ϵ_i) is an i.i.d noise sequence, β_{CAPM} is a parameter reflecting the sensitivity of the asset's excess return (over the risk free rate) to the market's excess return and α_{CAPM} is a parameter reflecting any abnormal gain (if positive) or loss (if negative). The parameter α_{CAPM} can be used as a metric to evaluate the performance of an asset or portfolio relative to the market. The parameter β_{CAPM} is a measure of how volatile the asset is in comparison to the market.

Traditionally, the CAPM is performed on daily or even lower frequency returns, however the analysis here will be performed using high frequency one-minute returns. Furthermore, we drop the i.i.d assumption on (ϵ_i) modelling it as a GARCH(1,1) series with standard normal innovations. For convenience, in this illustration we ignore $R_{f,i}$ and perform the linear model on returns rather than excess returns.

One-minute price data for the S&P 500 index, Apple Inc. (AAPL), Boeing Co. (BA) and Chevron Corp. (CVX) was obtained from https://www.finam.ru/ for the period January 2015 to June 2015. The S&P 500 is taken as the market portfolio. We fit a high frequency CAPM model, regressing the market portfolio return multiplied by 100 against the respective stock return multiplied by 100, for each individual month in that period, skipping over the minutes when the market is closed, that is only minutes when the market is open that have missing price information will be treated as missing. It is well known (cf. Harris [1986]) that

intraday volatility in equity returns are higher around the open and close of the market than the middle of the trading day. As this is a trait common to both response and predictor we choose not to adjust out this effect.

For reasons unknown, there are missing values in these price series, Figure 3.7 displays the one-minute returns (multiplied by 100) for these price series during the month of February.

Tables 3.10, 3.11 and 3.12 are the respective, SMC obtained, estimates of the GARCH and CAPM parameters for the AAPL, BA and CVX stock assets. The right most column of those tables indicate the percentage of response and predictor pairs missing in the respective data sets. Figure 3.8 displays the time series of the estimated GARCH and CAPM parameters for the three stocks.

The β_{CAPM} estimates for AAPL in each month are close to 1, indication that this stock moves close in tandem with the market, which is not surprising given that AAPL is a relatively large constituent of the S&P 500 index[1]. BA which is a relatively small constituent of the S&P 500 index, is less sensitive to market moves with β_{CAPM} ranging from approximately 0.4 to 0.6. Interestingly CVX, a relatively moderate constituent of the S&P 500 index, has higher β_{CAPM} ranging approximately from 0.8 to 1.1 during the first quarter of 2015, but then appears to be less sensitive to the market for the second quarter with β_{CAPM} ranging approximately from 0.4 to 0.65.

In terms of α_{CAPM}, for the six months under consideration none of the assets have a statistically positive α_{CAPM}. AAPL for January, April and June along with BA for March and May had statistically negative α_{CAPM}.

Regarding the GARCH parameters; BA and CVX have relatively similar α and β parameters across the six months, while these parameters for AAPL are relatively similar across January to May, a comparatively larger α and lower β is exhibited for June. Besides February for AAPL, February and April for BA

[1] http://siblisresearch.com/data/market-caps-sp-100-us/

along with January and February for CVX, statistically significant α were found. In terms of the weighted long-run variance ω, common to the three stocks we see that January yields higher ω estimates than the other months in the study.

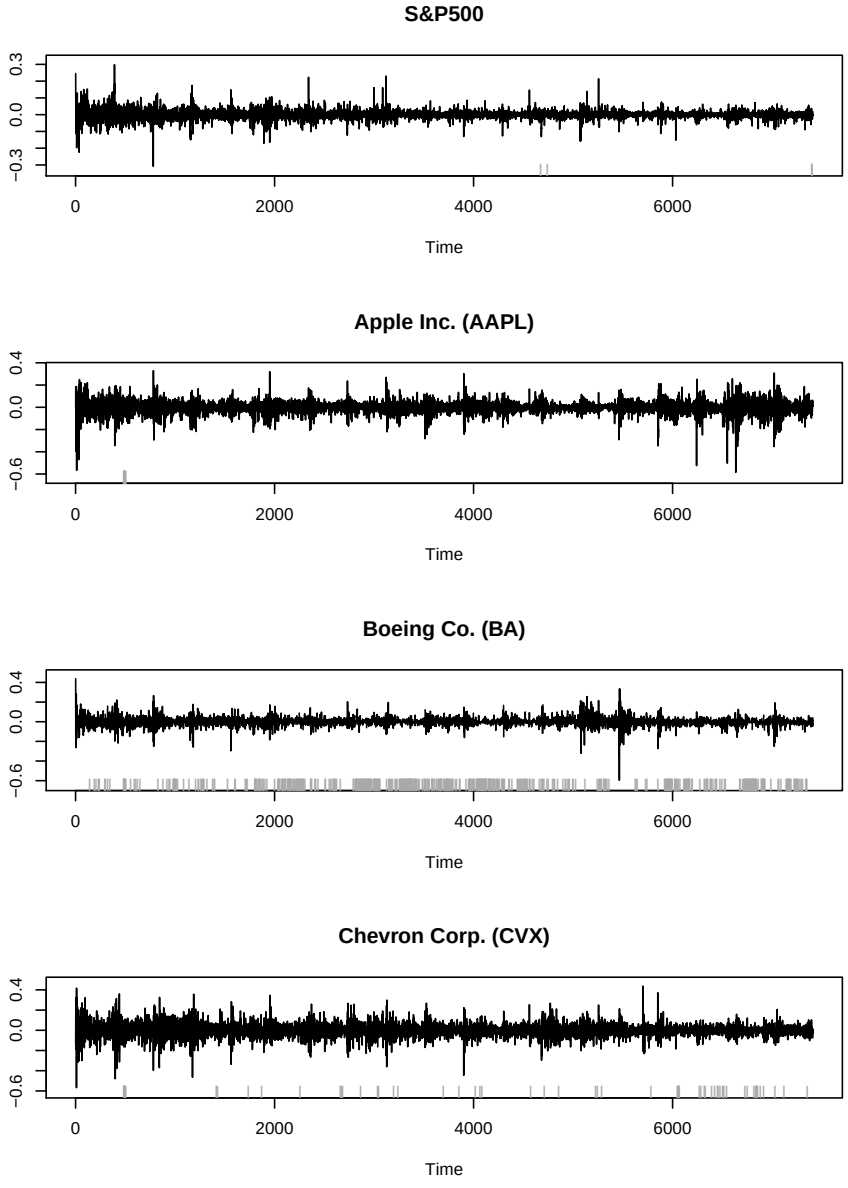


Figure 3.7: Time series of one-minute returns (multiplied by 100) for each trading minute in the month of February 2015 for the S&P 500 index, AAPL, BA and CVX. The grey "rug" plot in each panel shows where data are missing.

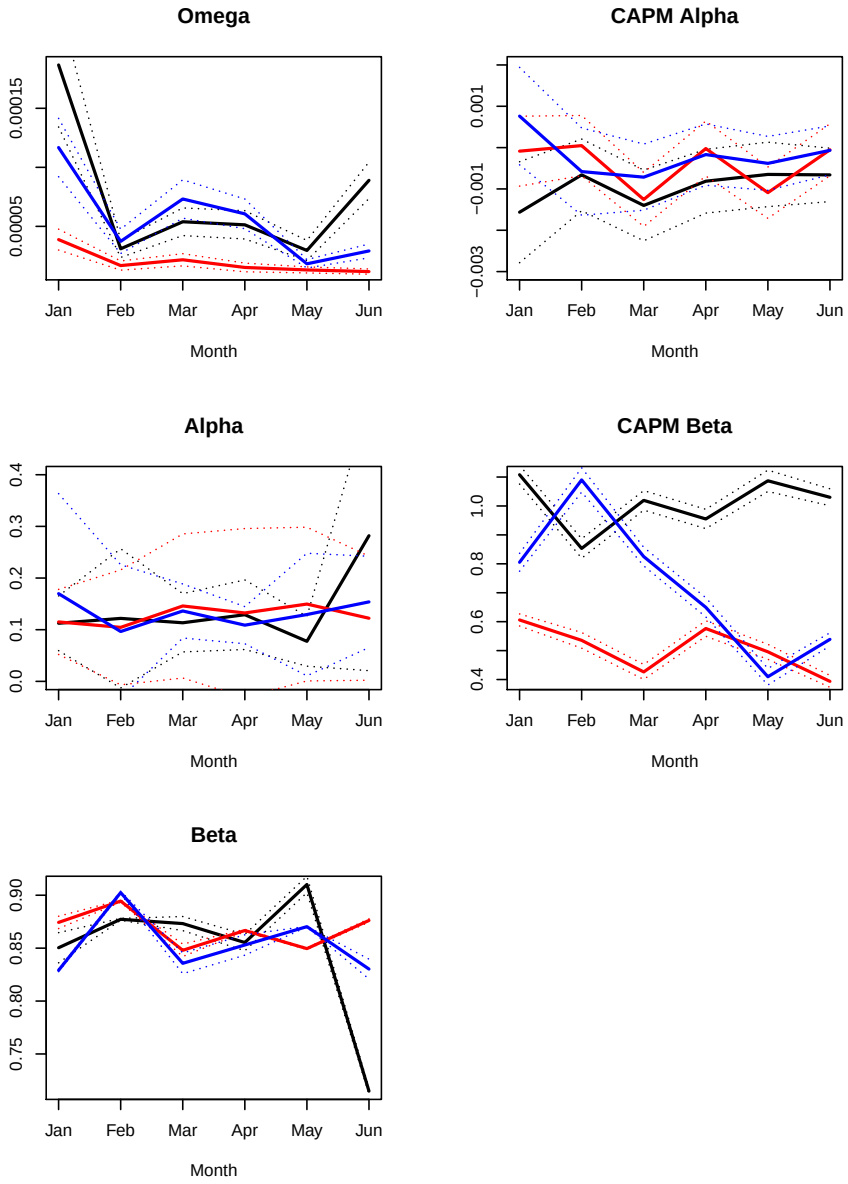


Figure 3.8: SMC estimates of GARCH and CAPM parameters for Apple Inc. (black), Boeing Co. (red) and Chevron Corp. (blue). The 95% confidence bands are indicated by the dotted lines.

Table 3.10: SMC estimates of GARCH and CAPM parameters along with their estimated standard errors (SE) for Apple Inc. (AAPL).

		$\omega \times 1000$	α	β	CAPM α	CAPM β	% Missing
Jan	Est	0.1867	0.1125	0.8504	-0.0016	1.1080	5.51%
	SE	0.0268	0.0270	0.0073	0.0006	0.0165	
Feb	Est	0.0310	0.1219	0.8773	-0.0007	0.8531	0.19%
	SE	0.0037	0.0687	0.0005	0.0004	0.0171	
Mar	Est	0.0539	0.1134	0.8732	-0.0014	1.0194	1.18%
	SE	0.0060	0.0288	0.0034	0.0004	0.0172	
Apr	Est	0.0513	0.1290	0.8553	-0.0008	0.9550	0.32%
	SE	0.0061	0.0344	0.0040	0.0004	0.0170	
May	Est	0.0296	0.0775	0.9101	-0.0006	1.0870	0.03%
	SE	0.0043	0.0245	0.0040	0.0004	0.0188	
Jun	Est	0.0890	0.2820	0.7149	-0.0007	1.0300	0.57%
	SE	0.0079	0.1333	0.0015	0.0003	0.0150	

Table 3.11: SMC estimates of GARCH and CAPM parameters along with their estimated standard errors (SE) for Boeing Co. (BA).

		$\omega \times 1000$	α	β	CAPM α	CAPM β	% Missing
Jan	Est	0.0388	0.1150	0.8743	-0.0001	0.6059	5.62%
	SE	0.0045	0.0321	0.0030	0.0004	0.0106	
Feb	Est	0.0168	0.1045	0.8945	0.0000	0.5354	7.68%
	SE	0.0020	0.0567	0.0005	0.0004	0.0138	
Mar	Est	0.0217	0.1459	0.8480	-0.0013	0.4263	18.43%
	SE	0.0026	0.0711	0.0030	0.0003	0.0132	
Apr	Est	0.0152	0.1324	0.8667	0.0000	0.5764	15.47%
	SE	0.0019	0.0834	0.0006	0.0003	0.0137	
May	Est	0.0131	0.1496	0.8495	-0.0011	0.4958	15.44%
	SE	0.0013	0.0759	0.0004	0.0003	0.0132	
Jun	Est	0.0117	0.1222	0.8761	-0.0001	0.3937	24.22%
	SE	0.0011	0.0612	0.0008	0.0003	0.0105	

Table 3.12: SMC estimates of GARCH and CAPM parameters along with their estimated standard errors (SE) for Chevron Corp. (CVX).

		$\omega \times 1000$	α	β	CAPM α	CAPM β	% Missing
Jan	Est	0.1168	0.1698	0.8291	0.0008	0.8053	1.04%
	SE	0.0126	0.0988	0.0007	0.0006	0.0157	
Feb	Est	0.0373	0.0965	0.9025	-0.0006	1.0900	0.93%
	SE	0.0053	0.0659	0.0007	0.0005	0.0214	
Mar	Est	0.0732	0.1362	0.8357	-0.0007	0.8250	2.28%
	SE	0.0083	0.0267	0.0050	0.0004	0.0163	
Apr	Est	0.0606	0.1088	0.8530	-0.0002	0.6496	3.79%
	SE	0.0064	0.0183	0.0049	0.0004	0.0169	
May	Est	0.0183	0.1293	0.8703	-0.0004	0.4096	6.33%
	SE	0.0018	0.0605	0.0002	0.0003	0.0146	
Jun	Est	0.0292	0.1538	0.8302	-0.0001	0.5389	6.75%
	SE	0.0030	0.0451	0.0048	0.0003	0.0115	

3.4 Concluding remarks

In summary, this section has illustrated use of the SMC methodology for providing a computationally feasible and reliable estimator (along with standard errors) to tackle the problem of parameter estimation of a partially observed GARCH(1,1) time series, outperforming common imputation methods, especially in regards to the persistence parameters α and β, when the proportion of missing observations is large. For the restricted case of a partially observed GARCH(1,1) series at equally spaced time points, the "Weak" GARCH approximation of Drost and Nijman [1993] is a very fast, for a small loss in efficiency, alternative.

While the analysis performed here have used standard normal innovations, other innovations that are easy to simulate and have a density function easy to evaluate can be simply substituted. It is straightforward to adapt the method for use with other GARCH variants such as the EGARCH(1,1) (Nelson [1991]), GJR-GARCH(1,1) (Glosten et al. [1992]), NGARCH(1,1) (Engle and Ng [1993]), QGARCH(1,1) (Sentana [1995]) etc.

While the general GARCH(p, q) models with $p > 1$ or $q > 1$ or both are of the form (1.8)-(1.9), the state vectors of these representations are not one-dimensional inhibiting the use of Pitt and Malik [2011]'s procedure. However, as "*the overwhelmingly most popular GARCH models in applications has been the GARCH(1, 1) model*" -Terasvirta [2009] methods for maximum likelihood estimation for this model are of considerable practical value.

With regards to the comment in Section 3.3.1 about the perceived difference in volatility behaviour between trading and non-trading days. One could explore this idea further by utilising the SMC procedure to fit a GARCH series that evolves switching between two sets of parameters, one for trading days and one for non-trading days, and gain an insight on the difference in volatility behaviour by comparing the two sets of estimated driving parameters.

Chapter 4

Applications: Likelihood inference for aggregated GARCH(1,1) series and partially observed AR(1)-GARCH(1,1) series

In some cases when dealing with missing observations it may be more appropriate to treat the data as aggregated rather than missing. To illustrate the difference between these two scenarios consider daily stock price returns. These are calculated from daily measurements of the stock price level usually taken at the same time each day. Now, if the market is shut on a particular day (say due to holidays or technical failure) then the returns on two successive days will be unavailable as a result of one day of missing stock price information.

For example, Bondon and Bahamonde [2012], in their analysis of fitting an ARCH model in the presence of missing observations used the Chilean IPSA stock index where for weekends they ignore the gap and treat returns from Friday to the next Monday as being on equal footing to daily returns from successive days in the working week. The IPSA stock index time series considered, denoted (P_i), covers

2869 calendar weekdays of which 127 are public holidays on which the Chilean stock exchange is closed and the IPSA index is not available. The percentage of missing data in the price series (P_i) is therefore equal to 4.427%. The actual series being analysed are the daily log returns (R_i) derived via $R_i = \ln P_i - \ln P_{i-1}$ and as a result of missing a single price P_i this will will mean two one day returns R_{i+1} and R_i will be missing, resulting in the percentage missing of log returns to be 8.438%.

Modelling of (R_i) calculated in this way, as was done in Bondon and Bahamonde [2012], corresponds to what we refer to as having "partial" observation of the series (R_i), or the series (R_i) with "missing" observations. This treatment of (R_i) was the convention utilised throughout Chapter 3.

However, when deriving returns from an incomplete price series (P_i) one can in fact utilise additional information from the underlying price series than was the case with "partial" observation. Suppose one were to observe the prices P_{i+1} and P_{i-1} but missed the price P_i, indeed R_{i+1} and R_i will be unknown, however as

$$\ln(P_{i+1}/P_{i-1}) = \ln(P_{i+1}/P_i) + \ln(P_i/P_{i-1}) = R_{i+1} + R_i \tag{4.1}$$

one in fact has record of the two-period log return, which is the aggregated value of the two missing one-period log returns. By taking log differences of successively available P_i one would observe an aggregated return over periods of missing price data, which corresponds to what we will refer to as having "aggregated" observation of the underlying series (R_i).

In the above case, both "partial" and "aggregated" observation treatments take R_{i+1} and R_i as unknown, however "aggregated" observation incorporates the available information that the aggregated value $R_{i+1} + R_i$ is observed, while "partial" observation ignores this piece of information.

In economics, measurements can be classified as either *stocks* or *flows* (c.f, Fisher [1896]). A stock measurement is taken at one specific time, and represents

a quantity existing at that point in time whereas a flow measurement is taken over an interval of time. As we have seen, if the quantity to be modelled is a flow measurement (ie. returns) which are derived from a stock measurement (ie. price) then one unnecessarily loses information utilising a "partial" observation treatment and thus would be recommended to use the "aggregated" observation treatment.

Although, if the quantity to be modelled is either a stock measurement or a flow measurement not derived from an underlying stock measurement, then the "partial" observation treatment is the only choice possible. For instance, if only the available one-period returns were provided to the practitioner, without access to the underlying price series, then the "partial" observation treatment is the only available route.

This chapter illustrates utilisation of SMC for obtaining parameter estimates of a GARCH(1,1) series from temporally aggregated observations, equivalent to partial observation of the cumulative sum series $(\sum_{j=1}^{i} y_j)$ of the underlying GARCH series $(y_i)_{i=1,\ldots,N}$. The task entails identifying a particular state space representation for ease of implementation of the SMC likelihood approximation from Section 2.1, as well as an extra step before application of the continuous resampling procedure from Section 2.3; these details are discussed in Section 4.1. Simulation studies are performed in Section 4.1.1 to assess the performance of this estimation method at varying levels of aggregation as well as for comparison against Drost and Nijman [1993]'s Weak GARCH formula for the special case when the data is consecutively aggregated in non-overlapping blocks of equal size. Utilisation of the method on a real world stock return application is illustrated in Section 4.1.2. In Section 4.2, an estimator for partially observed AR(1)-GARCH(1,1) series is presented, established through the notion that such a series can be seen as a general form of a temporally aggregated GARCH(1,1). A simulation study to test the performance of the AR(1)-GARCH(1,1) estimator at varying degrees of missingness

is conducted in Section 4.2.1 with a real world application of the method on wind data illustrated in Section 4.2.2. Finally Section 4.3 concludes with a discussion.

4.1 SMC for aggregated GARCH(1,1)

Assume $(y_i)_{i=1,\dots,N}$ is a GARCH(1,1) series and suppose one has as observations at time points $t_1 < \ldots < t_n = N$, where $t_j \in \{1, \dots, N-1\}$ for $j = 1, \dots, n-1$, the aggregated value of GARCH elements between times t_j and $t_{j-1}+1$ (with convention $t_0 \equiv 0$), that is $\breve{y}_j = \sum_{k=1}^{n_j} y_{t_{j-1}+k}$ where $n_j = t_j - t_{j-1}$.

Assuming one can simulate the noise z_j as well as evaluate its density q, we can again turn to SMC to approximate the otherwise intractable likelihood of $(\breve{y}_1, \dots, \breve{y}_n)$ by formulating the problem in the form of (1.8) – (1.9). Note however, when $n_j > 1$, $p(\breve{y}_j | \sigma^2_{t_{j-1}+1})$ is not analytical and thus $\sigma^2_{t_{j-1}+1}$ is not the best choice for X_j. Alternatively, consider for $j = 1, \dots, n$ the following 3-dimensional state process

$$X_j = \left(\sigma^2_{t_{j-1}+1}, \mathbb{I}(n_j > 1) \sum_{k=1}^{n_j - 1} z_{t_{j-1}+k} \sigma_{t_{j-1}+k}, \sigma^2_{t_j} \right). \tag{4.2}$$

Denote $X_{j,k}$ for $k = 1, \dots, 3$ the $k-th$ component of X_j. Now, taking for $j = 1, \dots, n$: $Y_j = \breve{y}_j$ and $Z_j = z_{t_j}$, we the have

$$Y_j = f(X_j, Z_j) = X_{j,2} + \sqrt{X_{j,3}} Z_j,$$

with the analytical quantity,

$$p(Y_j | X_j) = \frac{1}{\sqrt{X_{j,3}}} q\left(\frac{Y_j - X_{j,2}}{\sqrt{X_{j,3}}} \right). \tag{4.3}$$

Furthermore, with

$$W_j = \begin{cases} \{z_{t_{j-1}+1}, \ldots, z_{t_{j-1}+n_j-1}\} & \text{when } n_j > 1, \\ \emptyset & \text{when } n_j = 1, \end{cases} \tag{4.4}$$

we then have

$$X_j = g(X_{j-1}, Y_{j-1}, W_j) = \Big(g_1(X_{j-1}, Y_{j-1}), g_2(X_{j-1}, Y_{j-1}, W_j), g_3(X_{j-1}, Y_{j-1}, W_j)\Big)$$

with,

$$g_1(X_{j-1}, Y_{j-1}) = \omega + \alpha(Y_{j-1} - X_{j-1,2})^2 + \beta X_{j-1,3} \tag{4.5}$$

$$g_2(X_{j-1}, Y_{j-1}, W_j) = \mathbb{I}(n_j > 1) z_{t_{j-1}+1} \sqrt{g_1(X_{j-1}, Y_{j-1})} \tag{4.6}$$

$$+ \mathbb{I}(n_j > 2) \sum_{k=2}^{n_j-1} z_{t_{j-1}+k} \sqrt{\tilde{g}(X_{j-1}, Y_{j-1}, W_j, k)} \tag{4.7}$$

$$g_3(X_{j-1}, Y_{j-1}, W_j) = \mathbb{I}(n_j = 1) g_1(X_{j-1}, Y_{j-1})$$

$$+ \mathbb{I}(n_j > 1) \tilde{g}(X_{j-1}, Y_{j-1}, W_j, n_j) \tag{4.8}$$

where for $n_j > 1$ we define for $k = 2, \ldots, n_j$

$$\tilde{g}(X_{j-1}, Y_{j-1}, W_j, k) = \omega + \mathbb{I}(k > 2)\omega \Big(\sum_{i=2}^{k-1} \prod_{l=1}^{i-1} (\alpha z_{t_{j-1}+k-l}^2 + \beta) \Big)$$

$$+ g_1(X_{j-1}, Y_{j-1}) \prod_{i=1}^{k-1} (\alpha z_{t_{j-1}+i}^2 + \beta). \tag{4.9}$$

By construction $\{X_j\}$ is a vector Markov process. Using this alternative specification of the state (4.2), (4.3) is straightforward to evaluate, enabling implementation of the SMC likelihood approximation detailed in Section 2.1. Although, as the state is now a 3-dimensional quantity, the procedure in Section 2.3 for resampling the hidden state in a continuous manner cannot be directly applied.

However, as the model structure can be depicted as in Figure 4.1, we see that in fact a one dimensional quantity $X_{j,1}$ is sufficient to propagate the process and thus we really only need to find a way to continuously resample the $X_{j,1}$.

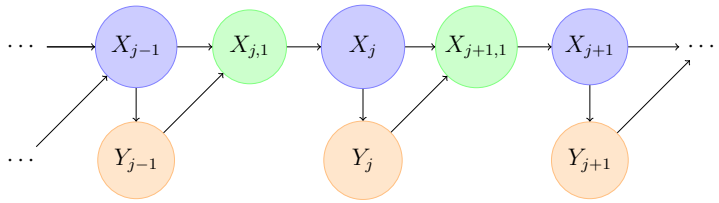


Figure 4.1: Aggregated GARCH model structure.

To begin, we first construct an empirical cumulative distribution of $X_{j+1,1}$ from the particles $\{X_j^{(k)}\}_{k=1,\dots,K}$ as

$$\hat{\mathbb{Q}}(X_{j+1,1} \leq x | Y_{1:j}) = \sum_{k=1}^{K} w_k \mathbb{I}(g_1(X_j^{(k)}, Y_j) \leq x) \tag{4.10}$$

where $w_k := \hat{\mathbb{P}}(X_j^{(k)} | Y_{1:j})$. Now, since the weights w_k are continuous in X_j and not $X_{j+1,1}$, even if $X_{j+1,1}^{(k)} = g_1(X_j^{(k)}, Y_j)$ and $X_{j+1,1}^{(k+1)} = g_1(X_j^{(k+1)}, Y_j)$ are close together, this does not guarantee that the weights w_k and w_{k+1} will be. Due to this, one then should not directly apply the continuous approximation of Section 2.3 to (4.10) with an illustration as to why provided in Figure 4.2.

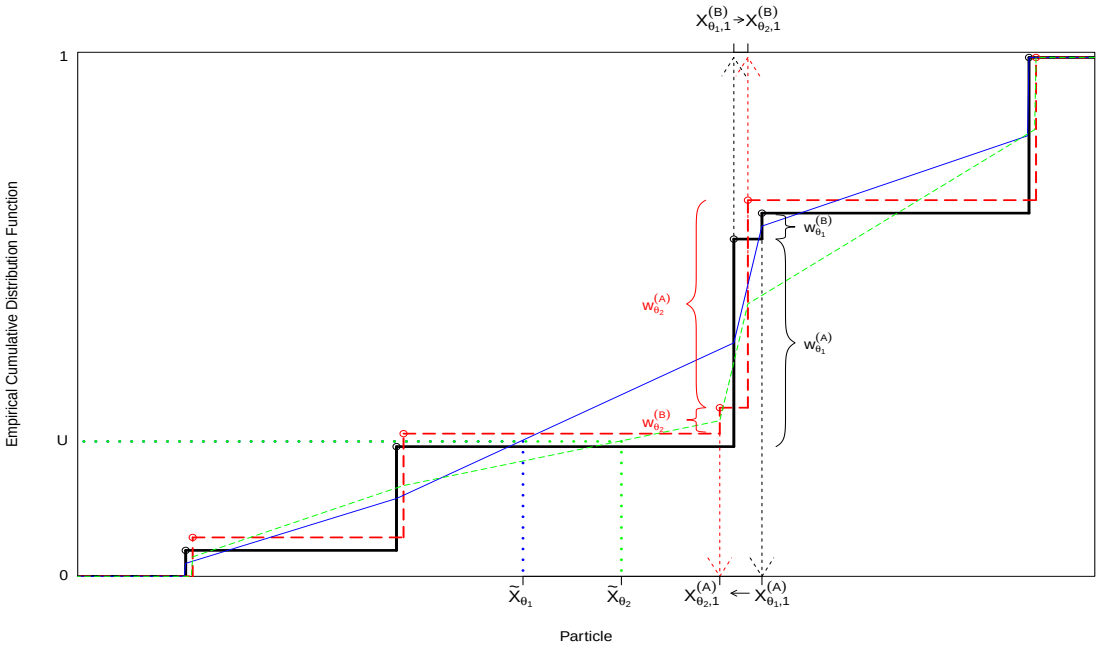


Figure 4.2: Assume the black and the red-dashed step functions are (4.10) constructed respectively with parameters θ_1 and θ_2 that are close to one another. Assume we have two particles denoted by $X_{\theta,1}^{(A)}$ and $X_{\theta,1}^{(B)}$ which are geographically close to one another but with respective weights $w_\theta^{(A)}$ and $w_\theta^{(B)}$ which are quite different in magnitude. Suppose that under θ_1 it is the case $X_{\theta_1,1}^{(A)} > X_{\theta_1,1}^{(B)}$, but for a small parameter change to θ_2 it is then the case that $X_{\theta_2,1}^{(A)} < X_{\theta_2,1}^{(B)}$, that is the order of particles has switched after the parameter change. As the weights are continuous in θ we will have $w_{\theta_1}^{(A)} - w_{\theta_1}^{(B)} \approx w_{\theta_2}^{(A)} - w_{\theta_2}^{(B)}$, that is the magnitude of difference of the weights still remains roughly the same after the particles have switched order, however a large structural change in the CDF functions has occurred; under θ_1 the CDF first increases by $w_{\theta_1}^{(A)}$ and then $w_{\theta_1}^{(B)}$, while under θ_2 the CDF now increases by $w_{\theta_2}^{(B)}$ before increasing by $w_{\theta_2}^{(A)}$. The blue function is the continuous approximation to the black step function and the green-dashed function is the continuous approximation to the red-dashed step function. As can be seen the structural change after the crossover carries through to the continuous approximations which can lead to large differences in resampled particles across small parameter changes (see $\tilde{X}_{\theta_1}$ and $\tilde{X}_{\theta_2}$ in the figure). The fact that $w_\theta^{(A)}$ and $w_\theta^{(B)}$ can be far apart despite $X_{\theta,1}^{(A)}$ and $X_{\theta,1}^{(B)}$ being in close proximity to one another is the root of the issue, in that if $w_\theta^{(A)} \approx w_\theta^{(B)}$ then essentially no structural change would result from a crossover (see Figure 4.3).

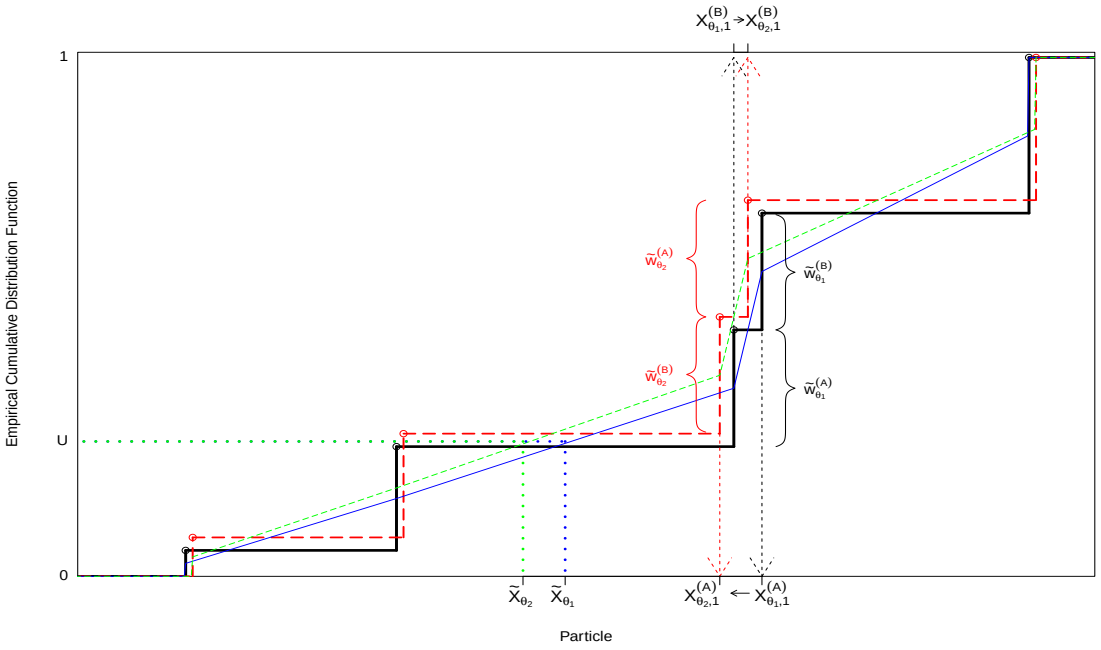


Figure 4.3: In contrast to Figure 4.2, when the weights $\tilde{w}_\theta^{(A)}$ and $\tilde{w}_\theta^{(B)}$ attached to respective particles $X_{\theta,1}^{(A)}$ and $X_{\theta,1}^{(B)}$ that are close in proximity to each others are also close in magnitude, then essentially no structural change would result if the particles $X_{\theta,1}^{(A)}$ and $X_{\theta,1}^{(B)}$ switched order under a parameter change.

In view of Figures 4.2 and 4.3 an operation to smooth the weights of (4.10) such that particles that are in close proximity to one another have attached weights of similar magnitude must be performed. The following kernel smoothing approach was proposed in Pitt and Malik [2011]

$$\bar{w}_k = \frac{\sum_{l=1}^{K} w_l \phi\Big((X_{j+1,1}^{(l)} - X_{j+1,1}^{(k)})/h\Big)}{\sum_{l=1}^{K} \phi\Big((X_{j+1,1}^{(l)} - X_{j+1,1}^{(k)})/h\Big)} \tag{4.11}$$

$$\tilde{w}_k = \frac{\bar{w}_k}{\sum_{l=1}^{K} \bar{w}_l} \text{ for } k = 1, \ldots, K \tag{4.12}$$

where $\phi(.)$ is the standard Gaussian density and $h = c/K$ for a very small number c. One then can proceed via the method in Section 2.3 to obtain a continuous approximation to

$$\hat{\mathbb{P}}(X_{j+1,1} \leq x | Y_{1:i}) = \sum_{k=1}^{K} \tilde{w}_k \mathbb{I}(X_{j+1,1}^{(k)} \leq x) \tag{4.13}$$

to continuously resample the $X_{j+1,1}$.

4.1.1 Simulation studies

4.1.1.1 Impact of level of aggregation on SMC performance

We define the term *aggregation percentage* (AP) to be

$$1 - \frac{\text{Total number of aggregation windows}}{\text{Total number of underlying observations}}. \tag{4.14}$$

In this section we test the performance of the SMC estimation method at varying aggregation percentage levels, while keeping the total number of aggregation windows at 2000 in order to study the impact of aggregation percentage in isolation from any impact on performance attributable to smaller sample sizes. For a given AP, we simulate 1000 different data sets of a length $N = \lfloor \frac{2000}{1-AP} \rfloor$ GARCH(1,1) series with standard normal innovations with true parameters $\omega = 0.1, \alpha = 0.08$ and $\beta = 0.9$, then for each data set we set as the last aggregation window end point $t_n = N$, where $n = 2000$ and the remaining $n - 1$ aggregation window end points chosen with equal probability (without replacement) from the set $\{1, \ldots, N - 1\}$. Note that the configuration of aggregations is different between the 1000 data sets.

Parameters were again optimised over the set defined in (3.4) using an integrated C++ and R implementation. Table 4.1 displays the results from utilising 250 and 1000 particles for 1000 simulations of the GARCH model at aggregation percentages of 5%, 10%, 20%, 35% and 50%. Refer to Section 3.2.1 for definitions of the summary statistics (**Mean, Bias** etc.) for the table. Some points about the table are:

- The average estimated standard error closely approximates the observed standard deviation suggesting that the use of the Hessian for obtaining standard errors is reliable.

- Across all three parameters the RMSE is primarily due to variance with

squared bias being small in comparison.

- The bias as a percentage of true value is largest for estimation of ω and is essentially negligible for estimation of α and β.

- Across all three parameters, RMSE is improved using more particles. The improvement is more pronounced at higher AP levels. At the 50% AP level, RMSE is roughly 10% lower for all parameters by increasing particles from 250 to 1000.

- Again, similar to Section 3.2.1, RMSE of ω decreases, as AP increases (while keeping n fixed), driven for the case of 250 particles by lower standard deviations and for the case of 1000 particles by both lower standard deviations and bias.

- In terms of RMSE for α and β, as AP increases (while keeping n fixed), at 250 particles there is a slight increase, while at 1000 particles RMSE is stable.

In summary, SMC utilising a modest 250 particles even at high AP levels is shown to be a reliable estimator, with the ability to improve estimation performance by utilisation of more particles.

Computational times using an Intel Xeon 3.06 GHz processor are provided in Table 4.2. The time to complete a SMC likelihood evaluation can be seen to be linear in N as well as the number of particles K.

Remark 4.1.1.1

Exact calculation of (4.11)-(4.12) would require proportionately K^2 operations. For computational efficiency, following Appendix 4 of Pitt and Malik [2011], the densities $\phi()$ were truncated at values where $|X^{(l)}_{j+1,1} - X^{(k)}_{j+1,1}| > 3h$, delivering the (almost) linear in K SMC likelihood evaluation times observed.

Table 4.1: SMC estimation of a GARCH(1,1) model from aggregated observations: Summary statistics from 1000 replications for each level of aggregation.

	250 Particles					1000 Particles				
AP	5%	10%	20%	35%	50%	5%	10%	20%	35%	50%
$\omega = 0.1$										
Mean	0.1186	0.1182	0.1176	0.1170	0.1175	0.1184	0.1177	0.1165	0.1150	0.1129
Bias	0.0186	0.0182	0.0176	0.0170	0.0175	0.0184	0.0177	0.0165	0.0150	0.0129
SD	0.0414	0.0424	0.0400	0.0402	0.0393	0.0413	0.0421	0.0395	0.0392	0.0359
Mean SE	0.0379	0.0371	0.0354	0.0328	0.0301	0.0380	0.0374	0.0360	0.0341	0.0319
RMSE	0.0454	0.0462	0.0437	0.0437	0.0430	0.0452	0.0457	0.0428	0.0420	0.0381
$\alpha = 0.08$										
Mean	0.0802	0.0801	0.0803	0.0817	0.0830	0.0801	0.0800	0.0800	0.0808	0.0809
Bias	0.0002	0.0001	0.0003	0.0017	0.0030	0.0001	<0.0001	<0.0001	0.0008	0.0009
SD	0.0131	0.0134	0.0134	0.0139	0.0146	0.0131	0.0133	0.0134	0.0137	0.0136
Mean SE	0.0134	0.0134	0.0133	0.0133	0.0131	0.0135	0.0135	0.0134	0.0134	0.0133
RMSE	0.0131	0.0134	0.0134	0.0140	0.0149	0.0131	0.0133	0.0134	0.0137	0.0136
$\beta = 0.9$										
Mean	0.8951	0.8954	0.8954	0.8946	0.8936	0.8952	0.8955	0.8958	0.8957	0.8962
Bias	-0.0049	-0.0046	-0.0046	-0.0054	-0.0064	-0.0048	-0.0045	-0.0042	-0.0043	-0.0038
SD	0.0179	0.0184	0.0183	0.0192	0.0195	0.0179	0.0183	0.0182	0.0189	0.0181
Mean SE	0.0179	0.0178	0.0175	0.0171	0.0166	0.0179	0.0179	0.0178	0.0176	0.0174
RMSE	0.0185	0.0189	0.0189	0.0199	0.0205	0.0185	0.0188	0.0187	0.0194	0.0185

Table 4.2: Summary, across the 1000 simulations for each level of aggregation, of computational times for the SMC optimisation using an Intel Xeon 3.06 GHz.

		250 particles			1000 particles		
AP	N	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)
5%	2105	23.69	222.47	0.106	93.61	211.31	0.443
10%	2222	27.23	238.45	0.114	105.87	223.37	0.474
20%	2500	33.94	261.87	0.130	131.45	242.89	0.541
35%	3076	44.22	277.10	0.160	172.79	258.02	0.670
50%	4000	60.35	303.10	0.199	234.79	278.33	0.844

4.1.1.2 Comparison with "Weak" GARCH approximation

Drost and Nijman [1993] established that if $(\epsilon_i)_{i=1,\dots,N}$ is a weak GARCH series with parameters ω, α, β and unconditional kurtosis κ then the m-period aggregation of that series denoted by $(\breve{\epsilon}_i)_{i=1,\dots,\lfloor \frac{N}{m} \rfloor}$, whereby $\breve{\epsilon}_i := \sum_{j=1}^{m} \epsilon_{(i-1)m+j}$, is also a weak GARCH series with unconditional kurtosis

$$\breve{\kappa} = 3 + \frac{\kappa - 3}{m} + 6(\kappa - 1)\frac{(m - 1 - m(\alpha+\beta) + (\alpha+\beta)^m)(\alpha - \beta\alpha(\alpha+\beta))}{m^2(1-\alpha-\beta)^2(1-\beta^2-2\beta\alpha)} \tag{4.15}$$

and parameters

$$\breve{\omega} = m\omega\frac{1-(\alpha+\beta)^m}{1-\alpha-\beta} \tag{4.16}$$

$$\breve{\alpha} = (\alpha+\beta)^m - \breve{\beta} \tag{4.17}$$

with $-1 < \breve{\beta} < 1$ as the solution to

$$\frac{\breve{\beta}}{1+\breve{\beta}^2} = \frac{\breve{a}(\alpha+\beta)^m - \breve{b}}{\breve{a}(1+(\alpha+\beta)^{2m}) - 2\breve{b}} \tag{4.18}$$

where

$$\begin{aligned}\breve{a} = {} & m(1-\beta)^2 + 2m(m-1)\frac{(1-\alpha-\beta)^2(1-\beta^2-2\alpha\beta)}{(\kappa-1)(1-(\alpha+\beta)^2)} \\ & + 4\frac{(m-1-m(\alpha+\beta)+(\alpha+\beta)^m)(\alpha-\alpha\beta(\alpha+\beta))}{1-(\alpha+\beta)^2}\end{aligned} \tag{4.19}$$

$$\breve{b} = (\alpha - \beta\alpha(\alpha+\beta))\frac{1-(\alpha+\beta)^{2m}}{1-(\alpha+\beta)^2}. \tag{4.20}$$

Thus, similarly as the case in Section 3.2.3, one can approximate the parameters of the higher frequency standard GARCH series $(y_i)_{i=1,\dots,N}$ from its m-period aggregation weak GARCH series $(\breve{y}_i)_{i=1,\dots,\lfloor \frac{N}{m} \rfloor}$ (whereby $\breve{y}_i = \sum_{j=1}^{m} y_{(i-1)m+j}$) utilising (4.16)-(4.20).

Below, we compare the performance of this approximation against that of the SMC estimates. We generate 1000 data sets of a GARCH(1,1) series with standard normal innovations and true parameters $\omega = 0.1, \alpha = 0.08$ and $\beta = 0.9$, taking $N = 10000$ and $m = 5$ which leads to an aggregation percentage of 80%.

Results from estimating the observed weak GARCH series by QML and then applying (4.16)-(4.20) to ascertain the underlying GARCH series, this process hereafter referred to as Drost QML, are displayed in Table 4.3. Large biases, in particular for the persistence parameters α and β are exhibited.

In light of these biases, before proceeding to compare the Drost approximation to the SMC estimates, we mention that while Drost and Nijman [1993] utilised the QML in weak GARCH estimation and found from their empirical experiments that "*the asymptotic bias of the QML, if there is any, is very small*", Nijman and Sentana [1996] found it to be the case that "*the QML estimator is approximately consistent in some cases and clearly inconsistent in others*". In Francq and Zakoian [2000] it was established that the weak GARCH "*estimators computed from the QML equations will unfortunately be inconsistent in general*".

Thus, in an attempt to see if the large biases in α and β are resultant from biased estimates of $\breve{\alpha}$ and $\breve{\beta}$ using the QML we also employ the least squares weak GARCH estimator (LSE) of Francq and Zakoian [2000] for which they proved to be a consistent and asymptotically normal estimator. The results from estimating the weak GARCH parameters by LSE and then application of (4.16)-(4.20), this process hereafter referred to as Drost LSE, are also displayed in Table 4.3. As stated by Francq and Zakoian [2000], the LSE "*is likely to be inefficient relative to the QML estimator*" and we do indeed notice this in Table 4.3 with the very large sample standard deviations of the Drost LSE estimates compared to the Drost QML. Furthermore, the same direction of bias using the consistent LSE as the *inconsistent in general* QML is observed and at an even higher level of magnitude. Thus, employing a consistent LSE does not eliminate the biases in the persistence

parameters α and β for this finite sample size either.

As mentioned in the previous chapter, to our knowledge, the appropriate covariance estimator for the QML applied to weak GARCH estimation has not been developed. Regarding the LSE, details for constructing the asymptotic covariance matrix have been provided in Francq and Zakoian [2000], for which approximate standard errors of $\breve{\omega}, \breve{\alpha}$ and $\breve{\beta}$ can be obtained, with application of the delta method (cf. Oehlert [1992]) required to obtain the standard errors of the higher frequency GARCH parameters ω, α and β. While estimated standard errors for the Drost LSE are in theory possible, we do not pursue this process as the focus of this section is on empirical performance of the Drost approximation to the SMC estimation method.

As the Drost QML is shown above to be a less biased and lower variance estimator than the Drost LSE we proceed to compare the performance of the SMC estimator to the Drost QML. The SMC estimates using 250, 1000 and 2000 particles are also displayed in Table 4.3. It is seen that even at 250 particles the SMC method produces superior estimates in terms of bias and variance for α and β compared to the Drost QML. As the number of particles increases, the RMSE of the SMC estimates improves, driven by both improvements in bias and standard deviation. The Drost QML however does produce superior estimates of ω compared to the SMC using 250 and 1000 particles. Although, increasing the number of particles to 2000 the SMC then overtakes the Drost QML in RMSE performance for ω.

Regarding performance of approximate standard errors for the SMC method, it appears at this high level of aggregation, 1000 particles are not enough to accurately estimate the variance observed in estimates across simulates. Utilising 2000 particles, better alignment between SD and Mean SE is observed.

Box plots of the Drost QML and SMC estimates are displayed in Figure 4.4. We see that the Drost QML estimates the parameter ω well, however has difficulty

accurately estimating the persistence parameters α and β. SMC on the other hand, even at a modest 250 particles generally does well estimating α and β with performance improving as more particles are utilised. In terms of ω however, it appears 250 particles may be insufficient requiring 2000 particles to best the Drost QML.

Using an Intel Xeon 3.06 GHz processor the Drost QML optimisation takes roughly 2 seconds, with computation times for the SMC optimisation provided in Table 4.4. While the Drost approximation is considerably faster, it yields unreliable estimates of α and β. The parameter α determines the reaction to market shocks with Alexander [2008] describing the parameter as the *volatility-of-volatility* and characterising markets that exhibit high α (>0.1) as being nervous or jumpy. The parameter β is a measure of how long elevation due to shocks persist in the variance.

The Drost approximation overestimates α and underestimates β considerably more than the SMC method. Overestimating α and underestimating β results in a comparatively more spiky volatility process than is otherwise the case. Not being able to accurately assess the impact of market shocks to future volatility, in particular not being able to distinguish between a volatility process where the impact of market shocks are large but short lived versus a volatility process in which the same market shock would lead to a less initial elevation in volatility but whose effects are longer lasting, can lead to different perceived effects when making financial decisions. For instance regarding options, if it is perceived that a market shock has a large but short lived impact on future volatility, one may have the impression that long dated option positions in one's portfolio are less impacted than is otherwise the case and short dated option positions are more impacted than is otherwise the case.

Table 4.3: Comparison of the estimation performance, across 1000 simulated data sets, of the Drost approximation against the SMC method utilising 250, 1000 and 2000 particles.

	SMC 250			SMC 1000			SMC 2000		
	ω	α	β	ω	α	β	ω	α	β
Mean	0.1507	0.1024	0.8701	0.1203	0.0882	0.8884	0.1133	0.0850	0.8925
Bias	0.0507	0.0224	-0.0299	0.0203	0.0082	-0.0116	0.0133	0.0050	-0.0075
SD	0.1077	0.0287	0.0464	0.0643	0.0195	0.0302	0.0357	0.0157	0.0208
RMSE	0.1190	0.0364	0.0552	0.0674	0.0211	0.0324	0.0381	0.0165	0.0221
Mean SE	0.0344	0.0139	0.0184	0.0286	0.0135	0.0174	0.0293	0.0138	0.0180

	Drost QML			Drost LSE		
	ω	α	β	ω	α	β
Mean	0.1142	0.1322	0.8447	0.1332	0.1488	0.8243
Bias	0.0142	0.0522	-0.0553	0.0332	0.0688	-0.0757
SD	0.0382	0.0935	0.0956	0.0973	0.1805	0.1906
RMSE	0.0407	0.1071	0.1105	0.1028	0.1932	0.2051

Table 4.4: Summary of the computational times on an Intel Xeon 3.06 GHz processor, across the 1000 simulated data sets, of the SMC optimisation utilising various particle sizes.

Number of particles	250	1000	2000
Time to complete optimisation (secs)	153.37	585.34	1144.14
Avg. number of function calls	362.38	332.48	319.32
Avg. time to evaluate function (secs)	0.42	1.76	3.58

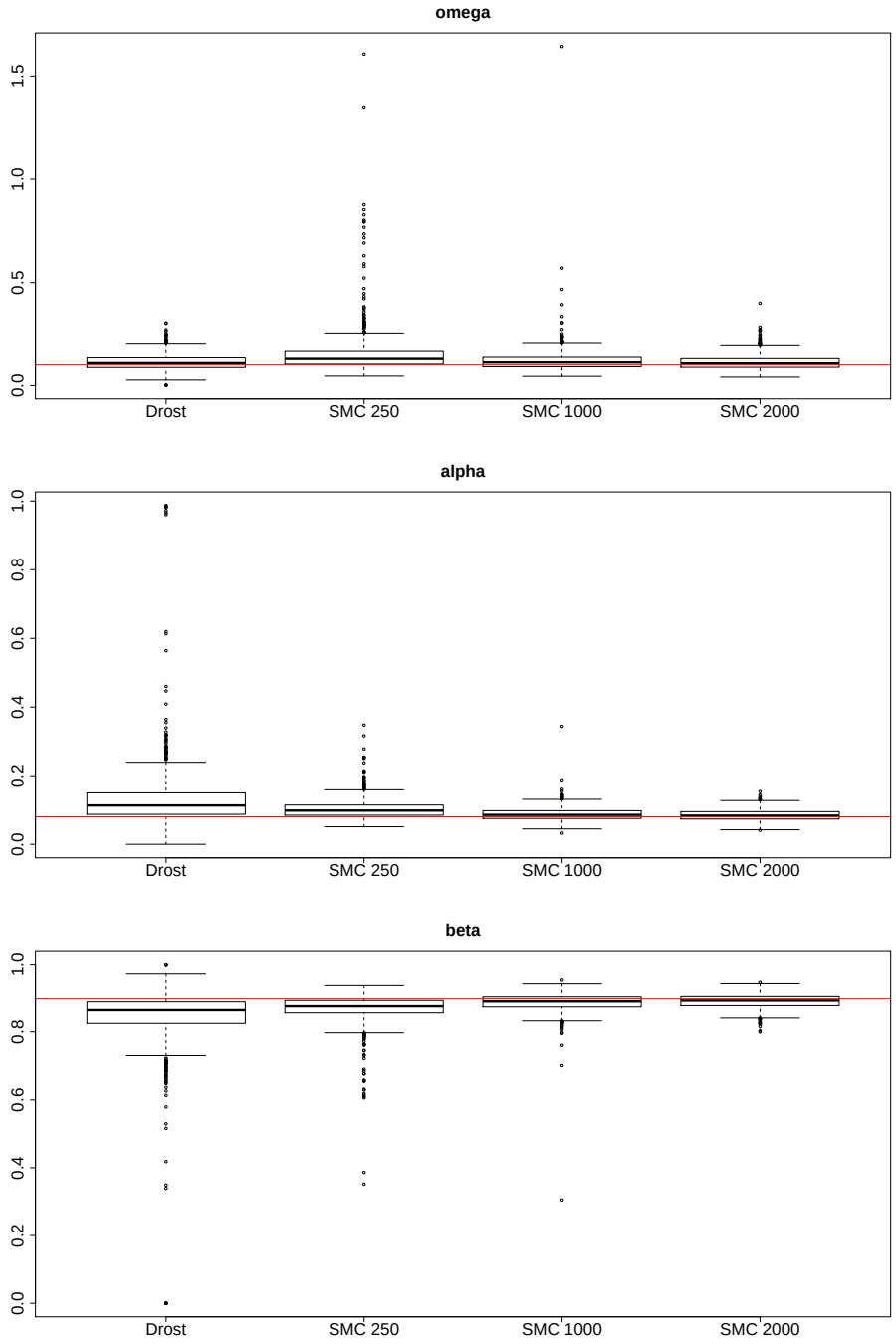


Figure 4.4: Boxplot summaries of the estimates obtained, from 1000 simulated data sets, from utilising the Drost QML and SMC method with 250, 1000 and 2000 particles.

4.1.2 Real data analysis

For each of the price series $(P_i^{(AXJO)})$, $(P_i^{(MXX)})$, $(P_i^{(N225)})$, $(\tilde{P}_i^{(AXJO)})$, $(\tilde{P}_i^{(MXX)})$ and $(\tilde{P}_i^{(N225)})$ defined in Section 3.3.1 we calculate the corresponding aggregated log return series between successively available prices. Details summarising the occurence of time spacings between sucessively available observations encountered in each price series are provided in Table 4.5. Regarding calender time spacings (CS), which is associated with the time convention of P_i, a regular weekend without additional public holidays corresponds to $n_i = 3$. Whereas for the time scale that ignores weekend spacings (IWS), which is associated with the time convention of $\tilde{P}_i$, regular weekends have $n_i = 1$.

Table 4.5: Summary of frequencies of occurence of time spacings between sucessively available observations encountered in each price series.

Spacings (n_i)	1	2	3	4	5	6	7
AXJO (CS)	2009	12	485	26	17	0	1
AXJO (IWS)	2491	38	20	0	1	0	0
MXX (CS)	1945	50	483	35	10	0	0
MXX (IWS)	2426	85	12	0	0	0	0
N225 (CS)	1921	37	449	61	7	7	3
N225 (IWS)	2370	97	7	8	3	0	0

Utilising the SMC methodology presented in this chapter, the underlying GARCH parameters implied by each of these aggregated return series are presented in Table 4.6. Comparing Table 4.6 to Table 3.6 we see that the point estimates from each table are within the corresponding 95% confidence band indicated in the other table, with the exception of the parameter μ for the MXX series with calendar time spacings. For the MXX returns series with calendar time spacings, the "missing" observation treatment yields an estimate of μ, the mean daily log return, of 0.1514 (0.0910, 0.2119), whereas the "aggregated" observation treatment here estimates a much lower mean daily log return of 0.0783 (0.0386, 0.1180).

Acting on a calender time scale, the "missing" observation treatment in Section 3.3.1 ignores price changes between the close of trade on the last trading day of a week and the close of trade on the first trading day of the following week. For instance, assuming a regular weekend, the change in price from the close of Friday to the close of the following Monday is not accounted for, as the three one-day log returns (Friday-Saturday, Saturday-Sunday, Sunday-Monday) are considered missing. Whereas the "aggregated" observation treatment (with calendar time spacings), incorporates said price change as an observation of the aggregated value of three successive one-day log returns.

The difference in estimated mean daily log return for the MXX price series (with calendar time spacings) between the "missing" and "aggregated" observation treatments may suggest that the MXX index has a tendency to decline over weekends.

Now, comparing CS and IWS treatments in Table 4.6, we see in all three indices, that estimates of α and β are not statistically different between treatments. However, in terms of ω, consistent across the three indices, we see that the ω estimate using IWS lie above the upper 95% level of the ω estimate using CS. In terms of μ and the long-run average variance $\frac{\omega}{1-\alpha-\beta}$, there does not seem to be any statistical difference in these quantities between treatments. Although, looking at the point estimates of μ and $\frac{\omega}{1-\alpha-\beta}$, these appear, for all three indices, in line with the conjecture of Section 3.3.1, that one would expect these quantities to rise as physical time is compressed.

Table 4.6: Estimated GARCH parameters with 95% confidence intervals for the aggregated returns data.

	ω	α	β	μ	$\omega/(1-\alpha-\beta)$
AXJO (CS)	0.019 (0.009, 0.028)	0.106 (0.075, 0.136)	0.862 (0.819, 0.904)	0.025 (0.005, 0.046)	0.570 (0.447, 0.693)
AXJO (IWS)	0.028 (0.013, 0.043)	0.096 (0.069, 0.124)	0.862 (0.818, 0.905)	0.035 (0.008, 0.063)	0.668 (0.534, 0.802)
MXX (CS)	0.039 (0.019, 0.058)	0.124 (0.092, 0.156)	0.866 (0.831, 0.900)	0.078 (0.039, 0.118)	3.660 (1.332, 5.987)
MXX (IWS)	0.061 (0.027, 0.095)	0.128 (0.091, 0.164)	0.858 (0.818, 0.899)	0.116 (0.064, 0.169)	4.349 (1.605, 7.093)
N225 (CS)	0.024 (0.011, 0.036)	0.074 (0.053, 0.095)	0.913 (0.889, 0.937)	0.001 (-0.035, 0.038)	1.836 (1.249, 2.422)
N225 (IWS)	0.046 (0.023, 0.068)	0.082 (0.059, 0.104)	0.897 (0.871, 0.924)	0.004 (-0.044, 0.052)	2.181 (1.558, 2.803)

4.2 SMC for partially observed AR(1)-GARCH(1,1)

Rather than model the phenomena as a differenced series, alternative first order difference relations can be utilised. One such is the Autoregressive (AR) process. The AR(1) process $(P_i)_{i=1,\dots,N}$ is defined as

$$P_i = c_0 + c_1 P_{i-1} + \epsilon_i \tag{4.21}$$

where $(\epsilon_i)_{i=1,\dots,N}$ is an i.i.d noise series.

Replacing $(\epsilon_i)_{i=1,\dots,N}$ with a GARCH(1,1) sequence, we arrive at the AR(1)-GARCH(1,1) model

$$P_i = c_0 + c_1 P_{i-1} + z_i \sigma_i \tag{4.22}$$

$$\sigma_i^2 = \omega + \alpha(P_{i-1} - c_0 - c_1 P_{i-2})^2 + \beta \sigma_{i-1}^2. \tag{4.23}$$

One may wish to model a series directly as an AR(1)-GARCH(1,1) process. Consider the case when P_i and P_{i-2} are observed, however P_{i-1} is not. In this instance we have

$$P_i = c_0 + c_1(c_0 + c_1 P_{i-2} + z_{i-1}\sigma_{i-1}) + z_i \sigma_i \tag{4.24}$$

$$= c_0(1 + c_1) + c_1^2 P_{i-2} + c_1 z_{i-1}\sigma_{i-1} + z_i \sigma_i. \tag{4.25}$$

In general, for an integer $r \geq 1$,

$$P_i - c_1^r P_{i-r} = \sum_{j=0}^{r-1} (z_{i-j}\sigma_{i-j} + c_0)c_1^j. \tag{4.26}$$

Thus, it can be seen that the treatment for "aggregated" observation is the special case of the AR(1)-GARCH(1,1) when $c_0 = 0$ and $c_1 = 1$.

Assume one partially observes $(P_j)_{j=1,\ldots,N}$ at time points $1 = t_1 < t_2 < \ldots < t_n = N$, where $t_i \in \{2, \ldots, N-1\}$ for $i = 2, \ldots, n-1$. To put this problem in the form of (1.8) – (1.9) we take

$$Y_i = P_{t_i} - c_1^{n_i} P_{t_{i-1}} \tag{4.27}$$

$$Z_i = z_{t_i} \tag{4.28}$$

$$X_i = \left(\sigma^2_{t_{i-1}+1}, \mathbb{I}(n_i > 1) \sum_{k=1}^{n_i-1} (z_{t_{i-1}+k}\sigma_{t_{i-1}+k} + c_0)c_1^{n_i-k}, \sigma^2_{t_i}\right), \tag{4.29}$$

where again $n_i = t_i - t_{i-1}$ and W_i the same as (4.4), with

$$Y_i = f(X_i, Z_i) = c_0 + X_{i,2} + \sqrt{X_{i,3}}Z_i \tag{4.30}$$
$$X_i = g(X_{i-1}, Y_{i-1}, W_i) = \Big(g_1(X_{i-1}, Y_{i-1}), g_2(X_{i-1}, Y_{i-1}, W_i), g_3(X_{i-1}, Y_{i-1}, W_i)\Big)$$

whereby,

$$g_1(X_{i-1}, Y_{i-1}) = \omega + \alpha(Y_{i-1} - X_{i-1,2} - c_0)^2 + \beta X_{i-1,3} \tag{4.31}$$

$$\begin{aligned} g_2(X_{i-1}, Y_{i-1}, W_i) = {} & \mathbb{I}(n_i > 1)(z_{t_{i-1}+1}\sqrt{g_1(X_{i-1}, Y_{i-1})} + c_0)c_1^{n_i-1} \\ & + \mathbb{I}(n_i > 2) \sum_{k=2}^{n_i-1} (z_{t_{i-1}+k}\sqrt{\tilde{g}(X_{i-1}, Y_{i-1}, W_i, k)} + c_0)c_1^{n_i-k} \end{aligned} \tag{4.32}$$

with $g_3(X_{i-1}, Y_{i-1}, W_i)$ and $\tilde{g}(X_{i-1}, Y_{i-1}, W_i, k)$ defined respectively as in (4.8) and (4.9) using the updated definitions (4.31)-(4.32). Then one can proceed as in the "aggregated" observation case to perform likelihood inference.

4.2.1 Simulation study

We repeat the simulation study described in Section 3.2.1, this time using an AR(1)-GARCH(1,1), standard normal innovations, true parameters $\omega = 0.1, \alpha = 0.08, \beta = 0.9, c_0 = 10$ and $c_1 = 0.65$. Using an integrated C++ and R implementation, the GARCH parameters were optimised over the set defined in (3.4) without any constraints on the AR parameters.

Table 4.7 displays the results from utilising 250 and 1000 particles for 1000 simulated AR-GARCH data sets at missing percentages of 5%, 10%, 20%, 35% and 50%. Regarding the GARCH parameters, the comments for Table 4.7 are identical to the comments made for Table 3.1. Thus we just summarise the following findings regarding the AR parameters:

- Like the GARCH parameters, the AR parameters are scarcely impacted by a four fold increase in particles.
- While the average bias in estimates of c_0 and c_1 as missing percentage increases (while keeping n fixed) is seen to improve, standard deviations remain roughly constant. RMSE, having variance as the dominating component, also exhibits the same behaviour, stable with increasing proportions of missingness. The RMSE at the 50% missingness scenario is on par with that of the no missingness scenario.

A summary, across the 1000 simulations for each level of missingness, of computational times using an Intel Xeon 3.06 GHz processor is provided in Table 4.8. The time to complete a SMC likelihood evaluation can be seen to be linear in N as well as the number of particles K.

Table 4.7: SMC estimation of the AR(1)-GARCH(1,1) model with missing data: Summary statistics across 1000 replications for each level of missingness.

		250 particles					1000 particles				
% Missing	0%	5%	10%	20%	35%	50%	5%	10%	20%	35%	50%
$\omega = 0.1$											
Mean	0.1179	0.1176	0.1171	0.1163	0.1147	0.1133	0.1168	0.1164	0.1152	0.1134	0.1107
Bias	0.0179	0.0176	0.0171	0.0163	0.0147	0.0133	0.0168	0.0164	0.0152	0.0134	0.0107
SD	0.0418	0.0427	0.0439	0.0411	0.0397	0.0364	0.0426	0.0435	0.0407	0.0391	0.0351
Mean SE	0.0384	0.0382	0.0375	0.0360	0.0338	0.0316	0.0382	0.0377	0.0364	0.0342	0.0323
RMSE	0.0455	0.0462	0.0471	0.0442	0.0423	0.0388	0.0458	0.0465	0.0434	0.0414	0.0367
$\alpha = 0.08$											
Mean	0.0802	0.0814	0.0813	0.0812	0.0818	0.0814	0.0814	0.0813	0.0811	0.0814	0.0808
Bias	0.0002	0.0014	0.0013	0.0012	0.0018	0.0014	0.0014	0.0013	0.0011	0.0014	0.0008
SD	0.0131	0.0135	0.0137	0.0136	0.0140	0.0133	0.0135	0.0136	0.0136	0.0140	0.0131
Mean SE	0.0135	0.0137	0.0136	0.0136	0.0135	0.0134	0.0137	0.0137	0.0136	0.0136	0.0135
RMSE	0.0131	0.0135	0.0137	0.0137	0.0141	0.0133	0.0136	0.0137	0.0137	0.0141	0.0132
$\beta = 0.9$											
Mean	0.8954	0.8948	0.8950	0.8952	0.8951	0.8958	0.8950	0.8952	0.8956	0.8957	0.8969
Bias	-0.0046	-0.0052	-0.0050	-0.0048	-0.0049	-0.0042	-0.0050	-0.0048	-0.0044	-0.0043	-0.0031
SD	0.0176	0.0182	0.0186	0.0185	0.0191	0.0178	0.0182	0.0185	0.0184	0.0190	0.0175
Mean SE	0.0179	0.0180	0.0179	0.0177	0.0175	0.0173	0.0180	0.0180	0.0179	0.0177	0.0176
RMSE	0.0182	0.0189	0.0193	0.0191	0.0197	0.0183	0.0188	0.0191	0.0189	0.0195	0.0177
$c_0 = 10$											
Mean	10.0624	9.8513	9.8649	9.8881	9.9002	9.9150	9.8423	9.8526	9.8774	9.8874	9.8983
Bias	0.0624	-0.1487	-0.1351	-0.1119	-0.0998	-0.0850	-0.1577	-0.1474	-0.1226	-0.1126	-0.1017
SD	0.4961	0.4870	0.4883	0.4747	0.4745	0.4958	0.4860	0.4884	0.4697	0.4710	0.4876
Mean SE	0.4560	0.4536	0.4465	0.4327	0.4218	0.4039	0.4535	0.4481	0.4344	0.4145	0.3946
RMSE	0.5000	0.5092	0.5066	0.4877	0.4849	0.5031	0.5110	0.5101	0.4855	0.4843	0.4981
$c_1 = 0.65$											
Mean	0.6479	0.6552	0.6547	0.6539	0.6535	0.6529	0.6555	0.6551	0.6543	0.6539	0.6535
Bias	-0.0021	0.0052	0.0047	0.0039	0.0035	0.0029	0.0055	0.0051	0.0043	0.0039	0.0035
SD	0.0173	0.0170	0.0171	0.0166	0.0166	0.0173	0.0170	0.0171	0.0164	0.0165	0.0170
Mean SE	0.0162	0.0162	0.0159	0.0154	0.0151	0.0145	0.0162	0.0160	0.0155	0.0148	0.0142
RMSE	0.0175	0.0178	0.0177	0.0170	0.0169	0.0175	0.0178	0.0178	0.0170	0.0169	0.0174

Table 4.8: Summary, across the 1000 simulations for each level of missingness, of computational times for the SMC optimisation using an Intel Xeon 3.06 GHz.

		250 particles			1000 particles		
% Missing	N	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)
5%	2105	83.56	912.80	0.092	347.47	912.33	0.381
10%	2222	90.34	919.99	0.098	371.88	911.29	0.408
20%	2500	101.53	896.98	0.113	414.32	882.47	0.470
35%	3076	129.71	924.42	0.140	531.13	917.80	0.579
50%	4000	184.30	1024.96	0.180	708.68	945.73	0.749

4.2.2 Real data analysis

We reanalyse the three wind time series considered in Section 3.3.2 estimating from them AR(1)-GARCH(1,1) models obtained in two ways. The first is the two-step estimation procedure employed in Section 3.3.2, using the `arima` command in R to obtain first the AR(1) parameters ignoring the GARCH effects, then fitting a GARCH(1,1) to the partially available AR(1) residuals. The second is the simultaneous estimation of the AR(1) and GARCH(1,1) parameters through the SMC configuration presented in Section 4.2.

Table 4.9 displays the fits obtained using both methods for each of the three wind time series. Comparing the two methods we see that the GARCH parameters are within each others respective confidence intervals and, for all but c_0 for Penrith, the AR parameters are outside each others confidence intervals.

In the case of a full observation series, it is known (cf. Gourieroux [1997]) that two-step estimation result in consistent although inefficient estimates. In the case of missing data it is not clear if these results carry over but it is likely that the two-step procedure continues to be inefficient.

Table 4.9: Fitted AR(1)-GARCH(1,1) models for the wind data. Top: Using a two-step estimation method - first fitting the AR(1) ignoring the GARCH effects then fitting a GARCH(1,1) to the partially available AR(1) residuals. Bottom: Simultaneous estimation of the AR(1) and GARCH(1,1) parameters.

Two-Stage Estimation			
	ω	α	β
North Head	0.8842 (0.7194, 1.0490)	0.2459 (0.2047, 0.2871)	0.5462 (0.4814, 0.6111)
Penrith	1.0619 (0.7792, 1.3446)	0.1630 (0.1336, 0.1923)	0.6551 (0.5867, 0.7235)
Fort Denison	0.2960 (0.2207, 0.3712)	0.1814 (0.1472, 0.2156)	0.7805 (0.7432, 0.8179)
	c_0	c_1	$\omega/(1-\alpha-\beta)$
North Head	-0.0085 (-0.0540, 0.0369)	0.9747 (0.9694, 0.9800)	4.2547 (3.8177, 4.6916)
Penrith	-0.1646 (-0.2226, -0.1067)	0.8836 (0.8722, 0.8950)	5.8372 (5.4074, 6.2670)
Fort Denison	-0.0383 (-0.0906, 0.0139)	0.9495 (0.9417, 0.9573)	7.7736 (5.1940, 10.353)

Simultaneous Estimation			
	ω	α	β
North Head	0.9259 (0.7626, 1.0893)	0.2770 (0.2324, 0.3216)	0.5146 (0.4505, 0.5786)
Penrith	1.1217 (0.8169, 1.4266)	0.1732 (0.1413, 0.2052)	0.6356 (0.5612, 0.7100)
Fort Denison	0.3291 (0.2368, 0.4214)	0.2054 (0.1609, 0.2500)	0.7559 (0.7076, 0.8042)
	c_0	c_1	$\omega/(1-\alpha-\beta)$
North Head	0.0733 (0.0344, 0.1122)	0.9831 (0.9783, 0.9879)	4.4416 (3.9218, 4.9615)
Penrith	-0.1690 (-0.2229, -0.1150)	0.9034 (0.8923, 0.9146)	5.8680 (5.4292, 6.3068)
Fort Denison	0.0375 (-0.0052, 0.0802)	0.9610 (0.9521, 0.9699)	8.5157 (5.3442, 11.687)

4.3 Concluding remarks

In summary, this section has illustrated use of the SMC methodology for providing computationally feasible and reliable estimators (along with standard errors) to tackle the problems of parameter estimation for temporally aggregated GARCH(1,1) series as well as partially observed AR(1)-GARCH(1,1) series.

For the restricted case when the data is consecutively aggregated in non-overlapping blocks of equal size - a condition required for application of the "Weak" GARCH approximation of Drost and Nijman [1993], the SMC method has been shown to provide better estimates, in particular for the persistence parameters α and β, for which the Drost approximation displays large biases.

Regarding partially observed AR(p)-GARCH(1,1), $p > 1$ series and temporally aggregated AR(1)-GARCH(1,1) series, the inability to propagate such series through a single variable at each time step means these problems cannot be solved using the technique utilised in this chapter.

Chapter 5

Application: Likelihood inference from discrete observation of a COGARCH process

5.1 Introduction

Seeking a continuous time counterpart to the discrete time GARCH(1,1) model, to be used in modelling increasingly available high frequency financial data sets as well as utilised in the well established continuous time option pricing framework (see for instance Harrison and Pliska [1991]), Kluppelberg et al. [2004] developed the COGARCH(1,1) process.

To arrive at the COGARCH model, the univariate noise series of a discrete time GARCH model is replaced by the increments of a Lévy process. The resulting COGARCH model is then a continuous time analogue of the discrete time GARCH model, both being systems driven by only a single source of noise and both incorporating a feedback mechanism whereby a perturbation in the current period continues to affect the variance of future perturbations.

Calibration of the COGARCH model to data is still an open topic. The cur-

rently available methods can be grouped into three categories. The first group employs the use of theoretical moments. Haug et al. [2007] equates theoretical moments with sample based estimates, while Bibbona and Negri [2015] use theoretical moments to construct prediction based estimating functions. Both these procedures however, have only been demonstrated under the assumption of equally spaced observation. To extend to irregularly spaced data requires overcoming the problem that the moments needed for both these methods depend on the time spacings between observations.

The second is the quasi maximum likelihood (QML) method, which maximises a quasi likelihood in place of the true likelihood. First suggested for COGARCH models in Maller et al. [2008], this method is based on a first jump approximation to the driving Lévy process and an assumption that the observed increments are Gaussian and conditionally independent. These simplifying assumptions lead to construction of a tractable likelihood very similar to that used for fitting GARCH models to discrete time data. The QML method can be applied to regularly and irregularly spaced observations. However, as pointed out by Muller [2010], the QML estimator is not consistent in general. Although Kim and Lee [2013] established the consistency and asymptotic normality of a QML estimator based on a slightly different form of quasi likelihood, they assumed an asymptotic scenario where the observation times grow infinitely dense, which is rather unrealistic from a practical point of view. Marin et al. [2015] uses a data cloning technique to explore the quasi likelihood surface by Markov chain Monte Carlo (MCMC) methods.

The third is the MCMC Bayesian procedure developed by Muller [2010], for the specific case of a compound Poisson driving Lévy process. It was reported that the posterior mode estimator outperformed the QML estimator in simulation studies. However, this MCMC procedure is computationally very expensive as the procedure attempts to uncover the complete sample path of the underlying compound Poisson process. Moreover, the choices of the prior distributions for

the parameters and hyper-parameters influences the performance of the estimators and there does not seem to be a principled approach to making these choices.

Inferences for the COGARCH model based on the genuine likelihood, however, have not been addressed, although as a general inference methodology, the maximum likelihood method is expected to have various advantages such as being applicable both to regularly spaced and to irregularly spaced data, ease of variance estimation, and asymptotic efficiency. The challenge in making inferences based on the genuine likelihood in COGARCH models is the computation of the likelihood function, as the likelihood lacks a tractable form and its computation entails the evaluation of infinite-dimensional integrals.

Development of a computationally feasible method for maximum likelihood estimation in the COGARCH model is the concern of this chapter. Since in many high frequency financial data sets, especially intraday data, time intervals without price changes are often observed, the model based on finite activity pure jump processes, or the compound Poisson process, seems more fitting. Therefore, as in Muller [2010], in this work we consider only the COGARCH model driven by a compound Poisson process. We will demonstrate that the sequential Monte Carlo (SMC) technique can be applied with the bootstrap particle filter [Gordon et al., 1993] to obtain an unbiased estimate of the likelihood of the compound Poisson COGARCH process based on either regularly or irregularly spaced discrete observations. The approximated likelihood function by the bootstrap particle filter is not continuous in the model parameters in general, even if the randomness in the Monte Carlo simulation is controlled. This creates difficulty in numerically optimising the likelihood function to obtain the maximum likelihood estimator (MLE). To overcome this issue, we use the continuous resampling procedure developed by Pitt and Malik [2011] to produce a smooth likelihood surface which is amenable to numerical optimisation. Unlike the MCMC procedure of Muller [2010] which essentially attempts to fill in the missing information, the SMC procedure

effectively integrates out the missing information. We demonstrate superior finite sample performance in comparison with the QML method through simulations. We also present numeric evidence that the Hessian matrix of the negative logarithm of the approximate likelihood can be inverted to estimate the variance of the MLE.

This chapter is organised as follows. A brief description of the COGARCH process is provided in Section 5.2. The intractable likelihood of a COGARCH process based on discrete observations is detailed in Section 5.3. Parameter estimation through simulated maximum likelihood using SMC is addressed in Section 5.4. A real world application of the procedure is illustrated on high frequency intraday financial data in Section 5.5. Simulation studies are conducted in Section 5.6. The chapter concludes in Section 5.7 with a discussion.

A paper based on the contents of this chapter is currently under revision with the *Journal of Financial Econometrics*.

5.2 The COGARCH(1,1) model

The COGARCH(1,1) model introduced in Kluppelberg et al. [2004] is the following system of stochastic differential equations

$$\begin{aligned} \mathrm{d}G_t &= \sigma_t \, \mathrm{d}L_t, \\ \mathrm{d}\sigma_{t+}^2 &= (\beta - \eta\sigma_t^2) \, \mathrm{d}t + \varphi\sigma_t^2 \, \mathrm{d}[L, L]_t^{(d)}, \end{aligned} \tag{5.1}$$

where L_t is a Lévy process, $\beta > 0$, $\eta > 0$ and $\varphi > 0$ are model parameters, and $[L, L]^{(d)}$ denotes the discrete quadratic variation of the Lévy process (cf. Protter [2005], p.66). Here and hereafter, $t\pm$ indicates the right/left-hand limit at t.

For a finite activity pure jump Lévy process, if we denote N_t as the number of jumps on the interval $[0, t]$, $\tau_1, \ldots, \tau_{N_t}$ as the ordered jump times and $z_1, \ldots, z_{N_t}$ as the corresponding i.i.d. (independent and identically distributed) jump sizes, then

the COGARCH(1,1) process, with an initial volatility value σ_0, can be represented as

$$G_t = G_0 + \int_0^t \sigma_s \, \mathrm{d}L_s = G_0 + \sum_{i=1}^{N_t} \sigma_{\tau_i} z_i,$$

where the skeleton volatility process $(\sigma_{\tau_i})_{i=1,\ldots,N_t}$ is given recursively by

$$\sigma^2_{\tau_i} = \frac{\beta}{\eta}\left(1 - e^{-\eta(\tau_i - \tau_{i-1})}\right) + e^{-\eta(\tau_i - \tau_{i-1})}\sigma^2_{\tau_{i-1}+}, \tag{5.2}$$

$$\sigma^2_{\tau_i+} = \sigma^2_{\tau_i}(1 + \varphi z_i^2), \; i = 1, \ldots, N_t, \text{ and } \sigma^2_{\tau_0+} := \sigma^2_0. \tag{5.3}$$

We see from the representation above that β/η bounds the squared volatility process from below, and can be interpreted as a baseline value that the squared volatility decays to in the absence of jumps. The parameter η determines the speed of this decay, with φ controlling the degree of impact new jumps have on the volatility. For parameter identifiability, it is common, such as was done in Haug et al. [2007] and Maller et al. [2008], to standardise the mean and variance per unit time of the driving Lévy process to be 0 and 1 respectively (that is $\mathbb{E}(L_1) = 0$ and $\mathbb{E}(L_1^2) = 1$), which we assume hereafter. The initial value of the squared volatility is often set at its long term mean $\sigma_0^2 = \beta/(\eta - \varphi)$ (see Proposition 4.2 of Kluppelberg et al. [2004]). Existence of a stationary distribution for the volatility process $(\sigma_t)_{t\geq 0}$ is guaranteed if $\eta - \varphi > 0$ (see Theorem 3.2 of Kluppelberg et al. [2004]).

5.3 Likelihood based on discrete observations of the COGARCH process

Assume the driving Lévy process is compound Poisson with arrival rate λ and a jump size distribution with density $q(\cdot)$ relative to Lebesgue measure μ. Note that this implies the jump size distribution has no point mass at 0. If one observes the complete sample path of the process G_t on the observation interval $[0, T]$, or equivalently the number N_T of jumps of the process in the interval together with the jump times $\tau_1, \ldots, \tau_{N_T}$ and the respective jump sizes $g_i = G_{\tau_i} - G_{\tau_i -}$, then the likelihood of the model parameters $\theta = (\beta, \eta, \varphi, \lambda)$, is the density function of the distribution of the compound Poisson COGARCH process, interpreted as a function of the parameter vector θ, given by

$$\text{lik}_c(\theta) = \lambda^{N_T} e^{-\lambda T} \prod_{i=1}^{N_T} \frac{1}{\sigma_{\tau_i}} q\left(\frac{g_i}{\sigma_{\tau_i}}\right),$$

where the σ_{τ_i}'s depend on the parameters θ and the initial volatility $\sigma_0^2 = \beta/(\eta - \varphi)$ through the recursions (5.2)-(5.3), with the z_i in (5.3) replaced by g_i/σ_{τ_i}. The parameter space is $\Theta = \{(\beta, \eta, \varphi, \lambda) : \beta > 0, \varphi > 0, \eta > \varphi, \lambda > 0\}$.

While the complete data likelihood is easy to compute and is a smooth function of the parameter vector, in practice we do not normally have continuous observations of the COGARCH process. The typical observations we have are the values of the process at a set of discrete time points $0 = t_0 < t_1 < \ldots < t_n = T$. It is assumed that the observation times $t_i, i = 1, \ldots, n$ are independent of the COGARCH process and their choice does not depend on the parameters of the COGARCH model. Assume the COGARCH process has a known initial value G_0, then the observations are equivalent to the observed increments of G over successive observation times, $\bar{G}_i := G_{t_i} - G_{t_{i-1}}$, $i = 1, \ldots, n$. Since each $\bar{G}_i$ has a positive probability to be 0 (due to the compound Poisson Lévy process), we define the likelihood to be the density $p_\theta(\bar{G}_{1:n})$ of the observations $(\bar{G}_1, \ldots, \bar{G}_n) =: \bar{G}_{1:n}$

with respect to the product measure $(\delta_0 + \mu)^{\otimes n}$, interpreted as a function of the parameter vector θ. Here, and hereafter, we use δ_x to denote the Dirac measure at x. Formally, the likelihood can be written in the following form,

$$\begin{aligned}\text{lik}(\theta) = p_\theta(\bar{G}_{1:n}) &= \prod_{i=1}^{n} p_\theta(\bar{G}_i|\bar{G}_{1:i-1}) \\ &= \prod_{i=1}^{n} e^{-\lambda\Delta t_i \mathbb{I}(\bar{G}_i=0)} \Big\{ (1-e^{-\lambda\Delta t_i}) \int \frac{1}{\sigma_{\tau_{N_{t_i}}}} q\left(\frac{G_{t_i} - G_{\tau_{N_{t_i}}-}}{\sigma_{\tau_{N_{t_i}}}}\right) \\ &\quad \times \mathbb{P}(\,\mathrm{d}G_{\tau_{N_{t_i}}-},\, \mathrm{d}\sigma_{\tau_{N_{t_i}}}|\bar{G}_{1:i-1}) \Big\}^{\mathbb{I}(\bar{G}_i\neq 0)},\end{aligned} \tag{5.4}$$

where $\Delta t_i = t_i - t_{i-1}$, $\mathbb{P}(\,\mathrm{d}G_{\tau_{N_{t_i}}-},\, \mathrm{d}\sigma_{\tau_{N_{t_i}}}|\bar{G}_{1:i-1})$ denotes the conditional joint distribution of $G_{\tau_{N_{t_i}}-}$ and $\sigma_{\tau_{N_{t_i}}}$ given the observations $\bar{G}_{1:i-1}$. Recall that, when $\bar{G}_i \neq 0$, there is at least one jump in the interval $(t_{i-1}, t_i]$ and $G_{\tau_{N_{t_i}}-}$ represents the value of the process G just before the last jump time in the interval, $G_{t_i} - G_{\tau_{N_{t_i}}-}$ the size of the last jump of G in the interval, and $\sigma_{\tau_{N_{t_i}}}$ the volatility just before the last jump in the interval.

The conditional distributions $\mathbb{P}(\,\mathrm{d}G_{\tau_{N_{t_i}}-},\, \mathrm{d}\sigma_{\tau_{N_{t_i}}}|\bar{G}_{1:i-1})$ in (5.4) are intractable in general, and therefore the likelihood can't be directly computed. We will show in the next section how this intractable likelihood can be approximated using SMC.

5.4 SMC for discretely observed COGARCH(1,1)

In this section we outline the procedure for parameter estimation through simulated maximum likelihood using SMC. To begin, we first formulate the observed non-zero increments of the COGARCH process as the observations of a hidden Markov process model (of the form (1.8)-(1.9)). While there are various ways to do this, we do it in such a way that the resulting model is suitable for SMC implementation.

5.4.1 State space representation

For notational convenience, let n_i be the number of jumps of the COGARCH process that occur in the interval $(t_{i-1}, t_i]$, $t_{i-1} < \tau_{i,1} < \ldots < \tau_{i,n_i} < t_i$ be the n_i jump times, and $z_{i,1}, \ldots, z_{i,n_i}$ the corresponding n_i jump sizes of the underlying compound Poisson process. That is, $n_i = N_{t_i} - N_{t_{i-1}}$, and $\tau_{i,j} = \tau_{N_{t_{i-1}}+j}$, $z_{i,j} = z_{N_{t_{i-1}}+j}$, $j = 1, \ldots, n_i$. When $n_i = 0$ we define $\tau_{i,0} := t_{i-1}$. Recall that

$$\bar{G}_i = G_{t_i} - G_{t_{i-1}} = \sum_{j=1}^{n_i} \sigma_{\tau_{i,j}} z_{i,j}, \tag{5.5}$$

where the $\sigma_{\tau_{i,j}}$, $j = 1, \ldots, n_i$ are recursively given as in (5.2) and (5.3), with initial value $\sigma^2_{\tau_{i,0}+} = \sigma^2_{t_{i-1}}$, and the sum is interpreted as 0 when $n_i < 1$. Let

$$x_i = \left(\sigma^2_{t_{i-1}}, \sum_{j=1}^{n_i-1} \sigma_{\tau_{i,j}} z_{i,j}, \sigma^2_{\tau_{i,n_i}}, \tau_{i,n_i} \right), \tag{5.6}$$

and denote $x_{i,j}$ the j-th element of x_i. Then $x_{i,1}$ is the volatility at the beginning of the i-th observation interval, $x_{i,2}$ is the increment of the COGARCH process in the interval before the last jump time in the interval, $x_{i,3}$ is the volatility at the last jump time, and $x_{i,4}$ is either the last jump time or the beginning of the observation interval, depending on whether $n_i > 0$ or $n_i = 0$. Note that the distribution of $x_{i,2:4}$ is fully determined once $x_{i,1}$ is given. Note also from (5.5) that the distribution of $\bar{G}_i$ is fully determined by $x_{i,2:4}$ via

$$\bar{G}_i = \begin{cases} 0, & x_{i,4} = t_{i-1} \Leftrightarrow n_i = 0, \\ x_{i,2} + \sqrt{x_{i,3}} z_{i,n_i}, & x_{i,4} > t_{i-1} \Leftrightarrow n_i > 0. \end{cases} \tag{5.7}$$

The conditional density of $\bar{G}_i$ given x_i, relative to the measure $\delta_0 + \mu$, has this explicit form,

$$p(\bar{G}_i|x_i) = \begin{cases} \mathbb{I}(\bar{G}_i = 0), & x_{i,4} = t_{i-1}, \\ \frac{1}{\sqrt{x_{i,3}}} q\left(\frac{\bar{G}_i - x_{i,2}}{\sqrt{x_{i,3}}}\right) \mathbb{I}(\bar{G}_i \neq 0), & x_{i,4} > t_{i-1}, \end{cases} \tag{5.8}$$

where we recall that q is the μ-density of the jump sizes of the underlying compound Poisson process. Now, observe that

$$\begin{aligned} p(\bar{G}_i|\bar{G}_{1:i-1}) &= p(\bar{G}_i|\bar{G}_{1:i-1}, x_{i,4} = t_{i-1})\mathbb{P}(x_{i,4} = t_{i-1}|\bar{G}_{1:i-1}) \\ &\quad + p(\bar{G}_i|\bar{G}_{1:i-1}, x_{i,4} > t_{i-1})\mathbb{P}(x_{i,4} > t_{i-1}|\bar{G}_{1:i-1}) \end{aligned} \tag{5.9}$$

$$= \begin{cases} e^{-\lambda \Delta t_i}, & \bar{G}_i = 0, \\ (1 - e^{-\lambda \Delta t_i}) \int p(\bar{G}_i|x_i)\mathbb{P}(\,\mathrm{d}x_i|\bar{G}_{1:i-1}, x_{i,4} > t_{i-1}), & \bar{G}_i \neq 0. \end{cases} \tag{5.10}$$

where we have used (5.8) and the facts $\mathbb{P}(x_{i,4} = t_{i-1}|\bar{G}_{1:i-1}) = \mathbb{P}(n_i = 0) = e^{-\lambda \Delta t_i}$ and $\mathbb{P}(x_{i,4} > t_{i-1}|\bar{G}_{1:i-1}) = \mathbb{P}(n_i > 0) = 1 - e^{-\lambda \Delta t_i}$. The likelihood of $\bar{G}_1, \ldots, \bar{G}_n$ is given by

$$\mathrm{lik}(\theta) = \prod_{i=1}^{n} p(\bar{G}_i|\bar{G}_{1:i-1}) \tag{5.11}$$

$$\begin{aligned} &= \prod_{i=1}^{n} \Big(\mathbb{I}(\bar{G}_i = 0) e^{-\lambda \Delta t_i} \\ &\quad + \mathbb{I}(\bar{G}_i \neq 0)(1 - e^{-\lambda \Delta t_i}) \int p(\bar{G}_i|x_i)\mathbb{P}(\,\mathrm{d}x_i|\bar{G}_{1:i-1}, x_{i,4} > t_{i-1}) \Big). \end{aligned} \tag{5.12}$$

In general the conditional distributions $\mathbb{P}(\,\mathrm{d}x_i|\bar{G}_{1:i-1}, x_{i,4} > t_{i-1})$ are intractable, which hinders direct evaluation of (5.12). The likelihood contributions for observations that are zero can be evaluated exactly, however we resort to a SMC approach to approximate the likelihood contributions of the non-zero observations. Let $\kappa_m := \min\{k : \sum_{j=1}^{k} \mathbb{I}(\bar{G}_j \neq 0) = m\}$ denote the index of the m-th non-zero

observation, for $m = 1, 2, \ldots, \tilde{n} := \sum_{j=1}^{n} \mathbb{I}(\bar{G}_j \neq 0)$. Then we can approximate the likelihood contribution for the non-zero increments by framing the the evolution of the conditionally non-zero increments in the form of (1.8)-(1.9) by setting for $i = 1, \ldots, \tilde{n}$

$$Y_i = \bar{G}_{\kappa_i}, \; X_i = x_{\kappa_i}, \; Z_i = z_{\kappa_i, n_{\kappa_i}} \tag{5.13}$$

$$W_i = \{n_{\kappa_i}, \tau_{\kappa_i,1}, \ldots, \tau_{\kappa_i,n_{\kappa_i}}, \zeta(n_{\kappa_i})\} \tag{5.14}$$

where

$$\zeta(n_{\kappa_i}) = \begin{cases} \{z_{\kappa_i,1}, \ldots, z_{\kappa_i,n_{\kappa_i}-1}\} & \text{when } n_{\kappa_i} > 1, \\ \emptyset & \text{when } n_{\kappa_i} = 1, \end{cases} \tag{5.15}$$

with

$$Y_i = f(X_i, Z_i) = X_{i,2} + \sqrt{X_{i,3}} Z_i \tag{5.16}$$

$$\begin{aligned} X_{i,1} &= \beta/\eta + e^{-\eta\left(t_{\kappa_i - 1} - X_{i-1,4}\right)} \left(X_{i-1,3} + \varphi(Y_{i-1} - X_{i-1,2})^2 - \beta/\eta\right) \\ &=: g_1(X_{i-1}, Y_{i-1}, t_{\kappa_i - 1}) \end{aligned} \tag{5.17}$$

$$X_i = [g_1(X_{i-1}, Y_{i-1}, t_{\kappa_i - 1}), g_2(X_{i,1}, W_i), g_3(X_{i,1}, W_i), g_4(W_i)]^T \tag{5.18}$$

for some functions $g_2(X_{i,1}, W_i) = X_{i,2}, g_3(X_{i,1}, W_i) = X_{i,3}$ and $g_4(W_i) = X_{i,4}$. With the specifications (5.13)-(5.18), proceeding as outlined in Section 2.1, yields a consistent and unbiased, K particle, SMC approximation of the likelihood, given by

$$\begin{aligned} \widehat{\text{lik}}_K(\theta) = &\prod_{i=1}^{n} \left(\mathbb{I}(\bar{G}_i = 0) e^{-\lambda \Delta t_i} + \mathbb{I}(\bar{G}_i \neq 0)\right) \\ &\times \prod_{i=1}^{\tilde{n}} \left((1 - e^{-\lambda \Delta t_{\kappa_i}}) \frac{1}{K} \sum_{k=1}^{K} p(Y_i | X_i^{(k)})\right), \end{aligned} \tag{5.19}$$

whereby $X_i^{(k)}$ $i = 1, \ldots, \tilde{n}$, $k = 1, \ldots, K$ are the SMC obtained particles.

The ease of evaluating (5.8) is a big advantage of our choice of the state variable X_i over other possible choices. For instance, if the state variable is chosen as $\sigma^2_{t_{\kappa_i-1}}$, then the resulting $p(\bar{G}_{\kappa_i}|X_i)$ does not admit an analytical form and is very difficult to evaluate, although we still have a hidden Markov model.

The noise W_i encompasses the following components of the compound Poisson process in the interval $[t_{\kappa_i-1}, t_{\kappa_i}]$: the number of jumps in the interval n_{κ_i} (conditional that $n_{\kappa_i} \geq 1$), the times at which these jumps occurred $t_{\kappa_i-1} < \tau_{\kappa_i,1} < \ldots < \tau_{\kappa_i,n_{\kappa_i}} < t_{\kappa_i}$ and the size of these jumps (except for the last jump) $z_{\kappa_i,1}, \ldots, z_{\kappa_i,n_{\kappa_i}-1}$.

A procedure to simulate W_i is as follows. Define $u_i = \tau_{\kappa_i,1} - t_{\kappa_i-1}$, then u_i follows a truncated exponential distribution with rate λ whereby $\mathbb{P}(u_i < t) = \frac{1-e^{-\lambda t}}{1-e^{-\lambda \Delta t_{\kappa_i}}}$ for $t < \Delta t_{\kappa_i}$. Thus, one can simulate a Uniform(0,1) random variable U_1 to obtain

$$\tau_{\kappa_i,1} = t_{\kappa_i-1} - \frac{1}{\lambda}\log(1 - (1 - e^{-\lambda \Delta t_{\kappa_i}})U_1). \tag{5.20}$$

Then one can simulate a Poisson random variable r with mean $\lambda(t_{\kappa_i} - \tau_{\kappa_i,1})$ to obtain the number of jumps, conditional that at least one jump occurred, as $n_{\kappa_i} = 1 + r$. Proceeding, simulate r i.i.d Uniform(0,1) random variables, denote by $U_2, \ldots, U_{n_{\kappa_i}}$ the sorted uniforms, and then obtain

$$\tau_{\kappa_i,s} = \tau_{\kappa_i,1} + U_s(t_{\kappa_i} - \tau_{\kappa_i,1}), \;\; s = 2, \ldots, n_{\kappa_i}. \tag{5.21}$$

Finally, the $z_{\kappa_i,1}, \ldots, z_{\kappa_i,n_{\kappa_i}-1}$ are r i.i.d simulates from the density q.

Pseudo code summarising the above steps for obtaining $\log(\widehat{\text{lik}}_K(\theta))$ is provided in Appendix A.1 (Algorithm 1).

5.4.2 Estimating parameters

The aim is to estimate the COGARCH parameters β, η, φ and the jump intensity λ. Recall, for parameter identifiability, $\mathbb{E}(L_1) = 0$ and $\mathbb{E}(L_1^2) = 1$. In this way, the variance of the jump distribution will be set as the inverse of the jump rate λ.

5.4.2.1 Estimating λ

Now, unlike the parameters β, η and φ, it is impossible to fix randomness while varying λ, as λ directly drives the noise components W_i. To overcome this, we proceed with a two-step approach. Taking advantage of the fact that in a given interval $[t_{i-1}, t_i]$ the probability that no jumps occurs is $e^{-\lambda\Delta t_i}$ and the probability that at least one jump occurred is $1 - e^{-\lambda\Delta t_i}$ we first estimate λ, denoting this estimate $\hat{\lambda}$, by maximising the partial likelihood

$$L(\lambda) = \prod_{i=1}^{n} e^{-\lambda\Delta t_i \mathbb{I}(\bar{G}_i=0)}(1 - e^{-\lambda\Delta t_i})^{\mathbb{I}(\bar{G}_i\neq 0)}. \tag{5.22}$$

In practice, the high frequency data sets usually associated with COGARCH modelling are sufficiently large for (5.22) to yield very precise estimates of λ.

5.4.2.2 Estimating β, η and φ

Even with randomness fixed across parameter changes of β, η and φ, recall from Section 2.2, that a procedure for resampling the hidden states X_i in a continuous manner is required to obtain a SMC approximated likelihood surface that is continuous in the parameters (β, η and φ) and thereby amenable to numerical optimisation. Now, in our case, X_i is a 4-dimensional vector, so we cannot (directly) apply the 1-dimensional continuous resampling procedure from Section 2.3. However, like the aggregated GARCH(1,1) case (see Section 4.1), the model structure here can also be depicted as in Figure 4.1, thus we really only need to find a way to continuously resample the $X_{i+1,1} = \sigma^2_{t_{\kappa_{i+1}-1}}$ a 1-dimensional quantity instead. This can be achieved in fact, the same way as the aggregated GARCH(1,1) case, with pseudo code for implementing the continuous resampling procedure provided in Appendix A.1 (Algorithm 2). The log-likelihood surface, obtained using SMC with continuous resampling and noise components driven by $\hat{\lambda}$, is maximised to estimate β, η and φ. Figure 5.1 illustrates the contrast between the jagged

log-likelihood surface obtained using bootstrap resampling (Algorithm 1) and the smooth log-likelihood surface obtained using continuous resampling (Algorithm 2).

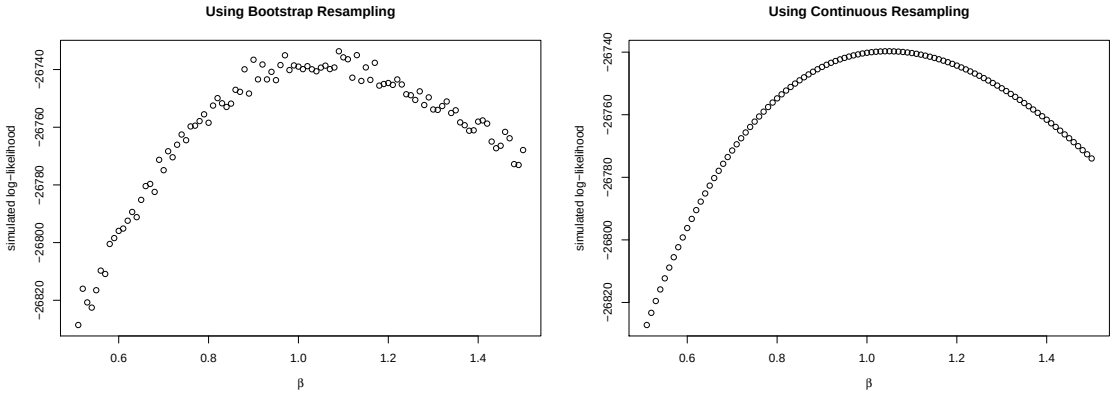


Figure 5.1: A data set with 5000 equally spaced observations of a COGARCH model with true parameters $\beta = 1$, $\eta = 0.06$ and $\varphi = 0.0425$ was generated. The above plots show the simulated log-likelihoods of this data set as β is varied (keeping η and φ fixed at their true values). LHS utilises bootstrap resampling and RHS utilises the continuous resampling procedure.

5.5 Real data analysis

In this section we use the SMC and QML procedures to model the intra-week volatility of the AUD/USD exchange rate for 50 consecutive trading weeks of 2015. The Forex market is international and a trading week commences Sunday 22:00 GMT at the weekly opening of the financial centre in Sydney, Australia and runs continuously until Friday 22:00 GMT at the weekly close of the financial centre in New York.

One-minute price data was obtained from www.histdata.com. Due to the presence of missing observations, the data is not exactly equally spaced, however a large percentage (99.6%) of observations are spaced 1 minute apart, with the remainder spaced from 2 to 41 minutes apart. For each of the 50 consecutive weeks a separate COGARCH model will be fitted. The first trading week runs from the 4th of January 2015 22:00 GMT to the 9th of January 2015 22:00 GMT and the last trading week runs from from the 13th of December 2015 22:00 GMT to the

18th of December 2015 22:00 GMT.

Two data pre-processing steps were required so as to render the data appropriate for modelling with the COGARCH model. The first step used Winsorising to exclude possible outliers, large isolated return spikes that were not followed by an elevated volatility expected under the COGARCH model. The second step adjusted for intra-week trends in hourly volatility. Previous empirical studies on foreign exchange rates have documented intra-week volatility patterns (cf. Muller et al. [1990]). The COGARCH model cannot meaningfully be applied to non-stationary data and as such; to account for intra-week hourly volatility patterns a transformation from the physical time scale to a virtual business time scale is applied leading to irregularly spaced data as trading hours with higher than normal volatility are expanded while trading hours with lower than normal volatility are compressed. This acts to balance out the volatility per unit of virtual time. Details of these data preprocessing steps are provided in Appendix A.2.

The likelihood maximisation was performed using the Broyden - Fletcher - Goldfarb - Shanno (BFGS) algorithm as implemented in the *optim* package of the *stats* library in R.

5.5.1 Results

We fit for each of the 50 trading weeks a separate COGARCH model driven by a compound Poisson process with mean zero normally distributed jumps using both the SMC and QML methodologies. Recall, for parameter identifiability we assume $\mathbb{E}(L_1^2) = 1$. In this way, the variance of the jump distribution will be set as the inverse of the jump rate λ. Thus, a total of four parameters β, η, φ and λ are estimated for each trading week. The jump rate λ was estimated via (5.22). As $\mathbb{E}(L_1^2) = 1$ is fixed, the estimated λ does not factor into the QML method. However, for the SMC method it specifies the parameters of the driving compound Poisson process. For the SMC procedure, $K = 1000$ particles were employed and

this choice seemed reasonable in light of simulation evidence in Section 5.6. The fits obtained for the 50 trading weeks analysed are displayed in Figure 5.2. It is seen that estimates for β, η and φ are generally higher using the SMC method compared to the QML.

Unfortunately, in the current literature, diagnostics to determine which set of parameter values provide the better fit for the data have not been substantially developed. However, one could compare summary statistics such as the mean stationary squared volatility $\mathbb{E}(\sigma_\infty^2) := \frac{\beta}{\eta-\varphi}$ (see Proposition 4.2 of Kluppelberg et al. [2004]) as well as the lower bound of the squared volatility process $\sigma_{\min}^2 := \frac{\beta}{\eta}$ (see Proposition 2 of Kuppelberg et al. [2006]).

An estimate of $\mathbb{E}(\sigma_\infty^2)$ can also be obtained empirically by $\sum_{i=1}^n \frac{\bar{G}_i^2}{\Delta t_i}$ (see Proposition 5.1 of Kluppelberg et al. [2004]). This, along with the estimated mean stationary squared volatilities given by the QML and SMC, are displayed in Figure 5.3a. In general, we observe strong agreement between these three quantities. The QML and SMC estimates of $\sigma_{\min}^2$ are displayed in Figure 5.3b. Both quantities seem to track each other well, however the QML estimates are higher almost all of the time.

Clearly there is a discrepancy between estimates from the two methods. To better determine which of the two methods one would place more confidence in, we turn to a simulation study in the next section, where we can examine the performance of each estimator. What is found is that the QML method's estimates, in particular for η and φ, are substantially biased in comparison to those from the SMC method.

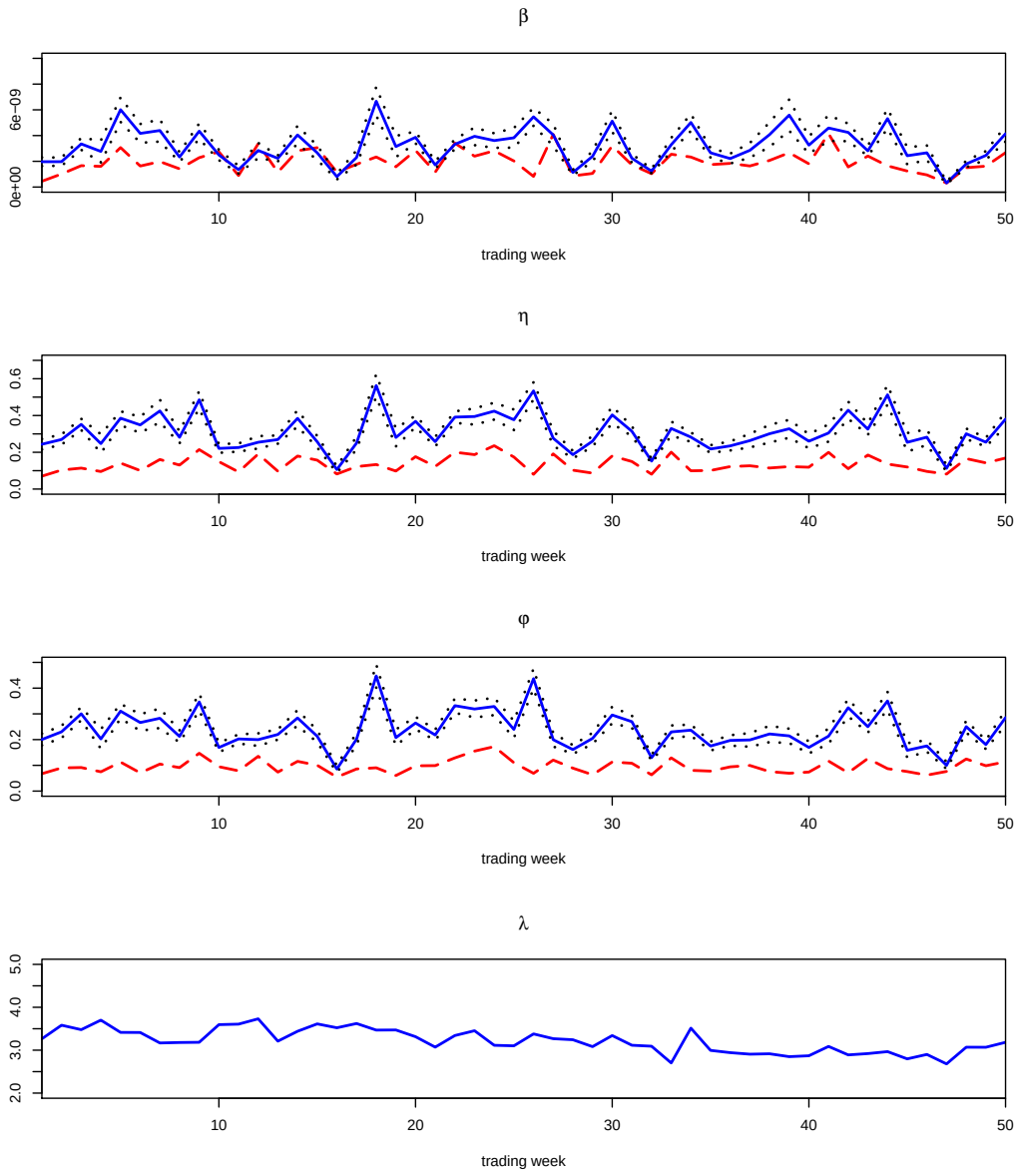
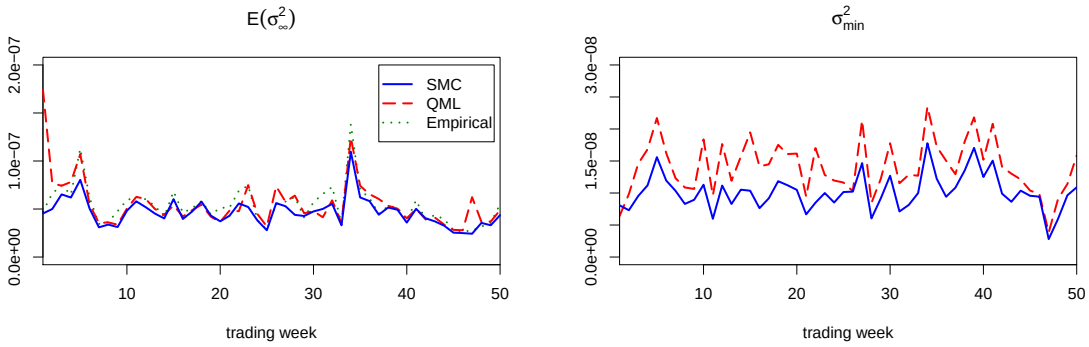


Figure 5.2: Estimated COGARCH parameters for the AUD/USD exchange rate for the 50 consecutive trading weeks. The SMC parameter estimates are given by the solid blue line with the dotted black lines giving the 95% confidence intervals. The QML parameter estimates are given by the dashed red line.



(a) QML (dashed red), SMC (solid blue) and empirically estimated (dotted green) mean stationary squared volatility $\mathbb{E}(\sigma_\infty^2)$.

(b) QML (dashed red) and SMC (solid blue) estimates for $\sigma^2_{\min}$.

Figure 5.3

5.6 Simulation studies

In this section, we perform a total of four simulation studies. The first two are based on the real data analysis of the previous section, while the last two are based on experiments in Maller et al. [2008].

5.6.1 Inspired from the real data analysis

We perform two simulation studies based on the parameter estimates and circumstances of the first trading week, the 4th of January 2015 22:00 GMT to the 9th of January 2015 22:00 GMT. This is a week where the difference between the QML and SMC are close to typical in the 50 week period under study. For both studies the driving compound Poisson process has arrival intensity $\lambda = 3.265474$ and jump size variance $\frac{1}{3.265474} = 0.3062343$. Each simulated COGARCH process is observed at exactly those times at which observations were made during the first trading week (on the adjusted time scale). Thus each data set has $n = 7183$ irregularly spaced observations with $T = 7199.495$.

For the first simulation study we simulate 1000 data sets of a COGARCH

with $\beta = 1.96771 \times 10^{-9}, \eta = 0.24372$ and $\varphi = 0.200345$ which are based on the SMC parameter estimates and for the second, we simulate 1000 data sets of a COGARCH with $\beta = 4.497526 \times 10^{-10}, \eta = 0.07017203$ and $\varphi = 0.06759063$ which are based on the QML parameter estimates.

The parameter λ was estimated using (5.22). The estimated λ's had a mean of 3.268242, a standard deviation of 0.059794 and mean estimated standard error of 0.058795 with the estimated 95% confidence intervals yielding a 94.4% empirical coverage probability (ECP_{95}), that is, the proportion of estimated 95% confidence intervals (assuming normality) that contained the true λ.

The SMC procedure was run with 250, 500, 1000, 2500 and 5000 particles allowing us to study the impact of choice in the number of particles. The results are summarised in Tables 5.1 and 5.2. Refer to Section 3.2.1 for definitions of the summary statistics (**Mean, Bias, SD, Mean SE** etc.) in these tables. As can be seen, as the number of particles increases, Bias and SD both decrease giving an overall reduction in RMSE.

For the first simulation study: In terms of β the QML estimates are less biased with lower variance in comparison to the SMC with 250 and 500 particles. As the SMC's particles are increased to 1000, the bias becomes on par with the QML although the variance of the estimates is still higher. As the particles increase to 2500, the SMC starts to become superior to the QML with less biased estimates that have a lower variance. Further improvements in bias and variance are achieved by increasing to 5000 particles. In terms of η and φ, even at 250 particles, a large reduction in bias is achieved from using the SMC instead of the QML. The variance in the QML estimates is always lower than the SMC even with 5000 particles, however even with 250 particles the SMC is a superior estimator in terms to RMSE due to the significant biases of the QML. As can be seen in the kernel density plots in Figure 5.4, the QML exhibits a substantial downward bias in estimates of η and φ while the SMC with 5000 particles is considerably less

biased but with slightly higher variability.

For the second simulation study: In terms of β the QML estimates are less biased with lower variance in comparison to the SMC with 250 particles. As the SMC's particles are increased to 500, the bias becomes on par with the QML although the variance of the estimates is still higher. As the particles increase to 1000, the SMC starts to yield less biased estimates than the QML and RMSE becomes on par. At 2500 particles, the SMC starts to become superior to the QML with less biased estimates that have a lower variance. Further improvements in bias and variance are achieved by increasing to 5000 particles. In terms of η and φ, even at 250 particles, a large reduction in bias is achieved from using the SMC instead of the QML. The variance in the QML estimates is always lower than the SMC even with 5000 particles, however even with 250 particles the SMC is a superior estimator in terms to RMSE due to the significant biases of the QML. As can be seen in the kernel density plots in Figure 5.5, the QML again exhibits a substantial downward bias in estimates of η and φ while the SMC with 5000 particles is considerably less biased but with slightly higher variability.

Regarding standard errors for the QML, Maller et al. [2008] use the inverse of the numerical Hessian. However, as the QML does not maximise the true likelihood, it is typically the case that a *sandwich* covariance estimator (see Heyde [1997]) is required. To date, the appropriate covariance estimator for the QML has not been developed and thus we do not provide standard error information for this method.

Adjustment to the inverted numerical Hessian of the SMC log-likelihood function is also in theory merited for standard error estimation due to the fact that a two-stage estimation procedure was used (see Murphy and Topel [1985]). However, given that it is generally the case and indeed exhibited here, that very precise estimates of λ are obtained from (5.22) any impact of such adjustments would be expected to be minimal to negligible. Thus the fields Mean SE and ECP_{95} in

Tables 5.1 and 5.2 are based on the standard errors obtained from the inverted numerical Hessian of the SMC log-likelihood function. We find that, as the number of particles increase:

- Mean SE and SD move closer to one another, being within close proximity of each other, for all parameters, by 5000 particles.
- ECP_{95} increases towards 95%, for all parameters, driven by lower bias and variability.

ECP_{95} not yet reaching the 95% mark by 5000 particles could be in part due to normality being asymptotic in the number of observations, not just particles.

Tables 5.3 and 5.4 display respectively for these first two simulation studies the average number of function calls and gradient approximations used per BFGS optimisation, as well as the required time to evaluate a simulated likelihood, with an integrated R and C++ implementation, using an Intel Xeon 3.06 GHz processor.

TRUE β = 1.96771E-09						
	QML	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Mean	2.17E-09	2.36E-09	2.26E-09	2.17E-09	2.11E-09	2.07E-09
Bias	2.00E-10	3.95E-10	2.94E-10	2.02E-10	1.41E-10	1.07E-10
SD	4.18E-10	6.40E-10	5.27E-10	4.53E-10	3.81E-10	3.71E-10
RMSE	4.63E-10	7.52E-10	6.03E-10	4.96E-10	4.06E-10	3.87E-10
Mean SE	NA	3.35E-10	3.45E-10	3.35E-10	3.30E-10	3.31E-10
ECP_{95}	NA	63.40%	77.90%	85.40%	91.40%	93.30%

TRUE η = 0.24372						
	QML	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Mean	0.1588	0.2840	0.2723	0.2630	0.2558	0.2521
Bias	-0.0850	0.0403	0.0286	0.0193	0.0121	0.0084
SD	0.0223	0.0453	0.0355	0.0312	0.0277	0.0264
RMSE	0.0878	0.0606	0.0456	0.0367	0.0302	0.0277
Mean SE	NA	0.0244	0.0247	0.0240	0.0237	0.0238
ECP_{95}	NA	54.30%	72.70%	83.20%	89.60%	91.80%

TRUE φ = 0.200345						
	QML	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Mean	0.1158	0.2428	0.2286	0.2186	0.2108	0.2071
Bias	-0.0845	0.0425	0.0282	0.0183	0.0105	0.0068
SD	0.0174	0.0375	0.0286	0.0250	0.0228	0.0213
RMSE	0.0863	0.0566	0.0402	0.0310	0.0250	0.0224
Mean SE	NA	0.0208	0.0203	0.0196	0.0193	0.0193
ECP_{95}	NA	47.90%	68.10%	80.30%	89.20%	90.30%

Table 5.1: Simulation study 1: Even with a small number of 250 particles, maximising the SMC log-likelihood produces superior estimates of η and φ in terms of RMSE than the QML method.

TRUE β = 4.50E-10						
	QML	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Mean	5.74E-10	6.64E-10	5.89E-10	5.41E-10	5.11E-10	5.11E-10
Bias	1.25E-10	2.15E-10	1.39E-10	9.11E-11	6.15E-11	6.14E-11
SD	2.05E-10	3.17E-10	2.66E-10	2.23E-10	2.04E-10	1.91E-10
RMSE	2.40E-10	3.83E-10	3.01E-10	2.41E-10	2.13E-10	2.01E-10
Mean SE	NA	1.42E-10	1.51E-10	1.46E-10	1.44E-10	1.46E-10
ECP_{95}	NA	47.30%	66.20%	81.20%	86.10%	89.80%

TRUE η = 0.07017203						
	QML	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Mean	0.0409	0.0805	0.0772	0.0746	0.0724	0.0719
Bias	-0.0292	0.0104	0.0071	0.0044	0.0023	0.0017
SD	0.0054	0.0141	0.0113	0.0094	0.0084	0.0082
RMSE	0.0297	0.0175	0.0133	0.0104	0.0087	0.0084
Mean SE	NA	0.0075	0.0077	0.0075	0.0074	0.0075
ECP_{95}	NA	58.30%	73.90%	86.00%	92.10%	92.80%

TRUE φ = 0.06759063						
	QML	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Mean	0.0378	0.0777	0.0743	0.0717	0.0696	0.0689
Bias	-0.0298	0.0101	0.0067	0.0041	0.0020	0.0014
SD	0.0049	0.0134	0.0106	0.0088	0.0078	0.0076
RMSE	0.0302	0.0168	0.0125	0.0097	0.0080	0.0078
Mean SE	NA	0.0073	0.0074	0.0072	0.0071	0.0071
ECP_{95}	NA	58.70%	76.10%	86.90%	92.90%	93.50%

Table 5.2: Simulation study 2: Even with a small number of 250 particles, maximising the SMC log-likelihood produces superior estimates of η and φ in terms of RMSE than the QML method.

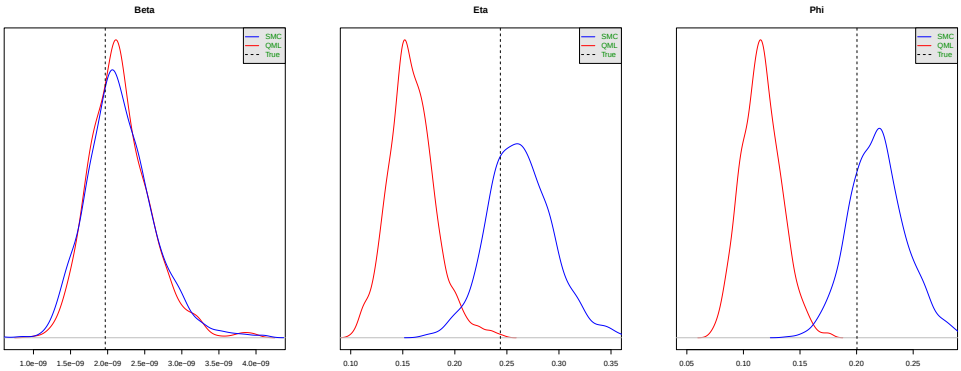


Figure 5.4: Kernel density plots for simulation study 1 - QML (red) compared to SMC (blue) with 5000 particles. The percentage bias of the mean of the QML estimates relative to the true values, respectively for β, η and φ, are 10%, -35% and -42%. The respective percentage biases obtained for the SMC estimates are considerably lower at 5.4%, 3.4% and 3.4%.

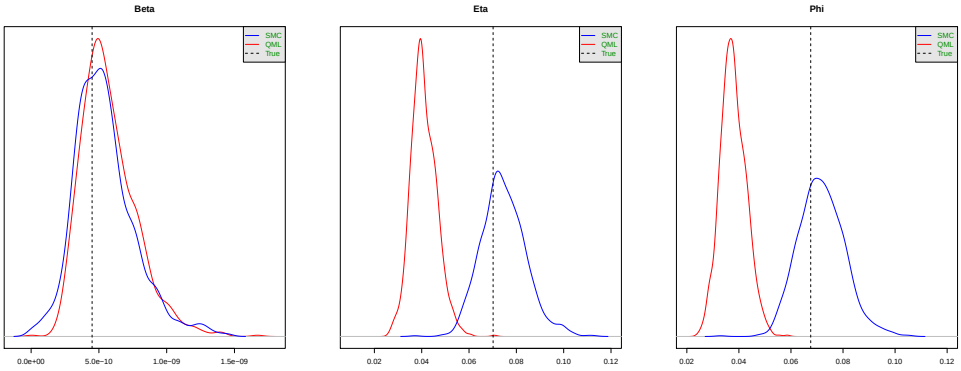


Figure 5.5: Kernel density plots for simulation study 2 - QML (red) compared to SMC (blue) with 5000 particles. The percentage bias of the mean of the QML estimates relative to the true values, respectively for β, η and φ, are 28%, -42% and -44%. The respective percentage biases obtained for the SMC estimates are considerably lower at 13.6%, 2.4% and 2.1%.

	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Average number of function calls	150.02	138.13	125.77	101.49	86.17
Average number of gradient calls	18.15	17.86	17.23	36.65	13.95
Avg. time req. for one function call (secs)	2.39	4.87	9.40	23.69	48.64

Table 5.3: Simulation study 1 - Computational speed information based on an Intel Xeon 3.06 GHz processor. Optimisation was performed using the BFGS method implemented in R with likelihood evaluations performed in C++.

	SMC 250	SMC 500	SMC 1000	SMC 2500	SMC 5000
Average number of function calls	146.08	135.80	128.57	111.93	95.18
Average number of gradient calls	14.88	15.52	16.43	16.58	16.12
Avg. time req. for one function call (secs)	2.40	4.68	9.27	23.81	48.10

Table 5.4: Simulation study 2 - Computational speed information based on an Intel Xeon 3.06 GHz processor. Optimisation was performed using the BFGS method implemented in R with likelihood evaluations performed in C++.

5.6.2 Parameter set of Maller et al. [2008]

Interestingly, the QML displayed a similar downward bias in η and φ for the simulation studies presented in Maller et al. [2008]. For comparison we repeat those experiments, presenting the results in Tables 5.5 and 5.6. In both cases a large reduction in bias is achieved using the SMC over the QML. Note that the conditional variance formula used for the QML was misspecified in Maller et al. [2008] and should actually be

$$\mathbb{E}(\bar{G}_i^2|\sigma_{t_{i-1}}^2) = \left(\sigma_{t_{i-1}}^2 - \frac{\beta}{\eta-\varphi}\right)\left(\frac{1-e^{-(\eta-\varphi)\Delta t_i}}{\eta-\varphi}\right) + \frac{\beta\Delta t_i}{\eta-\varphi}, \tag{5.23}$$

this was confirmed through correspondence with two of the authors. Our QML results have been implemented using the correct formula, which accounts for variation between the QML results presented here and those reported in Maller et al. [2008].

	$\beta = 1$		$\eta = 0.06$		$\varphi = 0.0425$		$\lambda = 1$
	QML	SMC 1000	QML	SMC 1000	QML	SMC 1000	
Mean	1.1622	1.1035	0.0535	0.0621	0.0334	0.0428	1.0010
Bias	0.1622	0.1035	-0.0065	0.0021	-0.0091	0.0003	0.0010
SD	0.5242	0.3888	0.0150	0.0122	0.0078	0.0074	0.0187
RMSE	0.5486	0.4023	0.0164	0.0124	0.0120	0.0074	0.0187
Mean SE	NA	0.3319	NA	0.0112	NA	0.0078	0.0186
ECP_{95}	NA	94.1%	NA	95.3%	NA	94.3%	94.1%

Table 5.5: Based on the equally spaced simulation study of Maller et al. [2008]. The SMC estimates are significantly less biased and have a lower variance than the QML estimates.

	$\beta = 1.5$		$\eta = 0.085$		$\varphi = 0.069$		$\lambda = 1$
	QML	SMC 1000	QML	SMC 1000	QML	SMC 1000	
Mean	1.7048	1.6438	0.0647	0.0852	0.0477	0.0679	1.0013
Bias	0.2048	0.1438	-0.0203	0.0002	-0.0213	-0.0011	0.0013
SD	1.0212	0.7878	0.0286	0.0291	0.0188	0.0228	0.0269
RMSE	1.0416	0.8008	0.0351	0.0291	0.0284	0.0228	0.0269
Mean SE	NA	0.5189	NA	0.0158	NA	0.0125	0.0267
ECP_{95}	NA	88.3%	NA	87.2%	NA	87.3%	95.2%

Table 5.6: Based on the irregularly spaced simulation study of Maller et al. [2008]. The SMC estimates are significantly less biased than the QML estimates. The variance of the QML estimates for η and φ are lower than those of the SMC, however the SMC still is the superior estimator in terms to RMSE due to the significant biases of the QML.

5.7 Concluding remarks

In conclusion, the QML, although simple to implement and fast to compute, has been shown in a number of simulation studies to exhibit a considerable downward bias in the parameters η and φ, whereas very little bias is observed in the SMC estimates of all model parameters. The downward bias in the QML estimates of η and φ are not isolated to our simulation studies and have also been observed in the simulation studies presented in Maller et al. [2008] and Bibbona and Negri [2015].

The QML employs both a first jump approximation, wherein any change in the observed process is assumed to come from one jump at the start of observation interval; as well as a normal distribution in place of the actual conditional distribution of the observed return. Which of these two approximations has the biggest impact on bias is yet to be determined.

As the QML estimates β and the long-run variance $\beta/(\eta-\varphi)$ reasonably well, by underestimating η the minimum variance β/η is overestimated, while the peaks in volatility at each jump time will be underestimated when φ is underestimated. In totality, the bias in QML will imply that the estimated volatility process in continuous time is flatter than it should be. This could having implications if one is calibrating the COGARCH model for use in option pricing, risk management or trading strategies.

In this chapter we have focussed on the case where the jump distribution is the normal distribution. Extension to other jump size distributions is straightforward. The ideas presented here could be easily adapted to the compound Poisson driven exponential COGARCH(1,1) process of Huag and Czado [2007] and the compound Poisson driven asymmetric COGARCH(1,1) process of Behme et al. [2014].

Observing financial phenomena at higher frequencies increases the presence of zeroes in the data. This supports the use of a compound Poisson driver for the increasing available high frequency financial data sets. Furthermore, any other pure jump Lévy process can be approximated by a sequence of compound Poisson processes.

Although the QML makes no assumptions about the form of the driving Lévy process, in practice if one were to model under the assumption of COGARCH dynamics, knowledge of exactly what the underlying Lévy process is, would be required to simulate sample paths for the pricing and hedging of exotic options, performance evaluation of trading strategies as well as forecasts required for risk management proposes.

Chapter 6

Extension: Likelihood inference for Markov switching GARCH(1,1) models

6.1 Introduction

It has been argued that artificially high persistence (measured as $\alpha + \beta$) obtained in empirical studies utilising the standard GARCH(1,1) specification (1.1)-(1.2) can be avoided by allowing the GARCH parameters to evolve over time; see for instance Dieobold [1986], Lamoureux and Lastrapes [1990] and Mikosch and Starica [2004]. Indeed, studies such as Schwert [1989], Hamilton and Lin [1996] and Perez-Quiros and Timmermann [2001] 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 determinis-

tically 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 [1989], 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 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 Hamilton [1989], yields what Francq and Zakoian [2008], Bauwens et al. [2010] 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

$$Y_i = \sigma_i z_i + \mu(R_i) \tag{6.1}$$

$$\sigma_i^2 = \omega(R_i) + \alpha(R_i)(Y_{i-1} - \mu(R_{i-1}))^2 + \beta(R_i)\sigma_{i-1}^2 \tag{6.2}$$

whereby $\omega(R_i), \alpha(R_i), \beta(R_i)$ and $\mu(R_i)$ 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 $\mu(\mathcal{R}) := \{\mu_1, \dots, \mu_J\} \in \mathbb{R}^J$ with $(z_i)_{i=1,\dots,N}$ a sequence of i.i.d zero mean, unit variance noise. Denote the $J \times J$ transition matrix which governs the regime evolution by $\mathbf{P}$ with elements defined as $P_{k,l} = \mathbb{P}(R_i = l | R_{i-1} = k)$. Let $\pi_k := \mathbb{P}(R_1 = k), k = 1, \dots, J$ be the stationary distribution of the regime Markov chain. It was proven in Francq et al. [2001] that

a strictly stationary solution to (6.1)-(6.2) exists if and only if

$$\sum_{k=1}^{J} \pi_k \mathbb{E}(\log(\alpha_k z_1^2 + \beta_k)) < 0. \tag{6.3}$$

Therefore, for strict stationarity it need not be the case that under all regimes the stability condition of Nelson [1990] (see (1.3)) 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 Bauwens et al. [2010] that the higher the unconditional probabilities of being in regimes with stable GARCH parameters, the higher the persistence unstable regimes can assume whilst maintaining the condition (6.3).

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 [1996], Dueker [1997] and Klaassen [2002] 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. [2004] avoids path dependency by separating the GARCH dynamics from the regime process.

Despite these alternative specifications where MLE is feasible, the specification (6.1)-(6.2) as stated by Klaassen [2002] and Haas et al. [2004] 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 (6.1)-(6.2), which bypass exact evaluation of the likelihood, have been proposed. To clarify, hereafter by MS-GARCH we are referring to the specification

(6.1)-(6.2).

A number of Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures have been developed for these MS-GARCH models. This class of method 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. [2010] 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 Bauwens et al. [2014], 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 Billio et al. [2016].

In terms of frequentist methods for this class of model, the first was developed by Augustyniak [2014] who employed the Monte Carlo Expectation Maximisation algorithm, simulating regimes from the single-move Gibbs sampler of Bauwens et al. [2010]. Later, a non-simulation based frequentist approach was developed by Augustyniak et al. [2017] based on a collapsing procedure that generalises those of Gray [1996], Dueker [1997] and Klaassen [2002].

Other developments towards estimation are the works of Francq and Zakoian [2008] who propose a generalised method of moments approach, Bildirici and Ersin [2014] who utilise neural networks and Elliott et al. [2012] who propose a Viterbi based technique for sampling the regimes.

In employing a particle filter approach, Bauwens et al. [2014] was able to obtain, through simulation, an estimated likelihood function. However, particle filter (a.k.a SMC) likelihood estimates although consistent and unbiased (cf. Del Moral [1996]) are simulated in such a way that make them prone to being discontinuous functions of the parameters (as explained in Section 2.2), thus inhibiting their use in numerical optimisers, which is why Bauwens et al. [2014] proceeded with a MCMC approach.

The purpose of this chapter is modification of the standard SMC likelihood estimation procedure to yield estimated likelihood surfaces that are amenable to numerical optimisation 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 Augustyniak [2014] and not resorting to the auxiliary model structure used in Augustyniak et al. [2017]. Furthermore, we illustrate the capability of this SMC approach in extending to parameter estimation of MS-GARCH(1,1) series for which one has missing observations - a so far unexplored real world application encountered every so often.

Organisation of this chapter is as follows. Section 6.2 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. An overview of the collapsed model structure used by Augustyniak et al. [2017] for approximating the MS-GARCH(1,1) model is provided in Section 6.3. Our modified SMC procedure, for parameter estimation of MS-GARCH(1,1) models through simulated maximum likelihood is introduced in Section 6.4. Extension of the SMC procedure to the, occasionally encountered, although unchartered problem of 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 6.5. Section 6.6.1 compares through simulation studies the performance of the SMC estimation method to the estimates obtained using the approximating model structure of Augustyniak et al. [2017]. 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 6.6.2. Two real world applications of the SMC method are illustrated. Section

6.7.1 analyses the volatility of stock returns allowing for structural changes brought about by phenomena such as changing business cycles. Section 6.7.2 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 chapter closes with some concluding remarks in Section 6.8.

6.2 Likelihood

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

$$h_i(Y_{1:i-1}, R_{1:i}) = \omega(R_i) + \alpha(R_i)\Big(Y_{i-1} - \mu(R_{i-1})\Big)^2 + \beta(R_i)h_{i-1}(Y_{1:i-2}, R_{1:i-1}) \tag{6.4}$$

for $i = 2, \ldots, n$ with $h_1(0, R_1) = \mathbb{E}(\sigma_1^2|R_1)$. From Francq and Zakoian [2008], $\mathbb{E}(\sigma_1^2|R_1 = r), r = 1, 2$ are obtained by solving the stationary equations

$$\begin{bmatrix} 1 - P_{1,1}(\alpha_1 + \beta_1) & -P_{2,1}(\alpha_1 + \beta_1) \\ -P_{1,2}(\alpha_2 + \beta_2) & 1 - P_{2,2}(\alpha_2 + \beta_2) \end{bmatrix} \begin{bmatrix} \pi_1 \mathbb{E}(\sigma_1^2|R_1 = 1) \\ \pi_2 \mathbb{E}(\sigma_1^2|R_1 = 2) \end{bmatrix} = \begin{bmatrix} P_{1,1}\omega_1 & P_{2,1}\omega_1 \\ P_{1,2}\omega_2 & P_{2,2}\omega_2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \tag{6.5}$$

whereby $\pi_1 := \frac{P_{2,1}}{P_{1,2}+P_{2,1}}$ and $\pi_2 := \frac{P_{1,2}}{P_{1,2}+P_{2,1}}$. Given $Y_{1:i-1}$, σ_i^2 has a discrete probability distribution with Figure 6.1 displaying (for i up to 3) how the sample space

of possible values for σ_i^2 expands as i increases. At $i = 1$ we have

$$\{\sigma_1^2, R_1\} = \begin{cases} \{h_1(0, \{1\}), 1\} & \text{w.p } \pi_1 \\ \{h_1(0, \{2\}), 2\} & \text{w.p } \pi_2. \end{cases} \tag{6.6}$$

Then at $i = 2$ we have

$$\{\sigma_2^2, R_2\}|Y_1 = \begin{cases} \{h_2(Y_{1:1}, \{1, 1\}), 1\} & \text{w.p } \pi_1 P_{1,1} \frac{p(Y_1|\sigma_1^2 = h_1(0,\{1\}), R_1=1)}{p(Y_1)} \\ \{h_2(Y_{1:1}, \{1, 2\}), 2\} & \text{w.p } \pi_1 P_{1,2} \frac{p(Y_1|\sigma_1^2 = h_1(0,\{1\}), R_1=1)}{p(Y_1)} \\ \{h_2(Y_{1:1}, \{2, 1\}), 1\} & \text{w.p } \pi_2 P_{2,1} \frac{p(Y_1|\sigma_1^2 = h_1(0,\{2\}), R_1=2)}{p(Y_1)} \\ \{h_2(Y_{1:1}, \{2, 2\}), 2\} & \text{w.p } \pi_2 P_{2,2} \frac{p(Y_1|\sigma_1^2 = h_1(0,\{2\}), R_1=2)}{p(Y_1)}, \end{cases} \tag{6.7}$$

with

$$p(Y_1) = \pi_1 p(Y_1|\sigma_1^2 = h_1(0, \{1\}), R_1 = 1) + \pi_2 p(Y_1|\sigma_1^2 = h_1(0, \{2\}), R_1 = 2). \tag{6.8}$$

For general i, there are 2^i possible realisations of $R_{1:i}$ which we denote by $R_{1:i}^{(k)}$, $k = 1, \ldots, 2^i$. These in turn, given $Y_{1:i-1}$, result in 2^i possible realisations of σ_i^2 given by $h_i(Y_{1:i-1}, R_{1:i}^{(k)})$, $k = 1, \ldots, 2^i$ that occur, conditional on $Y_{1:i-1}$, with probability $w_i^{(k)}$, whereby

$$w_i^{(k)} := \pi_{R_1^{(k)}} A_1^{(k)} \ldots A_{i-1}^{(k)} \tag{6.9}$$

$$A_{i-1}^{(k)} := P_{R_{i-1}^{(k)}, R_i^{(k)}} \frac{p(Y_{i-1}|\sigma_{i-1}^2 = h_{i-1}(Y_{1:i-2}, \{R_1^{(k)}, \ldots, R_{i-1}^{(k)}\}), R_{i-1} = R_{i-1}^{(k)})}{p(Y_{i-1}|Y_{1:i-2})} \tag{6.10}$$

with $R_j^{(k)}$, $j = 1, \ldots, i$ being the j-th component of $R_{1:i}^{(k)}$. That is,

$$\mathbb{P}(\sigma_i^2, R_i|Y_{1:i-1}) = \sum_{k=1}^{2^i} w_i^{(k)} \delta_{\{h_i(Y_{1:i-1}, R_{1:i}^{(k)}\}), R_i^{(k)}\}}(\sigma_i^2, R_i) \tag{6.11}$$

where $\delta_{\{a,b\}}$ is the bivariate Dirac measure at the point (a, b). The quantity (6.11) is a sum of size 2^i. When the number of observations n is not too large, one could compute the likelihood via

$$p(Y_{1:n}) = \prod_{i=2}^{n} p(Y_i|Y_{1:i-1})p(Y_1) \tag{6.12}$$

$$= \prod_{i=2}^{n} \Big(\sum_{\sigma_i^2, R_i} p(Y_i|\sigma_i^2, R_i)\mathbb{P}(\sigma_i^2, R_i|Y_{1:i-1}) \Big) p(Y_1). \tag{6.13}$$

However, in most practical situations n will be too large to evaluate (6.13) using (6.11). The next two sections detail methods of trimming, 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 in a manner such that the resultant approximated likelihood surfaces are amenable to numerical optimisation. The first procedure discussed is the general collapsing procedure (GCP) of Augustyniak et al. [2017] which trims the branches by combining pairs of branches. The second procedure is a modified SMC approach, trimming by resampling the branches, weighting them by how likely they would result in the given observation.

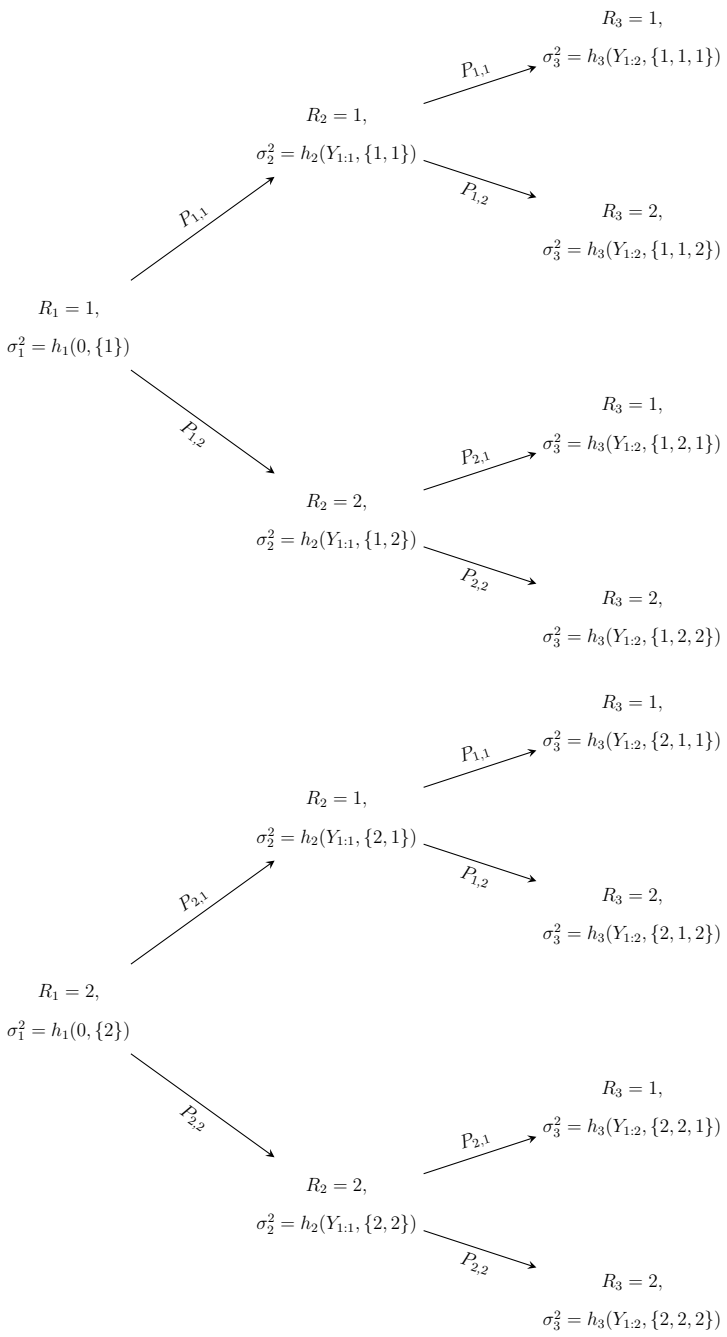


Figure 6.1: Sample space of $\sigma_i^2|Y_{1:i-1}$ for $i = 1, 2, 3$.

6.3 General collapsing procedure

The GCP of Augustyniak et al. [2017] is as follows. Assume at some time point q we have $h_q(Y_{1:q-1}, R_{1:q})$ and $p(R_{1:q}|Y_{1:q})$ for all the 2^q possible realisations of $R_{1:q}$. In order to prevent an expansion of the tree to 2^{q+1} branches at time $q+1$, the GCP will first collapse the tree into 2^{q-1} branches then expand back to 2^q branches for the next time step. To this end, note that there are 2^{q-1} permutations of $R_{2:q}$ which we will denote by $R_{2:q}^{(k)}$ for $k = 1, \ldots, 2^{q-1}$. For each $k = 1, \ldots, 2^{q-1}$, one then "collapses" the two realisations of $R_{1:q}$ that coincide with $R_{2:q}^{(k)}$ to the following quantity

$$\nu_q(Y_{1:q}, R_{2:q}^{(k)}) := \frac{1}{a_q^{(k)}} \sum_{i=1}^{2} p(R_{1:q} = \{i, R_{2:q}^{(k)}\}|Y_{1:q}) h_q(Y_{1:q-1}, \{i, R_{2:q}^{(k)}\}) \tag{6.14}$$

where $a_q^{(k)} := \sum_{i=1}^{2} p(R_{1:q} = \{i, R_{2:q}^{(k)}\}|Y_{1:q})$. The idea is then to expand the tree back to 2^q nodes $(\tilde{h}_{q+1}(Y_{1:q}, R_{2:q+1}^{(j)}), \tilde{w}_{q+1}^{(j)}, R_{2:q+1}^{(j)})$ for $j = 1, \ldots, 2^q$ at the next time step $q+1$, whereby

$$\tilde{h}_{q+1}(Y_{1:q}, \{R_{2:q}^{(k)}, l\}) := \omega(l) + \alpha(l)\Big(Y_q - \mu(R_q^{(k)})\Big)^2 + \beta(l)\nu_q(Y_{1:q}, R_{2:q}^{(k)}) \tag{6.15}$$

$$b_{q+1}^{(2(k-1)+l)} := p(Y_{q+1}|\sigma_{q+1}^2 = \tilde{h}_{q+1}(Y_{1:q}, \{R_{2:q}^{(k)}, l\}), R_{q+1} = l) P_{R_q^{(k)},l} a_q^{(k)} \tag{6.16}$$

$$\tilde{w}_{q+1}^{(2(k-1)+l)} := b_{q+1}^{(2(k-1)+l)} / \sum_{j=1}^{2^q} b_{q+1}^{(j)} \tag{6.17}$$

$$R_{2:q+1}^{(2(k-1)+l)} := \{R_{2:q}^{(k)}, l\} \tag{6.18}$$

for $l = 1, 2$ and $k = 1, \ldots, 2^{q-1}$. Then for $i \in \{q+2, \ldots, N\}$ one maintains a total of 2^q nodes $(\tilde{h}_i(Y_{1:i-1}, R_{i+1-q:i}^{(j)}), \tilde{w}_i^{(j)}, R_{i+1-q:i}^{(j)})$, $j = 1, \ldots, 2^q$ for each time point i, by collapsing the 2^q nodes at the previous time point $i-1$ to the 2^{q-1} auxiliary

nodes $(\nu_{i-1}(Y_{1:i-1}, R^{(k)}_{i+1-q:i-1}), a^{(k)}_{i-1})$, $k = 1, \ldots, 2^{q-1}$ whereby

$$\begin{aligned}\nu_{i-1}(Y_{1:i-1}, R^{(k)}_{i+1-q:i-1}) := \frac{1}{a^{(k)}_{i-1}} \sum_{j=1}^{2^q} \Big(& \tilde{h}_{i-1}(Y_{1:i-2}, R^{(j)}_{i-q:i-1}) \\ & \times \tilde{w}^{(j)}_{i-1} \mathbb{I}(R^{(j)}_{i+1-q:i-1} = R^{(k)}_{i+1-q:i-1}) \Big)\end{aligned} \tag{6.19}$$

$$a^{(k)}_{i-1} := \sum_{j=1}^{2^q} \tilde{w}^{(j)}_{i-1} \mathbb{I}(R^{(j)}_{i+1-q:i-1} = R^{(k)}_{i+1-q:i-1}) \tag{6.20}$$

Then expanding these auxiliary nodes to obtain $\tilde{h}_i(Y_{1:i-1}, R^{(j)}_{i+1-q:i})$ and $\tilde{w}^{(j)}_i$ for $j = 1, \ldots, 2^q$ using (with $i-1$ in place of q and $R^{(k)}_{i+1-q:i-1}$ in place of $R^{(k)}_{2:q}$) the equations (6.15)-(6.17). Furthermore, $R^{(2(j-1)+l)}_{i+1-q:i} = \{R^{(j)}_{i+1-q:i-1}, l\}$ for $l = 1, 2$ and $j = 1, \ldots, 2^{q-1}$.

Formally, the GCP model structure is as follows

$$\tilde{Y}_i = \tilde{\sigma}_i z_i + \mu(R_i) \tag{6.21}$$

$$\tilde{\sigma}^2_i = \begin{cases} h_i(\tilde{Y}_{1:i-1}, R_{1:i}) \text{ for } i = 1, \ldots, q \\ \tilde{h}_i(\tilde{Y}_{1:i-1}, R_{i+1-q:i}) \text{ for } i = q+1, \ldots, N. \end{cases} \tag{6.22}$$

Up to time q the GCP follows the MS-GARCH model structure, thereafter it follows the collapsed model structure. An illustration of the GCP for $q = 3$ is provided in Figure 6.2. The GCP generalises the collapsing procedures of Dueker [1997] and Klaassen [2002] which correspond to the GCP when $q = 2$ and $q = 1$ respectively.

The log-likelihood of the collapsed model is

$$\begin{aligned}\log(\mathcal{L}(\tilde{Y}_{1:N})) = & \sum_{i=1}^{q} \log\Big(\sum_{j=1}^{2^i} p(\tilde{Y}_i | \tilde{Y}_{1:i-1}, R_{1:i} = R^{(j)}_{1:i}) p(R_{1:i} = R^{(j)}_{1:i} | \tilde{Y}_{1:i-1}) \Big) \\ & + \sum_{i=q+1}^{N} \log(\sum_{j=1}^{2^q} b^{(j)}_i).\end{aligned} \tag{6.23}$$

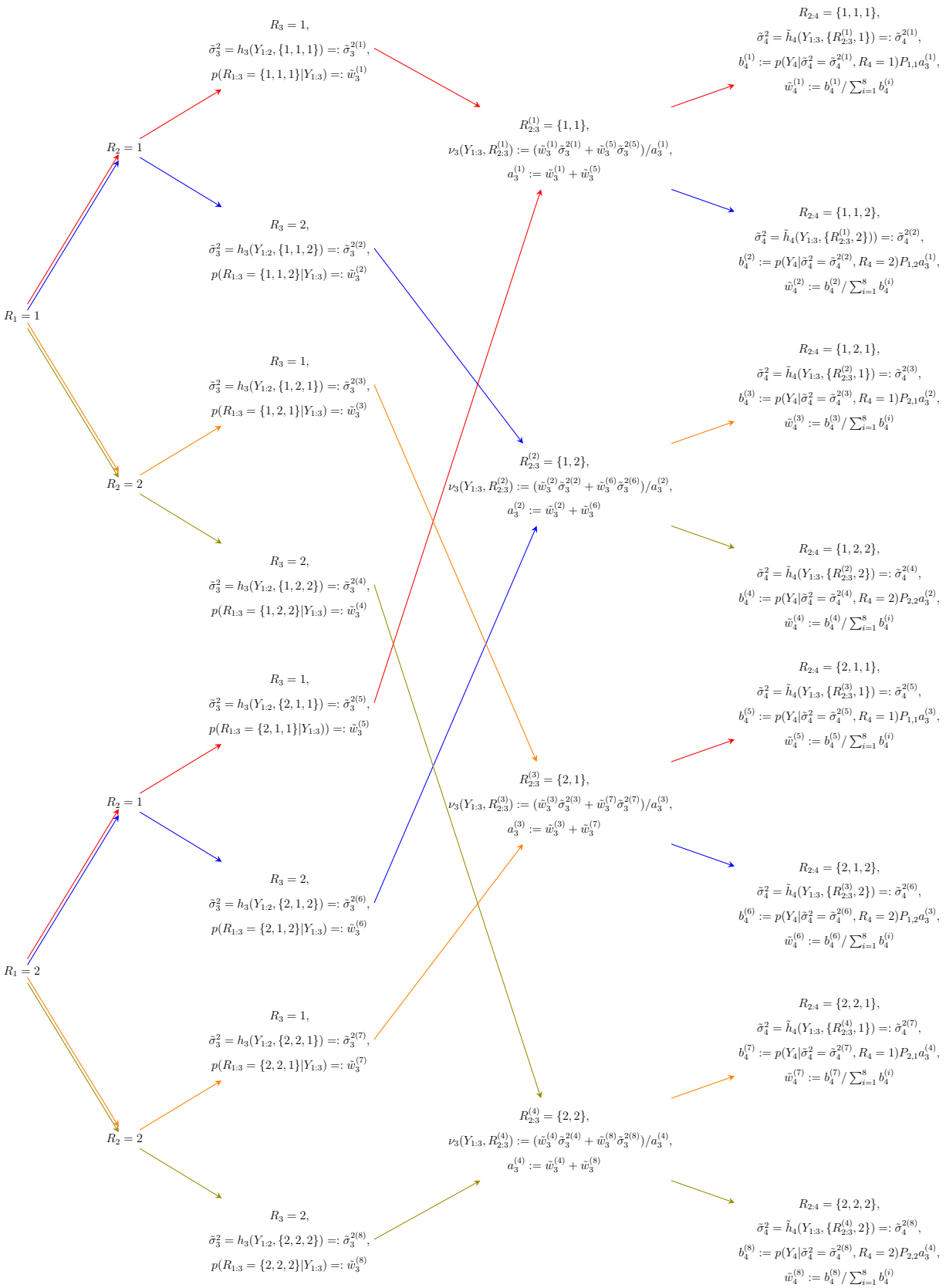


Figure 6.2: Illustration of the general collapsing procedure for $q = 3$.

Augustyniak et al. [2017] proposes obtaining parameter estimates by assuming the MS-GARCH observations $Y_{1:N}$ come from the collapsed model (6.21)-(6.22) and maximising (6.23). It was reasoned in Augustyniak et al. [2017] that (6.23) is a continuous function of the model parameters.

6.4 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

$$\mathbb{P}(R_q = r|Y_{1:q}) = \frac{\sum_{j=1}^{2^{q-1}} p(Y_q|\sigma_q^2 = \sigma_{q,r}^{2(j)}, R_q = r)\mathbb{P}(R_{1:q} = \{R_{1:q-1}^{(j)}, r\}|Y_{1:q-1})}{\sum_{s=1}^{2}\sum_{j=1}^{2^{q-1}} p(Y_q|\sigma_q^2 = \sigma_{q,s}^{2(j)}, R_q = s)\mathbb{P}(R_{1:q} = \{R_{1:q-1}^{(j)}, s\}|Y_{1:q-1})}, \tag{6.24}$$

$$\mathbb{P}(\sigma_q^2|Y_{1:q}, R_q = r) = \sum_{j=1}^{2^{q-1}} w_{q,r}^{(j)} \delta_{\{\sigma_{q,r}^{2(j)}\}}(\sigma_q^2) \tag{6.25}$$

for each $r = 1, 2$, where

$$w_{q,r}^{(j)} := \frac{p(Y_q|\sigma_q^2 = \sigma_{q,r}^{2(j)}, R_q = r)\mathbb{P}(R_{1:q} = \{R_{1:q-1}^{(j)}, r\}|Y_{1:q-1})}{\sum_{k=1}^{2^{q-1}} p(Y_q|\sigma_q^2 = \sigma_{q,r}^{2(k)}, R_q = r)\mathbb{P}(R_{1:q} = \{R_{1:q-1}^{(k)}, r\}|Y_{1:q-1})}, \tag{6.26}$$

$$\sigma_{q,r}^{2(j)} = h_q(Y_{1:q-1}, \{R_{1:q-1}^{(j)}, r\}) \tag{6.27}$$

for $j = 1, \ldots, 2^{q-1}$. The idea going forward is to approximate $\mathbb{P}(\sigma_{i+1}^2, R_{i+1}|Y_{1:i})$, $i = q, \ldots, N-1$, 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 $(\sigma_{i+1}^{2(j)}, R_{i+1}^{(j)})$, $j = 1, \ldots, 2^q$. Essentially we trim the tree so to speak, never allowing the possible branch nodes $(\sigma_{i+1}^2, R_{i+1})$, $i = q, \ldots, N-1$, 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

$$p(Y_{i+1}|Y_{1:i}) = \sum_{\sigma^2_{i+1},R_{i+1}} p(Y_{i+1}|\sigma^2_{i+1},R_{i+1})\mathbb{P}(\sigma^2_{i+1},R_{i+1}|Y_{1:i}) \tag{6.28}$$

$$\approx \sum_{j=1}^{2^q}\Big(p(Y_{i+1}|\sigma^2_{i+1}=\sigma^{2(j)}_{i+1},R_{i+1}=R^{(j)}_{i+1}) \times \hat{\mathbb{P}}(\sigma^2_{i+1}=\sigma^{2(j)}_{i+1},R_{i+1}=R^{(j)}_{i+1}|Y_{1:i})\Big) \tag{6.29}$$

$$=: \hat{p}(Y_{i+1}|Y_{1:i}). \tag{6.30}$$

Now, note that

$$\mathbb{P}(\sigma^2_{i+1},R_{i+1}|Y_{1:i}) = \sum_{\sigma^2_i,R_i} \mathbb{P}(\sigma^2_{i+1},R_{i+1}|\sigma^2_i,R_i,Y_i)\mathbb{P}(\sigma^2_i,R_i|Y_{1:i}) \tag{6.31}$$

$$\mathbb{P}(\sigma^2_{i+1},R_{i+1}|\sigma^2_i,R_i,Y_i) = \begin{cases} \big(h(Y_i,\sigma^2_i,R_i,1),1\big) & \text{w.p } P_{R_i,1} \\ \big(h(Y_i,\sigma^2_i,R_i,2),2\big) & \text{w.p } P_{R_i,2}=1-P_{R_i,1}, \end{cases} \tag{6.32}$$

where $h(y,\sigma^2,r,k) := \omega_k+\alpha_k(y-\mu_r)^2+\beta_k\sigma^2$. Thus to obtain $\hat{\mathbb{P}}(\sigma^2_{i+1},R_{i+1}|Y_{1:i})$ for which the couple (σ^2_{i+1},R_{i+1}) is restricted to 2^q possible values, due to the regime branching mechanism (6.32), one can first obtain $\hat{\mathbb{P}}(\sigma^2_i,R_i|Y_{1:i})$ for which (σ^2_i,R_i) is restricted to 2^{q-1} possible values, then replace $\mathbb{P}(\sigma^2_i,R_i|Y_{1:i})$ with $\hat{\mathbb{P}}(\sigma^2_i,R_i|Y_{1:i})$ in (6.31) to obtain the required $\hat{\mathbb{P}}(\sigma^2_{i+1},R_{i+1}|Y_{1:i})$. We begin at the first step $i=q$ and discuss several ways $\hat{\mathbb{P}}(\sigma^2_q,R_q|Y_{1:q})$ could be obtained.

- The first is to resample $K=2^{q-1}$ unique pairs of (σ^2_q,R_q) from the discrete distribution

$$\mathbb{P}(\sigma^2_q,R_q|Y_{1:q}) = \sum_{r=1}^{2}\Big(\mathbb{P}(R_q=r|Y_{1:q})\sum_{j=1}^{2^{q-1}} w^{(j)}_{q,r}\delta_{\{\sigma^{2(j)}_{q,r},r\}}(\sigma^2_q,R_q)\Big). \tag{6.33}$$

  This however, is the bootstrap resampling of Gordon et al. [1993] for which we have explained (see Section 2.2) produces SMC likelihood surfaces not amenable to numerical optimisers. The continuous resampling procedure of

Pitt and Malik [2011] cannot be applied to resample the pair (σ_q^2, R_q).

- The second is to 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 $\mathbb{P}(\sigma_q^2|Y_{1:q}, R_q = 2)$. However, K_1 and K_2 via $\mathbb{P}(R_q|Y_{1:q})$ are dependent on θ and drastically different σ_q^2 could be drawn 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.

- The third is to proceed as in the second, except, to prevent K_1 and K_2 changing across parameter changes, setting $K_1 = K_2 = K/2$. Eliminating the source of discontinuity of the second procedure, this third procedure is conducive in providing continuous SMC likelihood surfaces.

Proceeding as per the third procedure, initialise

$$\hat{\mathbb{P}}(\sigma_q^2|Y_{1:q}, R_q) = \mathbb{P}(\sigma_q^2|Y_{1:q}, R_q) \tag{6.34}$$

$$\hat{\mathbb{P}}(R_q|Y_{1:q}) = \mathbb{P}(R_q|Y_{1:q}) \tag{6.35}$$

and we detail how to proceed from $\hat{\mathbb{P}}(\sigma_i^2|Y_{1:i}, R_i)$ and $\hat{\mathbb{P}}(R_i|Y_{1:i})$, $i = q, \ldots, N-1$ 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[1] to the discrete distribution $\hat{\mathbb{P}}(\sigma_i^2|Y_{1:i}, R_i = r)$, denote these draws by $\hat{\sigma}_{i,r}^{2(j)}$, $j = 1, \ldots, 2^{q-2}$. Then construct the empirical distribution

$$\hat{\mathbb{P}}(\sigma_i^2, R_i|Y_{1:i}) = \sum_{r=1}^{2} \Big(\frac{\hat{\mathbb{P}}(R_i = r|Y_{1:i})}{K/2} \sum_{j=1}^{2^{q-2}} \delta_{\{\hat{\sigma}_{i,r}^{2(j)}, r\}}(\sigma_q^2, R_q) \Big), \tag{6.36}$$

[1] Details of how to construct this continuous approximation are provided later in Section 6.4.1.

for which replacing $\mathbb{P}(\sigma_i^2, R_i|Y_{1:i})$ with (6.36) in (6.31) and using (6.32) we arrive at

$$\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i}) = \sum_{j=1}^{2^{q-1}} \bar{w}_{i+1,r}^{(j)} \delta_{\{\sigma_{i+1,r}^{2(j)}, r\}}(\sigma_{i+1}^2, R_{i+1}), \tag{6.37}$$

for $r = 1, 2$, whereby

$$\bar{w}_{i+1,r}^{(j)} = \begin{cases} P_{1,r}\hat{\mathbb{P}}(R_i = 1|Y_{1:i})(2/K) & \text{for } j = 1, \ldots, 2^{q-2} \\ P_{2,r}\hat{\mathbb{P}}(R_i = 2|Y_{1:i})(2/K) & \text{for } j = 2^{q-2}+1, \ldots, 2^{q-1} \end{cases} \tag{6.38}$$

$$\sigma_{i+1,r}^{2(j)} = \begin{cases} h(Y_i, \hat{\sigma}_{i,1}^{2(j)}, 1, r) & \text{for } j = 1, \ldots, 2^{q-2} \\ h(Y_i, \hat{\sigma}_{i,2}^{2(j-2^{q-2})}, 2, r) & \text{for } j = 2^{q-2}+1, \ldots, 2^{q-1}. \end{cases} \tag{6.39}$$

From (6.29) and (6.37) we then obtain the approximated likelihood component

$$\hat{p}(Y_{i+1}|Y_{1:i}) = \sum_{r=1}^{2} \sum_{j=1}^{2^{q-1}} \bar{w}_{i+1,r}^{(j)} p(Y_{i+1}|\sigma_{i+1}^2 = \sigma_{i+1,r}^{2(j)}, R_{i+1} = r). \tag{6.40}$$

Next, using (6.37) and (6.40) we obtain $\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i+1})$ via

$$\mathbb{P}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i+1}) = \frac{p(Y_{i+1}|\sigma_{i+1}^2, R_{i+1})\mathbb{P}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i})}{p(Y_{i+1}|Y_{1:i})} \tag{6.41}$$

$$\approx \frac{p(Y_{i+1}|\sigma_{i+1}^2, R_{i+1})\hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i})}{\hat{p}(Y_{i+1}|Y_{1:i})} \tag{6.42}$$

$$= \sum_{j=1}^{2^{q-1}} \grave{w}_{i+1,r}^{(j)} \delta_{\{\sigma_{i+1,r}^{2(j)}, r\}}(\sigma_{i+1}^2, R_{i+1}) \tag{6.43}$$

$$:= \hat{\mathbb{P}}(\sigma_{i+1}^2, R_{i+1} = r|Y_{1:i+1}) \tag{6.44}$$

where

$$\grave{w}_{i+1,r}^{(j)} = \frac{\bar{w}_{i+1,r}^{(j)} p(Y_{i+1}|\sigma_{i+1}^2 = \sigma_{i+1,r}^{2(j)}, R_{i+1} = r)}{\sum_{s=1}^{2}\sum_{k=1}^{2^{q-1}} \bar{w}_{i+1,s}^{(k)} p(Y_{i+1}|\sigma_{i+1}^2 = \sigma_{i+1,s}^{2(k)}, R_{i+1} = s)}. \tag{6.45}$$

Furthermore,

$$\mathbb{P}(R_{i+1} = r|Y_{1:i+1}) = \sum_{\sigma^2_{i+1}} \mathbb{P}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1}) \tag{6.46}$$

$$\approx \sum_{\sigma^2_{i+1}} \hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1}) \tag{6.47}$$

$$= \sum_{j=1}^{2^{q-1}} \grave{w}^{(j)}_{i+1,r} =: \hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1}) \tag{6.48}$$

$$\hat{\mathbb{P}}(\sigma^2_{i+1}|Y_{1:i+1}, R_{i+1} = r) := \frac{\hat{\mathbb{P}}(\sigma^2_{i+1}, R_{i+1} = r|Y_{1:i+1})}{\hat{\mathbb{P}}(R_{i+1} = r|Y_{1:i+1})} \tag{6.49}$$

$$= \sum_{j=1}^{2^{q-1}} \hat{w}^{(j)}_{i+1,r} \delta_{\{\sigma^{2(j)}_{i+1,r}\}}(\sigma^2_{i+1}). \tag{6.50}$$

where

$$\hat{w}^{(j)}_{i+1,r} := \grave{w}^{(j)}_{i+1,r} / \sum_{k=1}^{2^{q-1}} \grave{w}^{(j)}_{i+1,r}, \ j = 1, \ldots, 2^{q-1}. \tag{6.51}$$

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

$$\hat{\mathcal{L}}(Y_{1:N}) = p(Y_{1:q}) \prod_{i=q}^{N-1} \hat{p}(Y_{i+1}|Y_{1:i}). \tag{6.52}$$

6.4.1 Continuous approximation to $\hat{\mathbb{P}}(\sigma^2_i|Y_{1:i}, R_i)$

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

$$\hat{\mathbb{P}}(\sigma^2_i|Y_{1:i}, R_i = r) = \begin{cases} \mathbb{P}(\sigma^2_q|Y_{1:q}, R_q = r) = \sum_{j=1}^{2^{q-1}} w^{(j)}_{q,r} \delta_{\{\sigma^{2(j)}_{q,r}\}}(\sigma^2_q) & ,i = q \\ \sum_{j=1}^{2^{q-1}} \hat{w}^{(j)}_{i,r} \delta_{\{\sigma^{2(j)}_{i,r}\}}(\sigma^2_i) & ,i = q+1 \ldots, N-1. \end{cases} \tag{6.53}$$

We detail here how these continuous approximations, for smooth SMC likelihoods, are constructed. For $i = q$ the weights $w_{q,r}^{(j)}$, see (6.26), for a given r are in addition to σ_q^2 also functions of $R_{1:i-1}$. Thus, as explained in Section 4.1 we need to smooth the weights $w_{q,r}^{(j)}$ such that $\sigma_{q,r}^{2(j)}$ that are close together have weights close in magnitude before application of the procedure from Pitt and Malik [2011]. The kernel smoothing approach of Section 4.1 achieves this. Proceeding, we calculate

$$\dot{w}_{q,r}^{(j)} = \frac{\sum_{k=1}^{2^{q-1}} w_{q,r}^{(j)} \phi\Big((\sigma_{q,r}^{2(k)} - \sigma_{q,r}^{2(j)})/h\Big)}{\sum_{k=1}^{2^{q-1}} \phi\Big((\sigma_{q,r}^{2(k)} - \sigma_{q,r}^{2(j)})/h\Big)} \tag{6.54}$$

$$\ddot{w}_{q,r}^{(j)} = \frac{\dot{w}_{q,r}^{(j)}}{\sum_{k=1}^{2^{q-1}} \dot{w}_{q,r}^{(k)}} \text{ for } j = 1, \ldots, 2^{q-1}, \tag{6.55}$$

where $\phi(.)$ is the standard Gaussian density and $h = c/2^{q-1}$ for a very small number c. One then can proceed by means of Pitt and Malik [2011], see Section 2.3, to obtain a continuous approximation to

$$\ddot{\mathbb{P}}(\sigma_q^2 \leq x | Y_{1:q}, R_q = r) = \sum_{j=1}^{2^{q-1}} \ddot{w}_{q,r}^{(j)} \mathbb{I}(\sigma_{q,r}^{2(j)} \leq x) \tag{6.56}$$

to continuously resample the σ_q^2, in a manner conducive with smooth SMC likelihoods.

Now, in the case of $i = q+1, \ldots, N-1$ note that the weights $\hat{w}_{i,r}^{(j)}$, see (6.51), (6.45) and (6.38) 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 from (6.51), (6.45) and (6.38) the weights $\hat{w}_{i,r}^{(j)}$, $j = 1, \ldots, 2^{q-2}$ correspond to $R_{i-1} = 1$ and differ only through $\sigma_{i,r}^{2(j)}$, whereby 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 $k, l = 1, \ldots, 2^{q-2}$ are close in proximity, so too are their respective weights $\hat{w}_{i,r}^{(k)}$ and $\hat{w}_{i,r}^{(l)}$. Likewise, the weights $\hat{w}_{i,r}^{(j)}$, $j = 2^{q-2}+1, \ldots, 2^{q-1}$ which correspond to $R_{i-1} = 2$ possess the same attribute.

Thus one can continuously approximate the individual quantities $A(x) :=$

$\sum_{j=1}^{2^{q-2}} \hat{w}_{i,r}^{(j)} \mathbb{I}(\sigma_{i,r}^{2(j)} \leq x)$ and $B(x) := \sum_{j=1}^{2^{q-2}} \hat{w}_{i,r}^{(j+2^{q-2})} \mathbb{I}(\sigma_{i,r}^{2(j+2^{q-2})} \leq x)$, then combine both continuous approximations, keeping in mind that the sum of continuous functions is also a continuous function.

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

$$\bar{A}(x) := \gamma_{0,1}\mathbb{I}(x \geq \sigma_{i,r,1}^{2(1)}) + \sum_{k=1}^{K-1} \gamma_{k,1} H\left(\frac{x - \sigma_{i,r,1}^{2(k)}}{\sigma_{i,r,1}^{2(k+1)} - \sigma_{i,r,1}^{2(k)}}\right) + \gamma_{K,1}\mathbb{I}(x \geq \sigma_{i,r,1}^{2(K)}) \quad (6.57)$$

$$\bar{B}(x) := \gamma_{0,2}\mathbb{I}(x \geq \sigma_{i,r,2}^{2(1)}) + \sum_{k=1}^{K-1} \gamma_{k,2} H\left(\frac{x - \sigma_{i,r,2}^{2(k)}}{\sigma_{i,r,2}^{2(k+1)} - \sigma_{i,r,2}^{2(k)}}\right) + \gamma_{K,2}\mathbb{I}(x \geq \sigma_{i,r,2}^{2(K)}) \quad (6.58)$$

where $K = 2^{q-2}, \gamma_{0,s} = w_{i,r,s}^{(1)}/2, \gamma_{K,s} = w_{i,r,s}^{(K)}/2$, $s = 1, 2$ and $\gamma_{k,s} = (w_{i,r,s}^{(k+1)} + w_{i,r,s}^{(k)})/2$ $k = 1, \ldots K-1$, $s = 1, 2$ with $H(z) := \max\Big(0, \min(z, 1)\Big)$. Now, the quantity $\bar{C}(x) = \bar{A}(x) + \bar{B}(x)$ will have four point masses at $x = \sigma_{i,r,1}^{2(1)}, \sigma_{i,r,2}^{2(1)}, \sigma_{i,r,1}^{2(K)}$ and $\sigma_{i,r,2}^{2(K)}$. Based on the Pitt and Malik [2011] procedure the two point masses at $\sigma_{\min}^{2(1)} := \min(\sigma_{i,r,1}^{2(1)}, \sigma_{i,r,2}^{2(1)})$ and $\sigma_{\max}^{2(K)} := \max(\sigma_{i,r,1}^{2(K)}, \sigma_{i,r,2}^{2(K)})$ 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 $\sigma_{\min}^{2(K)} := \min(\sigma_{i,r,1}^{2(K)}, \sigma_{i,r,2}^{2(K)})$ 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 $\sigma_{\max}^{2(1)} \leq \sigma_{\min}^{2(K)}$ it can be seen that $\bar{C}^{-1}(u)$ is continuous in the interval $[0, 1]$, thus the point masses of $\bar{C}(x)$ at $x = \sigma_{\max}^{2(1)}$ and $\sigma_{\min}^{2(K)}$ are admissible under these circumstances. However, in the event $\sigma_{\max}^{2(1)} > \sigma_{\min}^{2(K)}$, then $\bar{C}^{-1}(u)$ is not continuous in the interval $[0, 1]$, (see Figure 6.3) having a single discontinuity at

$$u = \begin{cases} \sum_{j=1}^{2^{q-2}} w_{i,r,1}^{(j)} \text{ if } \sigma_{i,r,1}^{2(K)} < \sigma_{i,r,2}^{2(1)} \\ \sum_{j=1}^{2^{q-2}} w_{i,r,2}^{(j)} \text{ if } \sigma_{i,r,2}^{2(K)} < \sigma_{i,r,1}^{2(1)} \end{cases} \quad (6.59)$$

leading to the same problem associated with resampling according to a discrete CDF. Thus, a continuous approximation for resampling σ_i^2, conducive for smooth SMC likelihoods, is

$$\bar{\mathbb{P}}(\sigma_i^2 \leq x | Y_{1:i}, R_i = r) := \bar{C}(x) + D(x) \tag{6.60}$$

$$D(x) = \begin{cases} 0, \quad \text{if } \sigma_{\max}^{2(1)} \leq \sigma_{\min}^{2(K)} \\ (\gamma_{K,1} + \gamma_{0,2}) H\big(\frac{x - \sigma_{\min}^{2(K)}}{\sigma_{\max}^{2(1)} - \sigma_{\min}^{2(K)}}\big) \\ \quad - \gamma_{K,1}\mathbb{I}(x \geq \sigma_{i,r,1}^{2(K)}) - \gamma_{0,2}\mathbb{I}(x \geq \sigma_{i,r,2}^{2(1)}), & \text{if } \sigma_{i,r,1}^{2(K)} < \sigma_{i,r,2}^{2(1)} \\ (\gamma_{K,2} + \gamma_{0,1}) H\big(\frac{x - \sigma_{\min}^{2(K)}}{\sigma_{\max}^{2(1)} - \sigma_{\min}^{2(K)}}\big) \\ \quad - \gamma_{K,2}\mathbb{I}(x \geq \sigma_{i,r,2}^{2(K)}) - \gamma_{0,1}\mathbb{I}(x \geq \sigma_{i,r,1}^{2(1)}). & \text{if } \sigma_{i,r,2}^{2(K)} < \sigma_{i,r,1}^{2(1)} \end{cases} \tag{6.61}$$

Resampling K values of σ_i^2 from (6.60) can be done in proportionate to K operations. Furthermore, we avoid the (proportionate to) K^2 overhead computational cost[1] that would have resulted if we had simply kernel smoothed the weights.

[1] While, see Remark 4.1.1.1, *approximately* kernel smoothing through truncation of the densities $\phi()$ can deliver *almost* linear in K computation times, proceeding by means of (6.60) is still slightly more computationally efficient.

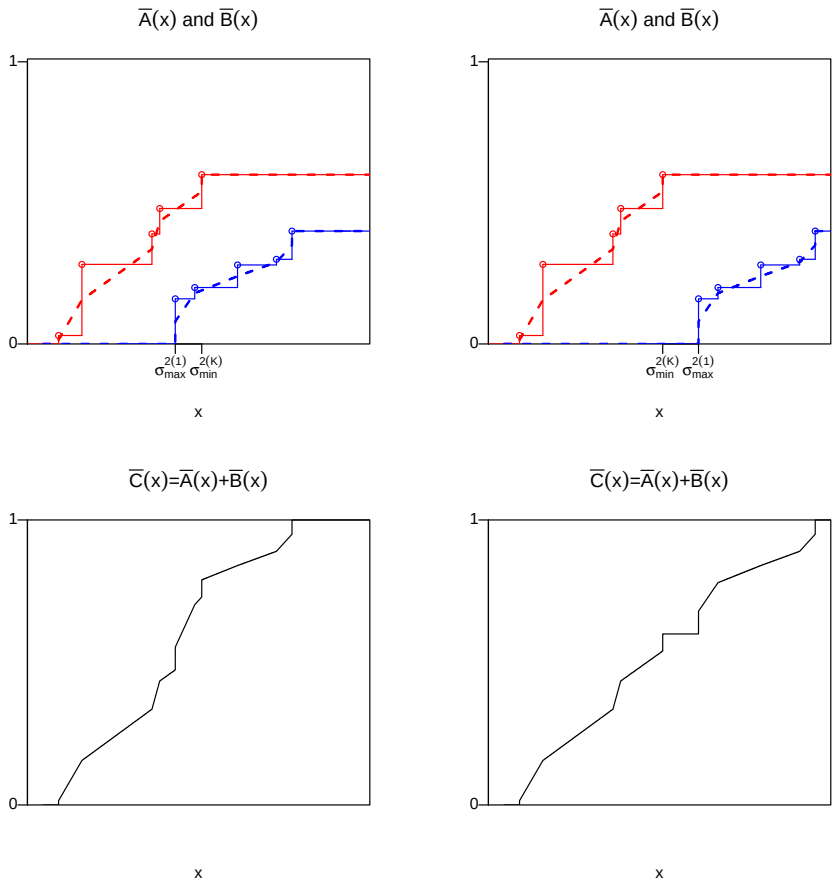


Figure 6.3: For the two top plots, the dashed red line construct is the continuous approximation $\bar{A}(x)$ to the solid red line function $A(x)$ and the dashed blue line construct is the continuous approximation $\bar{B}(x)$ to the solid blue line function $B(x)$. The top left plot corresponds to the case when $\sigma^{2(1)}_{\max} \leq \sigma^{2(K)}_{\min}$ 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 $\sigma^{2(1)}_{\max} > \sigma^{2(K)}_{\min}$ and the bottom right plot is the resultant $\bar{C}(x)$, for which $\bar{C}^{-1}(u)$ can be seen to exhibit a singular discontinuity in the interval $[0, 1]$.

6.5 Extension to the case of missing observations

Now, assume that we do not observe the full MS-GARCH series $Y_{1:N}$, we only have $n < N$ 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 $i = 2, \ldots, n-1$. One can adapt the SMC procedure from the previous section for parameter estimation of the partially observed MS-GARCH series $Y_{t_1:t_n} = \{Y_{t_1}, \ldots, Y_{t_n}\}$.

Figure 6.4 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 fixing randomness across parameter changes). On the other hand, Figure 6.5 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 6.5 over Figure 6.4, to the extent of appearing adequately smooth for use with numerical optimisers. Indeed, as will be seen in the simulation studies of Section 6.6.2, 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 procedure. Start by assuming that at time step i one has at hand the following quantities

$$\tilde{\mathbb{P}}(\sigma^2_{t_i+1}|Y_{t_1:t_i}, R_{t_i+1} = r) = \sum_{k=1}^{2^{q-1}} v^{(k)}_{i,r} \delta_{\{\sigma^{2(k)}_{t_i+1,r}\}}(\sigma^2_{t_i+1}) \text{ and } \hat{\mathbb{P}}(R_{t_i+1}|Y_{t_1:t_i}) \tag{6.62}$$

for which $v^{(k)}_{i,r}, \sigma^{2(k)}_{t_i+1,r}$, $r = 1, 2$, $k = 1, \ldots, 2^{q-1}$ and $\hat{\mathbb{P}}(R_{t_i+1}|Y_{t_1:t_i})$, are determined recursively from the previous time step $i-1$ and their derivation revealed as we proceed to obtain $v^{(k)}_{i+1,r}, \sigma^{2(k)}_{t_{i+1}+1,r}$, $r = 1, 2$, $k = 1, \ldots, 2^{q-1}$ 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 $\sigma^2_{t_i+1}$ from a continuous approximation (details given shortly) to the discrete distribution $\tilde{\mathbb{P}}(\sigma^2_{t_i+1}|Y_{t_1:t_i}, R_{t_i+1} = r)$. Denote these draws

by $\hat{\sigma}^{2(k)}_{t_i+1,r}$, $k = 1, \ldots, 2^{q-2}, r = 1, 2$ and construct

$$\hat{\mathbb{P}}(\sigma^2_{t_i+1}|Y_{t_1:t_i}, R_{t_i+1} = r) = \frac{1}{2^{q-2}} \sum_{k=1}^{2^{q-2}} \delta_{\{\hat{\sigma}^{2(k)}_{t_i+1,r}\}}(\sigma^2_{t_i+1}). \tag{6.63}$$

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 $k = 1, \ldots, 2^{q-2}, r = 1, 2$. 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 $R_{t_i+2}, \ldots, R_{t_{i+1}}$ given $R_{t_i+1} = r$, denote these $R^{(k,r)}_{t_i+2:t_{i+1}} := \{R^{(k,r)}_{t_i+2}, \ldots, R^{(k,r)}_{t_{i+1}}\}$, $k = 1, \ldots, 2^{q-2}$, along with 2^{q-2} sets of $n_{i+1} - 1$ innovations $z^{(k,r)}_{t_i+1:t_{i+1}-1} := \{z^{(k,r)}_{t_i+1}, \ldots, z^{(k,r)}_{t_{i+1}-1}\}$. Given $\hat{\sigma}^{2(k)}_{t_i+1,r}, R^{(k,r)}_{t_i+2:t_{i+1}}, R_{t_i+1} = r$ and $z^{(k,r)}_{t_i+1:t_{i+1}-1}$ obtain $\hat{\sigma}^{2(k)}_{t_{i+1},r}$ recursively from

$$\hat{\sigma}^{2(k)}_{t_i+1+j,r} = h(\sqrt{(z^{(k,r)}_{t_i+j})^2 \hat{\sigma}^{2(k)}_{t_i+j,r}}, \hat{\sigma}^{2(k)}_{t_i+j,r}, R^{(k,r)}_{t_i+j}, R^{(k,r)}_{t_i+1+j}) \tag{6.64}$$

for $j = 1, \ldots, n_i - 1$, where $R^{(k,r)}_{t_i+1} = r$. We then construct

$$\hat{\mathbb{P}}(\sigma^2_{t_{i+1}}, R_{t_{i+1}}|Y_{t_1:t_i}) = \sum_{r=1}^{2} \hat{\mathbb{P}}(R_{t_i+1} = r|Y_{t_1:t_i}) \frac{1}{2^{q-2}} \sum_{k=1}^{2^{q-2}} \delta_{\hat{\sigma}^{2(k)}_{t_{i+1},r}, R^{(k,r)}_{t_{i+1}}}(\sigma^2_{t_{i+1}}, R_{t_{i+1}}) \tag{6.65}$$

and yield the approximate likelihood component

$$\begin{aligned} \hat{p}(Y_{t_{i+1}}|Y_{t_1:t_i}) = \sum_{r=1}^{2} \Bigg(& \hat{\mathbb{P}}(R_{t_i+1} = r|Y_{t_1:t_i}) \frac{1}{2^{q-2}} \\ & \times \sum_{k=1}^{2^{q-2}} p(Y_{t_{i+1}}|\sigma^2_{t_{i+1}} = \hat{\sigma}^{2(k)}_{t_{i+1},r}, R_{t_{i+1}} = R^{(k,r)}_{t_{i+1}}) \Bigg). \end{aligned} \tag{6.66}$$

Then from (6.32) and

$$\mathbb{P}(\sigma^2_{t_{i+1}+1}|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1}) = \int_{\sigma^2_{t_{i+1}}, R_{t_{i+1}}} \left(\mathbb{P}(\sigma^2_{t_{i+1}+1}, R_{t_{i+1}+1}|\sigma^2_{t_{i+1}}, R_{t_{i+1}}, Y_{t_{i+1}}) \times \frac{p(Y_{t_{i+1}}|\sigma^2_{t_{i+1}}, R_{t_{i+1}})\mathbb{P}(\sigma^2_{t_{i+1}}, R_{t_{i+1}}|Y_{t_1:t_i})}{\mathbb{P}(R_{t_{i+1}+1}|Y_{t_1:t_{i+1}})p(Y_{t_{i+1}}|Y_{t_1:t_i})} \right) \tag{6.67}$$

we arrived at

$$\tilde{\mathbb{P}}(\sigma^2_{t_{i+1}+1}|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1} = r) = \sum_{k=1}^{2^{q-1}} v^{(k)}_{i+1,r} \delta_{\{\sigma^{2(k)}_{t_{i+1}+1,r}\}}(\sigma^2_{t_{i+1}+1}) \tag{6.68}$$

whereby

$$v^{(k)}_{i+1,r} = \begin{cases} \dfrac{p(Y_{t_{i+1}}|\sigma^2_{t_{i+1}} = \hat{\sigma}^{2(k)}_{t_{i+1},1}, R_{t_{i+1}} = R^{(k,1)}_{t_{i+1}})}{2^{q-2}\hat{p}(Y_{t_{i+1}}|Y_{t_1:t_i})} \times \dfrac{P_{R^{(k,1)}_{t_{i+1}},r}\hat{\mathbb{P}}(R_{t_i+1} = 1|Y_{t_1:t_i})}{\hat{\mathbb{P}}(R_{t_{i+1}+1} = r|Y_{t_1:t_{i+1}})} & \text{for } k = 1, \ldots, 2^{q-2} \\ \dfrac{p(Y_{t_{i+1}}|\sigma^2_{t_{i+1}} = \hat{\sigma}^{2(k-2^{q-2})}_{t_{i+1},2}, R_{t_{i+1}} = R^{(k-2^{q-2},2)}_{t_{i+1}})}{2^{q-2}\hat{p}(Y_{t_{i+1}}|Y_{t_1:t_i})} \times \dfrac{P_{R^{(k-2^{q-2},2)}_{t_{i+1}},r}\hat{\mathbb{P}}(R_{t_i+1} = 2|Y_{t_1:t_i})}{\hat{\mathbb{P}}(R_{t_{i+1}+1} = r|Y_{t_1:t_{i+1}})} & \text{for } k = 2^{q-2}+1, \ldots, 2^{q-1} \end{cases} \tag{6.69}$$

$$\sigma^{2(k)}_{t_{i+1}+1,r} = \begin{cases} h(Y_{t_{i+1}}, \hat{\sigma}^{2(k)}_{t_{i+1},1}, R^{(k,1)}_{t_{i+1}}, r) & \text{for } k = 1, \ldots, 2^{q-2} \\ h(Y_{t_{i+1}}, \hat{\sigma}^{2(k-2^{q-2})}_{t_{i+1},2}, R^{(k-2^{q-2},2)}_{t_{i+1}}, r) & \text{for } k = 2^{q-2}+1, \ldots, 2^{q-1} \end{cases} \tag{6.70}$$

for which

$$\hat{\mathbb{P}}(R_{t_{i+1}+1} = r|Y_{t_1:t_{i+1}}) = \sum_{s=1}^{2} \left(\frac{\hat{\mathbb{P}}(R_{t_i+1} = s|Y_{t_1:t_i})}{2^{q-2}\hat{p}(Y_{t_{i+1}}|Y_{t_1:t_i})} \times \sum_{k=1}^{2^{q-2}} P_{R^{(k,s)}_{t_{i+1}},r} p(Y_{t_{i+1}}|\sigma^2_{t_{i+1}} = \hat{\sigma}^{2(k)}_{t_{i+1},r}, R_{t_{i+1}} = R^{(k,s)}_{t_{i+1}}) \right). \tag{6.71}$$

The required quantities $\tilde{\mathbb{P}}(\sigma^2_{t_{i+1}+1}|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1})$ 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 $z^{(k,r)}_{t_i+1:t_{i+1}-1}$ which are invariant across parameter changes, the noise components $R^{(k,r)}_{t_i+2:t_{i+1}}$ vary as either $P_{1,2}$ or $P_{2,1}$ change.

The weights $v^{(k)}_{i+1,r}$, see (6.69), in addition to $\sigma^2_{t_{i+1}}$ have components that differ by R_{t_i+1} and $R_{t_{i+1}}$. Note that by design the component R_{t_i+1} is invariant across parameter changes. However, as mentioned, the component $R_{t_{i+1}}$ can vary as either $P_{1,2}$ or $P_{2,1}$ change. Despite this, sufficiently smooth likelihood surfaces for numerical optimisation (as confirmed in the simulation studies of Section 6.6.2) are able to be obtained from approximating a continuous version to the discrete $\tilde{\mathbb{P}}(\sigma^2_{t_{i+1}+1}|Y_{t_1:t_{i+1}}, R_{t_{i+1}+1})$ by partitioning by R_{t_i+1} into two step-wise functions and proceeding in the manner (6.60) was obtained.

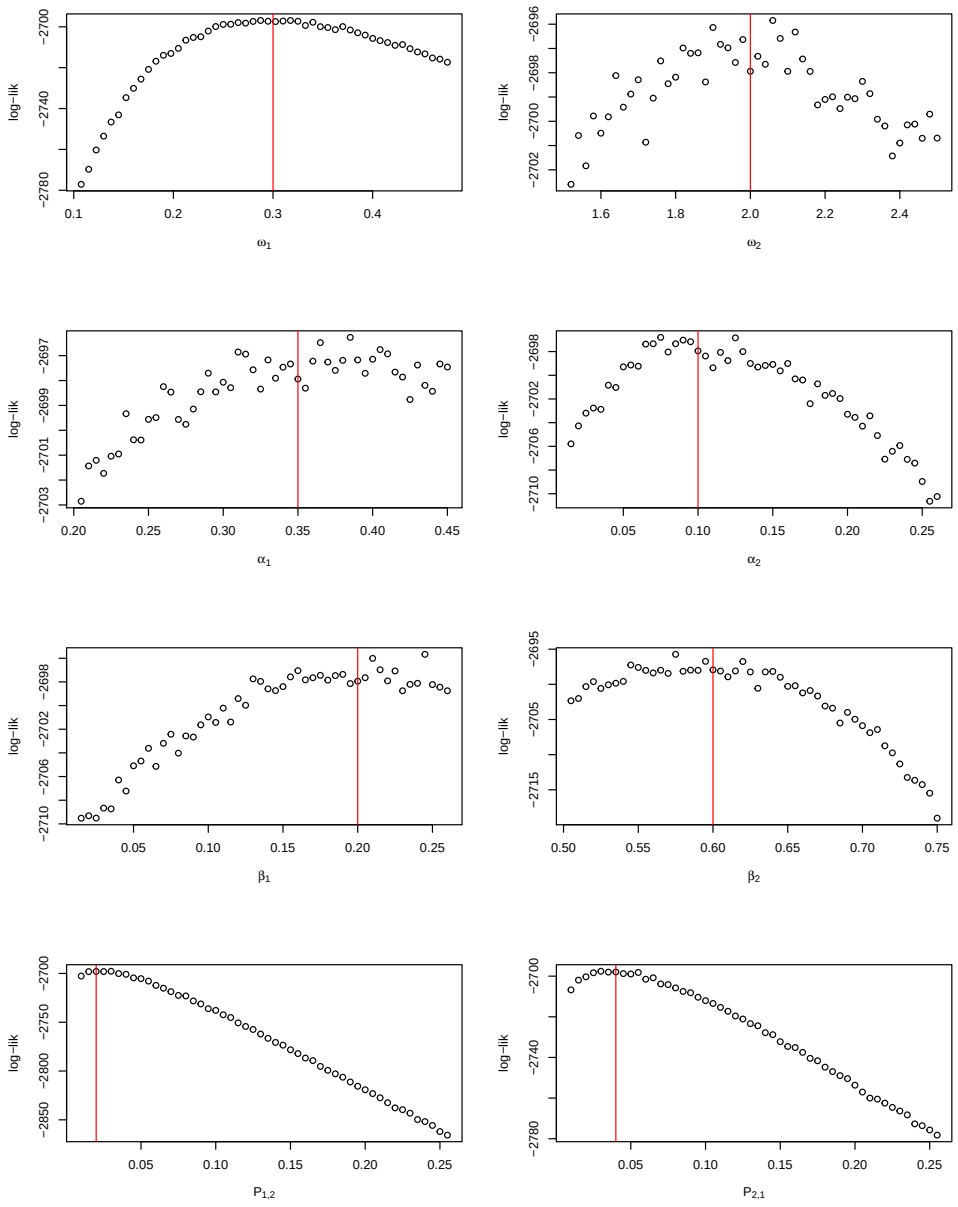


Figure 6.4: SMC approximated likelihood profile plots, using bootstrap resampling, for a partially observed MS-GARCH(1,1) series ($N = 2000$ with 20% of observations missing). The true parameters are indicated by the red vertical lines.

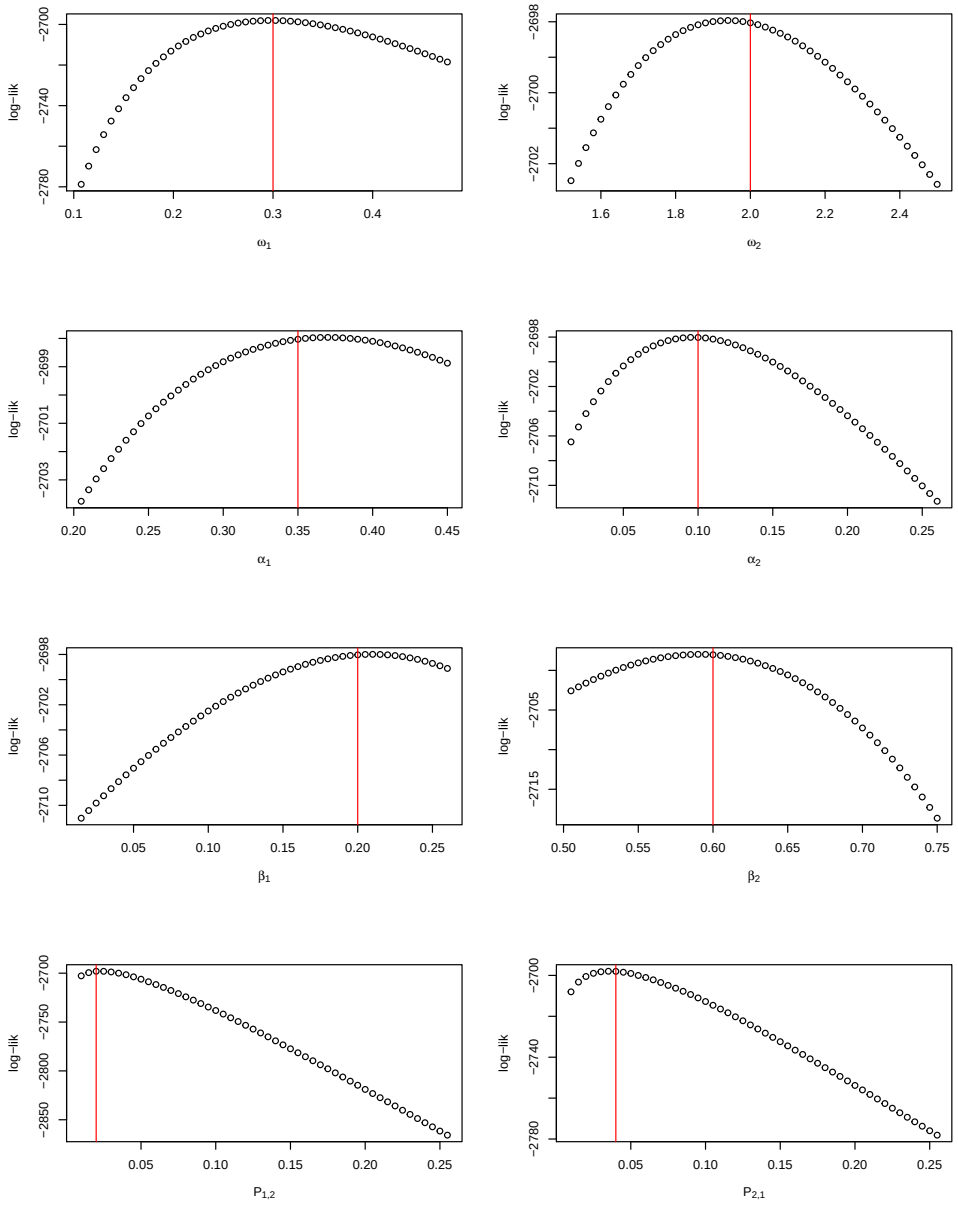


Figure 6.5: SMC approximated likelihood profile plots, using the resampling mechanism presented in this section, for a partially observed MS-GARCH(1,1) series (same data set as Figure 6.4). The true parameters are indicated by the red vertical lines.

6.6 Simulation studies

6.6.1 Full Y observation series

We simulate 1000 data sets of 1500 observations of a 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 $\mu_2 = -0.09$. This parameter set, inspired by empirical studies, was employed by Bauwens et al. [2010] for testing their Bayesian MCMC algorithm. Parameter estimates obtained utilising the SMC procedure and GCP approximation on these 1000 data sets are summarised respectively in Tables 6.1 and 6.2. Refer to Section 3.2.1 for definitions of the summary statistics (**Mean, Bias** etc.) in these tables. Some points about these tables are:

- The SMC and GCP methods are seen to be on par in terms of estimation performance, having very similar RMSE for all parameters. Both methods perform reasonably well, with low percentage biases for most of the parameters.

- Estimation performance of both methods appears stabilised by $q = 8$, with no substantial improvement in either Bias or SD 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 are seen to produce very similar Mean SE for all parameters. For both methods, for all parameters, there is little deviation in Mean SE from utilising $q = 8, 10$ or 12. In general, we see Mean SE to be reasonably close

to the observed SD.

Computational times using an Intel Xeon 3.06 GHz processor for both SMC and GCP optimisations both utilising integrated C++ and R implementations are presented in Table 6.3. Both methods display optimisation times roughly linear in 2^q. For a given q, the SMC optimisation takes on average 40% longer than the GCP.

Table 6.1: Summary of SMC 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 $\mu_2 = -0.09$.

	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	0.0523	0.0990	0.1444	1.1313	0.1251	0.1754	0.0059	0.0119	0.0237	0.1150
SD/ Mean SE	1.0478	0.7263	0.6683	1.0552	0.5028	1.2115	1.0508	1.1176	0.9705	1.0043
$q = 10$										
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	0.0527	0.0996	0.1667	1.1328	0.1253	0.1720	0.0059	0.0118	0.0237	0.1143
SD/ Mean SE	1.0455	0.7199	0.5837	1.0698	0.5044	1.2424	1.0508	1.0847	0.9662	1.0070
$q = 12$										
Mean	0.3000	0.3501	0.1997	2.4010	0.1001	0.5348	0.0205	0.0424	0.0612	-0.0942
Bias	0.0000	0.0001	-0.0003	0.4010	0.0001	-0.0652	0.0005	0.0024	0.0012	-0.0042
SD	0.0546	0.0718	0.0969	1.2280	0.0631	0.2165	0.0060	0.0192	0.0229	0.1162
RMSE	0.0546	0.0718	0.0969	1.2918	0.0631	0.2261	0.0061	0.0193	0.0230	0.1163
Mean SE	0.0523	0.0988	0.1489	1.1405	0.1242	0.1723	0.0059	0.0119	0.0237	0.1151
SD/ Mean SE	1.0440	0.7267	0.6508	1.0767	0.5081	1.2565	1.0169	1.6134	0.9662	1.0096

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

Table 6.2: Summary of 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 $\mu_2 = -0.09$.

	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	0.0525	0.0993	0.1531	1.1201	0.1229	0.1705	0.0059	0.0119	0.0237	0.1153
SD/ Mean SE	1.0419	0.7241	0.6342	1.0657	0.5142	1.2545	1.0169	1.0672	0.9620	0.9983
$q = 10$										
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	0.0526	0.0995	0.1527	1.1220	0.1234	0.1707	0.0059	0.0119	0.0237	0.1154
SD/ Mean SE	1.0399	0.7226	0.6359	1.0631	0.5122	1.2542	1.0169	1.0672	0.9662	0.9974
$q = 12$										
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	0.0526	0.0996	0.1526	1.1237	0.1234	0.1709	0.0059	0.0119	0.0237	0.1155
SD/ Mean SE	1.0380	0.7229	0.6363	1.0635	0.5122	1.2557	1.0169	1.0672	0.9662	0.9965

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

Table 6.3: Summary, across the 1000 simulated data sets, of computational times for the SMC and GCP optimisations utilising an Intel Xeon 3.06 GHz processor. The optimisations were performed with R's implementation of the Broyden - Fletcher - Goldfarb - Shanno (BFGS) algorithm with a relative tolerance set to 10^{-6}.

	SMC			GCP		
q	Avg. time to complete optimisation (secs)	Avg. number of function evaluations	Avg. number of gradient evaluations	Avg. time to complete optimisation (secs)	Avg. number of function evaluations	Avg. number of gradient evaluations
8	21.77	70.53	23.10	15.58	58.39	21.89
10	87.71	66.74	22.67	62.10	58.21	21.84
12	348.82	62.74	22.28	247.37	58.24	21.83

6.6.2 Partial Y observation series

In this section we test the performance of the SMC estimation method at 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 mp as the missing percentage. For each missing scenario we simulate 1000 different data sets of a length $N = \lfloor \frac{1500}{1-mp} \rfloor$ 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 $\lfloor mp \times N \rfloor$ 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. Tables 6.4, 6.5 and 6.6 provide summary statistics for the parameters estimated from 1000 simulated data sets at each missing percentage level using the SMC method with respective values of $q = 8, 10, 12$. A visual summary of Bias, SD and Mean SE

against missing percentage is provided in Figure 6.6. From these:

- Estimation performance in terms of RMSE for the dominant[1] regime's volatility and mean parameters ($\omega_1, \alpha_1, \beta_1$ and μ_1) is seen to slightly worsen as missingness levels increase. Whereas the transition probability parameters as well as the less dominant regime's volatility and mean parameters are seen to either slightly improve or maintain RMSE performance as missingness increases (while holding the overall sample size constant).

- By increasing q, substantial improvements in Bias for most of the parameters is achieved. However little to no improvement is observed in SD by increasing q from 8 to 12. RMSE, being dominated in composition by SD, displays the same behaviour.

- In general, Mean SE appear to be reasonably close to the observed SD. For the GARCH parameters of the second less dominant regime (ω_2, α_2 and β_2) Mean SE is perhaps mildly larger than the observed SD. Figure 6.7 displays the ratio of SD to Mean SE utilising $q = 8, 10$ and 12 at each level of missingness analysed. More noticeably, at the higher levels of missingness, 35% and 50%, it appears the SD to Mean SE ratio moves closer towards unity as q increases from 8 to 12. At the 10% missing level, it may be apparent that while increasing q, the SD to Mean SE ratio remains flat for some parameters, having converged to a number away from 1, this is perhaps not surprising given that one would expect normality to be asymptotic in the number of observations, not just particles/branches.

Computational times using an Intel Xeon 3.06 GHz processor for the SMC optimisations are presented in Table 6.7. The average time to complete a SMC likelhood evaluation can be seen to be linear in N as well as 2^q.

[1] Dominant in the sense $\mathbb{P}(R_i = 1) = 2/3 > 1/3 = \mathbb{P}(R_i = 2)$.

Table 6.4: SMC estimation with $q = 8$ for partially observed MS-GARCH(1,1) series: Summary statistics from 1000 replications for each level of missingness.

	$q = 8$									
	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
10% Missing										
Mean	0.2820	0.3487	0.2312	1.9531	0.0863	0.6254	0.0210	0.0446	0.0613	-0.0793
Bias	-0.0180	-0.0013	0.0312	-0.0469	-0.0137	0.0254	0.0010	0.0046	0.0013	0.0107
SD	0.0612	0.0832	0.1133	0.4934	0.0598	0.1052	0.0065	0.0148	0.0232	0.1222
Mean SE	0.0513	0.1029	0.1035	0.8747	0.0912	0.1223	0.0048	0.0106	0.0237	0.1026
RMSE	0.0638	0.0832	0.1175	0.4956	0.0614	0.1082	0.0066	0.0155	0.0232	0.1227
20 % Missing										
Mean	0.2716	0.3433	0.2527	1.9348	0.0845	0.6320	0.0215	0.0459	0.0611	-0.0759
Bias	-0.0284	-0.0067	0.0527	-0.0652	-0.0155	0.0320	0.0015	0.0059	0.0011	0.0141
SD	0.0650	0.0935	0.1276	0.3651	0.0602	0.0853	0.0066	0.0150	0.0241	0.1231
Mean SE	0.0491	0.1011	0.1381	0.8517	0.0866	0.1161	0.0043	0.0100	0.0237	0.0964
RMSE	0.0709	0.0937	0.1381	0.3709	0.0621	0.0911	0.0067	0.0162	0.0242	0.1239
35% Missing										
Mean	0.2516	0.3268	0.2966	1.9273	0.0873	0.6368	0.0221	0.0487	0.0617	-0.0771
Bias	-0.0484	-0.0232	0.0966	-0.0727	-0.0127	0.0368	0.0021	0.0087	0.0017	0.0129
SD	0.0679	0.0980	0.1390	0.3606	0.0666	0.0859	0.0071	0.0186	0.0255	0.1286
Mean SE	0.0492	0.1041	0.0821	0.9017	0.0876	0.1226	0.0039	0.0098	0.0243	0.1009
RMSE	0.0834	0.1007	0.1693	0.3679	0.0678	0.0935	0.0074	0.0206	0.0255	0.1293
50% Missing										
Mean	0.2372	0.3027	0.3344	1.9291	0.0877	0.6367	0.0233	0.0513	0.0617	-0.0682
Bias	-0.0628	-0.0473	0.1344	-0.0709	-0.0123	0.0367	0.0033	0.0113	0.0017	0.0218
SD	0.0703	0.1040	0.1475	0.3383	0.0667	0.0793	0.0076	0.0199	0.0264	0.1299
Mean SE	0.0525	0.1104	0.0865	0.9525	0.0900	0.1301	0.0038	0.0092	0.0253	0.1002
RMSE	0.0943	0.1143	0.1996	0.3457	0.0678	0.0873	0.0083	0.0228	0.0264	0.1317

Table 6.5: SMC estimation with $q = 10$ for partially observed MS-GARCH(1,1) series: Summary statistics from 1000 replications for each level of missingness.

	$q = 10$									
	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
10% Missing										
Mean	0.2867	0.3537	0.2218	1.9956	0.0889	0.6152	0.0204	0.0439	0.0610	-0.0822
Bias	-0.0133	0.0037	0.0218	-0.0044	-0.0111	0.0152	0.0004	0.0039	0.0010	0.0078
SD	0.0615	0.0891	0.1103	0.6005	0.0625	0.1217	0.0065	0.0159	0.0239	0.1273
Mean SE	0.0512	0.1011	0.1176	0.9766	0.1002	0.1416	0.0051	0.0111	0.0236	0.1070
RMSE	0.0629	0.0892	0.1124	0.6005	0.0635	0.1226	0.0066	0.0164	0.0240	0.1275
20 % Missing										
Mean	0.2827	0.3513	0.2293	1.9404	0.0878	0.6257	0.0208	0.0445	0.0608	-0.0827
Bias	-0.0173	0.0013	0.0293	-0.0596	-0.0122	0.0257	0.0008	0.0045	0.0008	0.0073
SD	0.0650	0.0935	0.1251	0.4803	0.0624	0.1033	0.0065	0.0153	0.0247	0.1298
Mean SE	0.0516	0.1046	0.1527	0.9744	0.0992	0.1352	0.0049	0.0108	0.0238	0.1070
RMSE	0.0672	0.0935	0.1285	0.4840	0.0636	0.1064	0.0066	0.0159	0.0247	0.1300
35% Missing										
Mean	0.2711	0.3445	0.2539	1.9753	0.0870	0.6254	0.0210	0.0456	0.0608	-0.0837
Bias	-0.0289	-0.0055	0.0539	-0.0247	-0.0130	0.0254	0.0010	0.0056	0.0008	0.0063
SD	0.0694	0.0990	0.1332	0.4444	0.0657	0.0953	0.0066	0.0181	0.0257	0.1308
Mean SE	0.0527	0.1085	0.1135	0.9932	0.1005	0.1334	0.0046	0.0101	0.0243	0.1056
RMSE	0.0752	0.0991	0.1437	0.4450	0.0669	0.0986	0.0066	0.0190	0.0257	0.1310
50% Missing										
Mean	0.2572	0.3333	0.2850	1.9616	0.0815	0.6319	0.0213	0.0469	0.0605	-0.0798
Bias	-0.0428	-0.0167	0.0850	-0.0384	-0.0185	0.0319	0.0013	0.0069	0.0005	0.0102
SD	0.0734	0.1120	0.1458	0.4095	0.0699	0.0957	0.0066	0.0159	0.0266	0.1320
Mean SE	0.0546	0.1200	0.1015	1.0777	0.1045	0.1464	0.0043	0.0097	0.0249	0.1096
RMSE	0.0850	0.1132	0.1688	0.4113	0.0723	0.1008	0.0067	0.0174	0.0266	0.1324

Table 6.6: SMC estimation with $q = 12$ for partially observed MS-GARCH(1,1) series: Summary statistics from 1000 replications for each level of missingness.

	$q = 12$									
	ω_1	α_1	β_1	ω_2	α_2	β_2	$P_{1,2}$	$P_{2,1}$	μ_1	μ_2
10% Missing										
Mean	0.2885	0.3571	0.2171	2.0143	0.0915	0.6092	0.0202	0.0433	0.0607	-0.0867
Bias	-0.0115	0.0071	0.0171	0.0143	-0.0085	0.0092	0.0002	0.0033	0.0007	0.0033
SD	0.0626	0.0980	0.1111	0.7010	0.0682	0.1387	0.0067	0.0162	0.0250	0.1320
Mean SE	0.0519	0.1054	0.1302	1.0077	0.1138	0.1521	0.0054	0.0114	0.0235	0.1124
RMSE	0.0637	0.0983	0.1124	0.7011	0.0687	0.1390	0.0067	0.0165	0.0250	0.1320
20 % Missing										
Mean	0.2881	0.3574	0.2167	1.9958	0.0906	0.6124	0.0203	0.0433	0.0604	-0.0886
Bias	-0.0119	0.0074	0.0167	-0.0042	-0.0094	0.0124	0.0003	0.0033	0.0004	0.0014
SD	0.0671	0.0988	0.1229	0.6549	0.0667	0.1299	0.0066	0.0151	0.0261	0.1308
Mean SE	0.0522	0.1049	0.1893	1.0204	0.1117	0.1620	0.0052	0.0110	0.0238	0.1128
RMSE	0.0681	0.0990	0.1241	0.6549	0.0673	0.1305	0.0066	0.0154	0.0261	0.1308
35% Missing										
Mean	0.2799	0.3536	0.2336	2.0240	0.0884	0.6128	0.0204	0.0441	0.0604	-0.0925
Bias	-0.0201	0.0036	0.0336	0.0240	-0.0116	0.0128	0.0004	0.0041	0.0004	-0.0025
SD	0.0700	0.1044	0.1311	0.6461	0.0680	0.1289	0.0064	0.0143	0.0267	0.1353
Mean SE	0.0548	0.1134	0.1449	1.0862	0.1135	0.1539	0.0049	0.0104	0.0242	0.1140
RMSE	0.0729	0.1044	0.1353	0.6466	0.0689	0.1296	0.0064	0.0149	0.0267	0.1354
50% Missing										
Mean	0.2723	0.3478	0.2500	1.9881	0.0829	0.6222	0.0206	0.0445	0.0600	-0.0882
Bias	-0.0277	-0.0022	0.0500	-0.0119	-0.0171	0.0222	0.0006	0.0045	0.0000	0.0018
SD	0.0762	0.1167	0.1453	0.5589	0.0710	0.1193	0.0063	0.0139	0.0277	0.1358
Mean SE	0.0577	0.1244	0.1494	1.2182	0.1190	0.1756	0.0047	0.0100	0.0248	0.1162
RMSE	0.0811	0.1167	0.1537	0.5591	0.0731	0.1214	0.0063	0.0147	0.0277	0.1358

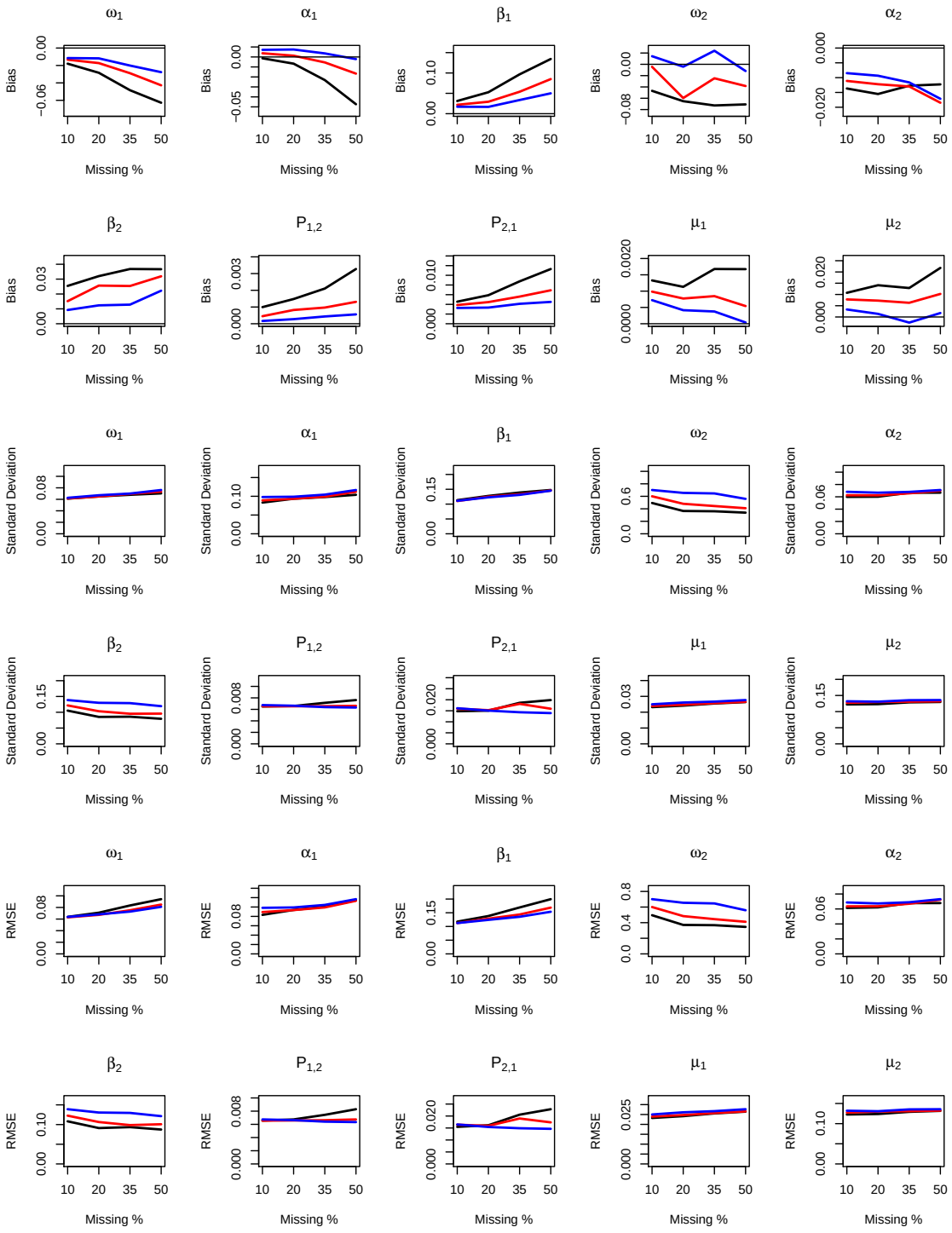


Figure 6.6: Visual summary from the simulation study of Bias, SD and RMSE at different levels of missingness utilising the SMC method with $q = 8$ (black), $q = 10$ (red) and $q = 12$ (blue).

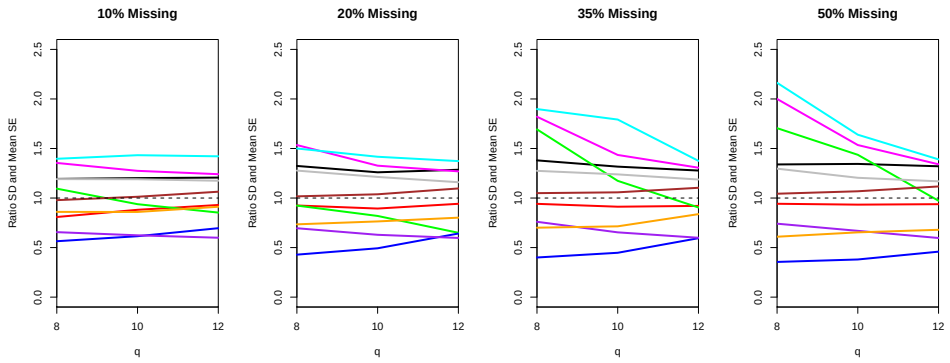


Figure 6.7: Ratio of SD to Mean SE, for the parameters ω_1 (black), α_1 (red), β_1 (green), ω_2 (blue), α_2 (purple), β_2 (orange), $P_{1,2}$ (magenta), $P_{2,1}$ (cyan), μ_1 (brown) and μ_2 (grey), utilising $q = 8, 10$ and 12 at each of the four levels of missingness analysed.

Table 6.7: Summary, across the 1000 simulated data sets at each missing percentage level, of computational times for the SMC optimisations utilising an Intel Xeon 3.06 GHz processor. The optimisations were performed with R's *constrOptim* function with a relative tolerance set to 10^{-6}.

		$q = 8$			$q = 10$		
Missing %	N	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)
10%	1667	59.93	1770.25	0.034	232.41	1667.86	0.139
20%	1875	68.89	1775.52	0.039	279.53	1748.96	0.160
35%	2308	88.51	1843.02	0.048	357.63	1811.98	0.197
50%	3000	110.47	1783.93	0.062	467.4	1821.25	0.257

		$q = 12$		
Missing %	N	Time (secs) to complete optimisation	Avg. Number of function calls	Avg. time to evaluate function (secs)
10%	1667	946.68	1629.72	0.581
20%	1875	1137.69	1698.47	0.670
35%	2308	1474.01	1772.13	0.832
50%	3000	1964.49	1849.78	1.062

6.7 Real data analysis

6.7.1 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 6.8. Returns were calculated on successive trading days, thus no returns are considered missing. A zero mean, standard normal innovation, two regime MS-GARCH(1,1) model was fit to this series multiplied by 100 using both the SMC and GCP methods. The fits, along with estimated 95% confidence intervals, for both methods are provided in Table 6.8. 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 6.8 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. Regime 2 remains prevalent until around perhaps October 2010; while indeed between May 2009 to September 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" began in April 1997 and ended in June 2003[1] this 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[2].

Table 6.8: The fits, along with estimated 95% confidence intervals, for the S&P 500 index returns data, using both the SMC and GCP methods. To clarify, in reference to regime 1 for instance, by **regime exit probability** this refers to the parameter $P_{1,2}$.

	Regime	ω	α	β	regime exit probability
SMC	1	0.0123 (0.0053, 0.0194)	0.0190 (-0.0020, 0.0399)	0.9541 (0.9400, 0.9681)	0.0015 (-0.0002, 0.0031)
	2	0.0538 (0.0286, 0.0791)	0.0941 (0.0351, 0.1531)	0.8846 (0.8730, 0.8961)	0.0011 (-0.0001, 0.0023)
GCP	1	0.0126 (0.0031, 0.0220)	0.0195 (-0.0117, 0.0508)	0.9533 (0.9351, 0.9715)	0.0014 (-0.0008, 0.0036)
	2	0.0531 (0.0256, 0.0805)	0.0938 (0.0287, 0.1589)	0.8845 (0.8715, 0.8975)	0.0011 (-0.0006, 0.0027)

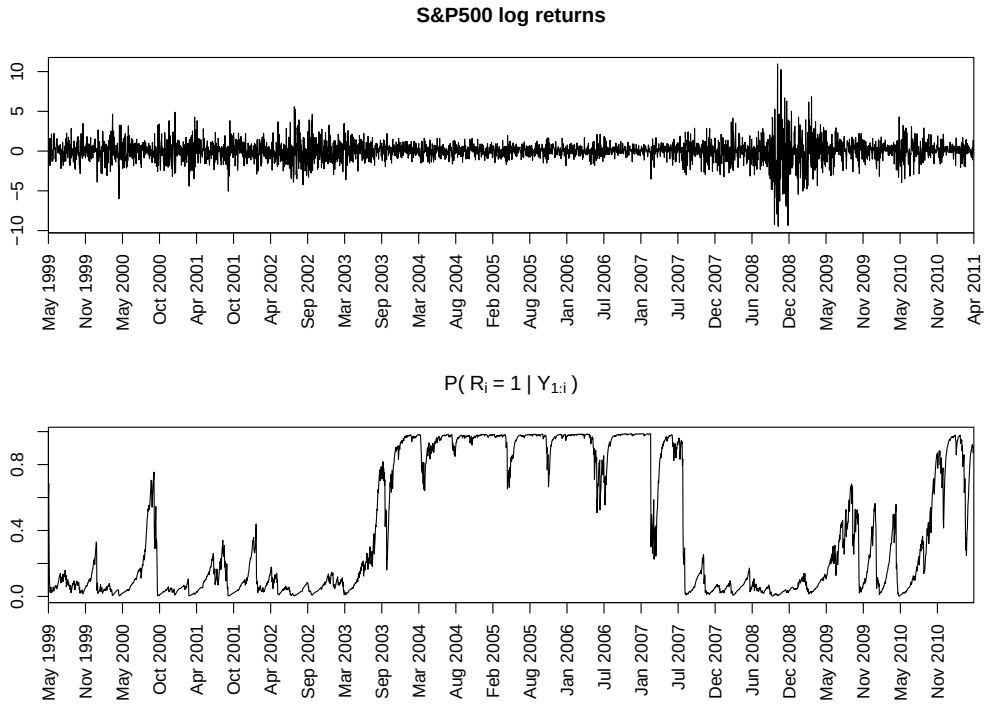


Figure 6.8: (Top:) Daily percentage log returns on the S&P 500 price index from May 20, 1999 to April 25, 2011. (Bottom:) SMC estimated probability of being in regime 1 given the current and past observations.

[1] https://www.businessinsider.com.au/heres-why-the-dot-com-bubble-began-and-why-it-popped-2010-12?r=US&IR=T

[2] https://en.wikipedia.org/wiki/Subprime_mortgage_crisis

6.7.2 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. Returns calculated from successive NYMEX trading days are displayed as the top plot in Figure 6.9. Due to hurricane Rita, Henry Hub was forced to shut down from September 23, 2005 to October 6, 2005 resulting in a 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 6.5. The fits obtained, along with estimated 95% confidence intervals, are provided in Table 6.9. Regime 1 has a stationary variance of 4.74 and stationary probability of 62.9% while regime 2 has a stationary variance of 499.1 and stationary probability of 37.1%. Regime 2 is the much more volatile regime. Both regimes are estimated to have statistically significant persistence parameters (α and β), although regime 1's α is very close to zero. The bottom plot of Figure 6.9 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 red line segment in that plot. This plot appears to indicate 10 regime switches during the study period. The natural gas market tends to have two annual phases, an injection season (November to March) and a withdrawal season (April to October)[1]. 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 avail-

[1] https://seekingalpha.com/article/4060690-logic-injection-season-need-know

able, 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 creates 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 sill 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*" -Alterman [2012]. These cycles of lower and higher periods of volatility appear to be identified in the conditional regime occurrence probabilities in Figure 6.9. 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. The relatively earlier than usual onset of the higher volatility regime in April 2006 potentially might have been induced by large speculative actions of the hedge fund Amaranth[1].

Table 6.9: 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	0.3840 (0.3381,0.4300)	0.0014 (0.0009,0.0018)	0.9176 (0.9117,0.9236)	0.0130 (0.0112,0.0147)
2	0.5810 (0.4709,0.6912)	0.0499 (0.0358,0.0641)	0.9489 (0.9484,0.9493)	0.0220 (0.0185,0.0255)

[1] http://uk.reuters.com/article/amaranth-senate-report-idUKN2425785820070625

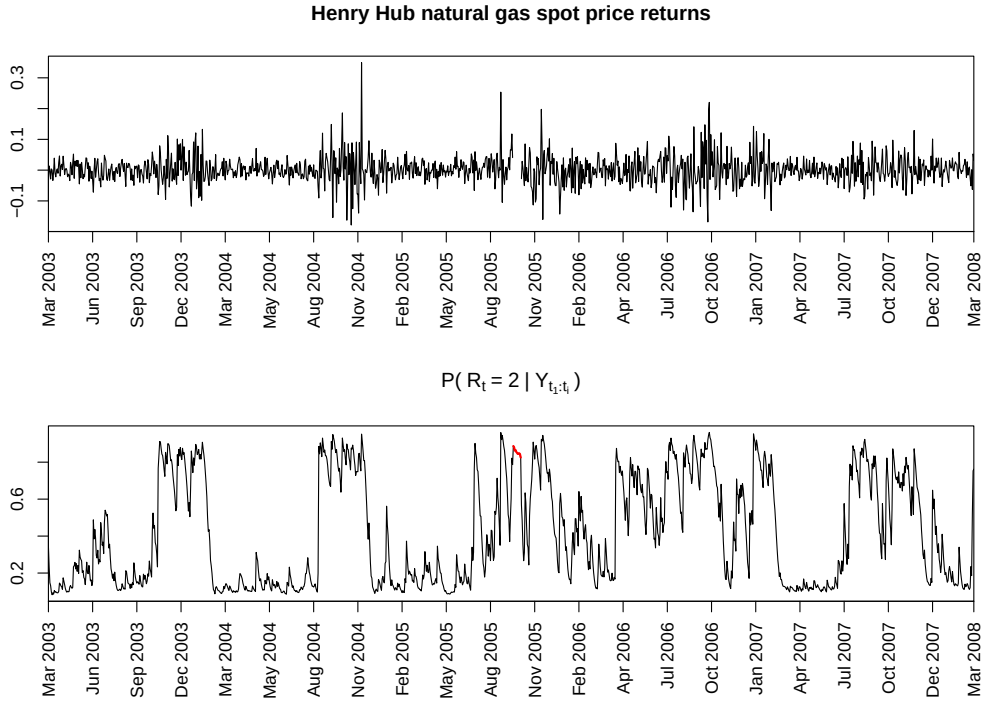


Figure 6.9: (Top:) Daily NYMEX trading day Henry Hub natural gas spot price returns from March 17, 2003 to March 24, 2008. Due to hurricane Rita, Henry Hub was forced to shut down from September 23, 2005 to October 6, 2005 resulting in a 11 day stretch of unavailable NYMEX trading day natural gas spot 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 red line segment.

6.8 Concluding remarks

In summary, this chapter has introduced a modified SMC approach, providing a computationally feasible and reliable frequentist method for obtaining parameter estimates (along with standard errors) for MS-GARCH(1,1) models, even when one does not observe the full MS-GARCH(1,1) series.

When provided full observation of the MS-GARCH(1,1) series, the GCP method of Augustyniak et al. [2017] is able to be utilised. However, unlike SMC, the GCP procedure is unequipped for dealing with partially observed MS-GARCH(1,1) data sets. Through simulation studies, the estimation performance (for fully observed MS-GARCH(1,1) series) of the SMC and GCP methods was seen to be on par, with the GCP method having a modest speed advantage.

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.

The technique presented in this chapter could be combined with that of Section 4.1 for parameter estimation of aggregated MS-GARCH(1,1) series, however this has not been investigated.

Chapter 7

Conclusions and future work

In this thesis we have utilised and further developed SMC with the smooth resampling procedure of Pitt and Malik [2011] (smooth bootstrap) as a statistically and computationally efficient method for providing likelihood inference in many applications for which exact likelihood evaluation is infeasible. While smooth bootstrap is already a well established procedure, for the stochastic volatility models considered in this thesis careful development of suitable state space representations as well as modifications to the smooth bootstrap procedure was required.

These developments have substantially expanded the range of practical applications and scope of smooth bootstrap SMC and allowed accurate likelihood based inference to be used for statistical models that, hitherto, had been estimated using biased and inefficient *adhoc* methods.

If one is able to identify a state space representation of the form (1.8)-(1.9) that has

i) a noise component W_i that is easy to simulate,

ii) a combination of Y_i and X_i for which $p(Y_i|X_i)$ is easy to evaluate,

iii) a state X_i that has continuous support on some interval of the real line,

then implementation of smooth bootstrap is fairly straightforward, as we saw

in Chapter 3 utilising this method for likelihood inference of partially observed GARCH(1,1) series.

For the case of temporally aggregated GARCH(1,1) series, the more familiar state space representation while satisfying i and iii did not satisfy ii. We identified in Chapter 4 an alternative 3-dimensional state specification for which $p(Y_i|X_i)$ was analytical. Despite this specification not satisfying iii, we deduced that this 3-dimensional state specification could be in fact still propagated using only a 1-dimensional quantity $X_{i,1}$. Thus with some slight adjustments we were still eventually able to proceed by applying the smooth bootstrap to resample at each time step $X_{i,1}$ instead of X_i.

The same principled approach used for temporally aggregated GARCH(1,1) series was extended to provide a method for likelihood inference of the COGARCH(1,1) model in Chapter 5.

We showed in Chapter 6, it was possible to modify the standard smooth bootstrap to provide a means for parameter estimation for MS-GARCH(1,1) models in which one needs two variables, volatility (which has a continuous support) and regime (which has a finite support), to evolve the hidden state process.

While the developments of this thesis have considerably expanded the range of models for stochastic volatility to which likelihood inference can now be applied, extensions to models with higher order lags in the autocorrelation and the GARCH recursions are required for more complex applications.

Development of a method for resampling in more than one continuous dimension conducive to smooth SMC likelihoods would significantly increase the scope for which the smooth bootstrap could be utilised. A step towards this is the work of Lee [2008] who utilises tree based resampling schemes, that although do not generate a continuous estimate of the likelihood function, has been shown able to provide a much "smoother" estimate. An alternative prototype approach based on copula is proposed in Appendix A.3. This method however, is still in development

and to be further investigated. Proof of concept is demonstrated for parameter estimation of ARMA(1,1)-GARCH(1,1) models in the presence of missing data which is an example of when resampling in two continuous dimensions is required. Making use of higher dimensional copulas this method naturally extends to higher dimensional state processes necessary for applications requiring for instance higher order lags in their model specification.

Throughout, our SMC procedures have utilised the system transition distribution $p(X_i|X_{i-1}, Y_{i-1})$ to simulate the hidden state process, "*this is the most widely used distribution, since it is simple to compute, but it can be inefficient, since it ignores the most recent evidence,* Y_i" -van der Merwe et al. [2000]. That being said, additional improvements in performance could be sought exploring a change of measure, simulating the state from something other than the system dynamics (cf. Doucet et al. [2000]).

While a fixed number of particles have been employed throughout each iteration of our SMC implementations, this is not essential. Indeed, it would be worthwhile future research to analyse the relationship between the number of consecutive missing (or aggregated) observations and particle degeneracy measures such as *effective sample size* (ESS; cf. Liu [2001]). The ESS of a set of particles has the interpretation that inference using this set of particles "*is approximately equivalent (in terms of estimator variance) to inference based on ESS perfect samples from the target distribution*" -Doucet and Johansen [2011]. Thus if it is found that iterations that involve a large block of consecutively missing (or aggregated) underlying series elements are synonymous with low ESS values, then it may be beneficial to utilise more particles at iterations where the "length" of the missing (or aggregated) block is large relative to when it is small, perhaps choosing the number of particles employed by some non-decreasing function of the block length.

Resampling was conducted at each iteration of our SMC procedures. There are schemes such as *adaptive resampling* (cf. Del Moral et al. [2012]) that only trigger

a resampling operation when some measure, commonly ESS, drops below some threshold. Note however that ESS is a function of the particle weights and thus the model parameters. Therefore resampling schedules that are dictated by ESS could be drastically different across model parameters imposing discontinuities across the approximated likelihood surface.

Other simulation based frequentist approaches that have been applied to hidden Markov models include Approximate Bayesian Computation Maximum Likelihood (ABC-ML; cf. Dean et al. [2014]) and online Expectation Maximisation (EM; cf. Cappe [2009], Cappe [2011]). Potentially some of the ideas and techniques developed in this thesis could be applied in these contexts, however the feasibility of implementing these approaches for the models we have considered in this thesis remains to be investigated. In regards to the online EM method, ease of implementation generally requires that the model structure possess sufficient statistics that are of an additive form as well as an inverse mapping of these sufficient statistics that can be used to recover the model parameters - these aspects would have to be explored for the models we have considered.

Appendix A

A.1 COGARCH SMC pseudo code

Algorithm 1 SMC log-likelihood approximation using bootstrap resampling

procedure
Initilisation:
  $X_{0,3}^{1:N} \leftarrow \frac{\beta}{\eta-\varphi}$
  $\text{loglik} \leftarrow Y_0 \leftarrow t_0 \leftarrow X_{0,2}^{1:N} \leftarrow X_{0,4}^{1:N} \leftarrow \kappa \leftarrow 0$
  Set Seed to fix randomness
Loop:
  for i in $1:n$ **do**
    if $Y_i \neq 0$ **then**
      $\text{loglik} \leftarrow \text{loglik} + \log(1 - e^{-\lambda \Delta t_i})$
      for j in $1:N$ **do**
        if $\kappa > 0$ **then**
          Sample ν from the set $\{1, \dots N\}$ *according to the probabilities* $\{w_\kappa^1, \dots w_\kappa^N\}$
        else
          $\nu \leftarrow j$
        $X_{i,1}^j \leftarrow \frac{\beta}{\eta} + (X_{\kappa,3}^\nu + \varphi(Y_\kappa - X_{\kappa,2}^\nu)^2 - \frac{\beta}{\eta}) e^{-\eta(t_{i-1} - X_{\kappa,4}^\nu)}$
        Simulate $U_1 \sim \text{Uniform}(0,1)$
        $\tau_{i,1}^j \leftarrow t_{i-1} - \frac{1}{\lambda} \log(1 - (1 - e^{-\lambda \Delta t_i}) U_1)$
        $X_{i,2}^j \leftarrow 0$
        $X_{i,3}^j \leftarrow \frac{\beta}{\eta} + (X_{i,1}^j - \frac{\beta}{\eta}) e^{-\eta(\tau_{i,1}^j - t_{i-1})}$
        Simulate $r \sim \text{Poisson}(\lambda(t_i - \tau_{i,1}^j))$
        if $r > 0$ **then**
          Simulate $z_1, \dots, z_r \sim$ *i.i.d random variables with density* q
          Simulate $U_2 < \dots < U_{r+1}$ as *i.i.d standard uniforms sorted in ascending order*
          for s in $1:r$ **do**
            $\tau_{i,s+1}^j \leftarrow \tau_{i,1}^j + U_{s+1}(t_i - \tau_{i,1}^j)$
            $X_{i,2}^j \leftarrow X_{i,2}^j + \sqrt{X_{i,3}^j} z_s$
            $X_{i,3}^j \leftarrow \frac{\beta}{\eta} + (X_{i,3}^j (1 + \varphi z_s^2) - \frac{\beta}{\eta}) e^{-\eta(\tau_{i,s+1}^j - \tau_{i,s}^j)}$
        $X_{i,4}^j \leftarrow \tau_{i,r+1}^j$
        $w_i^j \leftarrow \frac{1}{\sqrt{X_{i,3}^j}} q\Big(\frac{Y_i - X_{i,2}^j}{\sqrt{X_{i,3}^j}}\Big)$
      $\text{loglik} \leftarrow \text{loglik} + \log(\frac{1}{N} \sum_{j=1}^N w_i^j)$
      Normalise $w_i^j \leftarrow \frac{w_i^j}{\sum_{j=1}^N w_i^j}$
      $\kappa \leftarrow i$
    else
      $\text{loglik} \leftarrow \text{loglik} - \lambda \Delta t_i$

Algorithm 2 SMC log-likelihood approximation using continuous resampling

Lines 1-9 from Algorithm 1

for j in $1:N$ **do**
  if $\kappa > 0$ **then**
    $\gamma_1 \leftarrow \gamma_2 \leftarrow 0$
    for p in $1:N$ **do**
      if $j = 1$ **then**
        $X^p_{i,1} \leftarrow \frac{\beta}{\eta} + (X^p_{\kappa,3} + \varphi(Y_\kappa - X^p_{\kappa,2})^2 - \frac{\beta}{\eta})e^{-\eta(t_{i-1} - X^p_{\kappa,4})}$
      $\gamma_1 \leftarrow \gamma_1 + w^p_\kappa \phi\Big((X^j_{i,1} - X^p_{i,1})/h\Big)$
      $\gamma_2 \leftarrow \gamma_2 + \phi\Big((X^j_{i,1} - X^p_{i,1})/h\Big)$
    $\alpha^j \leftarrow \gamma_1/\gamma_2$
  else
    $X^j_{i,1} \leftarrow \frac{\beta}{\eta} + (X^j_{\kappa,3} + \varphi(Y_\kappa - X^j_{\kappa,2})^2 - \frac{\beta}{\eta})e^{-\eta(t_{i-1} - X^j_{\kappa,4})}$
if $\kappa > 0$ **then**
  Normalise $\alpha^j \leftarrow \alpha^j / \sum_{j=1}^N \alpha^j$
  Sort $\{X^j_{i,1}, \alpha^j\}_{j=1\ldots N}$ *ascending in* $X^j_{i,1}$.
  $\pi^0 \leftarrow \frac{1}{2}\alpha^1$
  $\pi^N \leftarrow \frac{1}{2}\alpha^N$
  Simulate $\tilde{U}_1 < \cdots < \tilde{U}_N$ as *i.i.d standard uniforms sorted in ascending order*
  $s \leftarrow 1$
  $t \leftarrow \pi^0$
  for j in $1:N$ **do**
    if $\tilde{U}^j < \pi^0$ **then**
      $\tilde{X}^j_{i,1} = X^1_{i,1}$
    else if $\tilde{U}^j > 1 - \pi^N$ **then**
      $\tilde{X}^j_{i,1} = X^N_{i,1}$
    else
      while $\tilde{U}^j > t$ **do**
        $\pi^s = \frac{1}{2}(\alpha^s + \alpha^{s+1})$
        $t \leftarrow t + \pi^s$
        $s \leftarrow s + 1$
      $\tilde{X}^j_{i,1} = X^{s-1}_{i,1} + \frac{\tilde{U}^j - (t - \pi^{s-1})}{\pi^{s-1}}(X^s_{i,1} - X^{s-1}_{i,1})$
  Set $X^{1:N}_{i,1} = \tilde{X}^{1:N}_{i,1}$
for j in $1:N$ **do**

Lines 16-34 from Algorithm 1

A.2 Data pre-processing for Section 5.5 data set

Let us denote

- t - as the number of minutes elapsed since the open of the trading week (Sunday 22:00 GMT),
- $S_i(t)$ the AUD/USD price on week i at time t,
- O_i the number of observations available on week i.

We then define the log returns

$$r^{(i)}(t_j) = \log\left(\frac{S_i(t_j)}{S_i(t_{j-1})}\right) \text{ for } j = 2, \ldots, O_i. \tag{A.1}$$

A.2.1 Winsorised sample

In our preliminary analysis, we found that the presence of outliers tended to inflate estimates of the decay parameter η. Intuitively, the reason behind this, lies at the essence of the GARCH type relationship in that perturbations in the current period flow on to affect future volatility. Thus, if a one off significant spike is observed without a subsequent increase in volatility for the next few observations - to make a COGARCH model conform to such a situation would be to have a COGARCH model that decays off that large spike quickly so that the volatility is back down again for subsequent observations.

A simple technique to reduce the effect of possibly spurious outliers is to work with a Winsorised sample. That is, for each trading week any time standardised log returns that are below the x-th percentile for that week are set to the value of the x-th percentile and any time standardised log returns that are above the $(100 - x)$-th percentile for that week are set to the value of the $(100 - x)$-th

percentile. Formally, the Winsorised log returns for our analysis are obtained as

$$r_w^{(i)}(t_j) = \max\left(\min\left(\frac{r^{(i)}(t_j)}{\Delta t_j}, r_{100-x}^{(i)}\right), r_x^{(i)}\right) \times \Delta t_j \tag{A.2}$$

where $r_x^{(i)}$ and $r_{100-x}^{(i)}$ are respectively the x-th and $(100-x)$-th percentiles from the sample $\frac{r^{(i)}(t_j)}{\Delta t_j}$ $j = 2, \ldots, O_i$. The time standardised log returns as opposed to simply the log returns is Winsorised to factor in that naturally a larger observation interval allows a larger magnitude of price change to manifest compared to a shorter observation interval. Our analysis is performed with $x = 0.01$, in this way approximately 99.8% of the original returns are preserved. Hereafter, by log returns we will implicitly mean the Winsorised log returns.

A.2.2 Virtual time scale

To account for intra-week volatility patterns a transformation from the physical time scale to the virtual time scale of Dacorogna et al. [1993] will be applied. The adjustment begins with the empirical scaling law of Muller et al. [1990], which relates $\overline{\mid r_{\Delta t} \mid}$ the mean absolute log return over a time interval to the size of this interval Δt via

$$\overline{\mid r_{\Delta t} \mid} = c\Delta t^{\frac{1}{E}}. \tag{A.3}$$

The mean absolute log return $\overline{\mid r_{\Delta t} \mid}$ for observations spaced 1 minute, 2 minutes, up to 10 minutes apart were computed. The $\overline{\mid r_{\Delta t} \mid}$ values for different interval sizes Δt are not totally independent, as the larger intervals are aggregated from the smaller intervals - however as an approximation using simple least squares regression we obtain almost perfect fits with $\hat{c} = 0.0001469908$ and $\hat{E} = 2.038489$.

There are 120 trading hours between Sunday 22:00 GMT to Friday 22:00 GMT.

Define the *Activity* of the h-th hour as

$$\hat{a}(h) = \left(\frac{\hat{\mu}_h}{\hat{c}*}\right)^{\hat{E}} \tag{A.4}$$

where

$$\hat{\mu}_h = \frac{\sum_{i=1}^{50}\sum_{j=2}^{O_i} \mid \frac{r_w^{(i)}(t_j)}{\Delta t_j^{\frac{1}{\hat{E}}}} \mid \mathbb{I}(\lfloor \frac{t_j}{60} \rfloor = h)}{\sum_{i=1}^{50}\sum_{j=2}^{O_i} \mathbb{I}(\lfloor \frac{t_j}{60} \rfloor = h)} \tag{A.5}$$

is the mean time standardised absolute log return for trading hour h over all the 50 trading weeks, $\lfloor . \rfloor$ is the floor function and $\hat{c}*$ is calibrated to satisfy the normalisation condition $\sum_{h=1}^{120} \hat{a}(h) = 120$. The new virtual observation times will then be given by

$$\begin{aligned} \vartheta_j &= \vartheta_{j-1} + \hat{a}\Big(\lfloor \frac{t_j}{60} \rfloor\Big)(t_j - t_{j-1}) \quad , j = 2, \ldots, O_i, \\ \vartheta_1 &= 0. \end{aligned}$$

A physical time trading hour associated with higher volatility results in a higher *Activity* for that hour. Thus, the new virtual time scale ϑ slows down during physical time trading hours associated with higher volatility as well as speeds up during physical time trading hours associated with lower volatility. In turn, the volatility per unit of virtual time is more balanced.

The estimated *Activities* are displayed in Figure A.1. One would expect three periods of higher than normal volatility each day corresponding to the opening hours of the Asian, European and American Forex markets. The results appear in line with this expectation as three spurts in *Activities* are observed to begin at 1:00 AM GMT, 7:00 AM GMT and 1:00 PM GMT each day corresponding to the start of trade of the aforementioned markets.

Physical time trading hour $h = 1$ would be expected to be relatively quiet as only the thinner Australian and New Zealand markets are open. We have $\hat{a}(1) =$

0.5052 so virtual time will be running approximately twice as fast during physical trading hour $h = 1$. On the other hand, trading activity would be expected to pick up during physical time trading hour $h = 4$ at the open of the Hong Kong market. We have $\hat{a}(4) = 1.2018$, so virtual time will be running at approximately 83% the speed of physical time during trading hour $h = 4$.

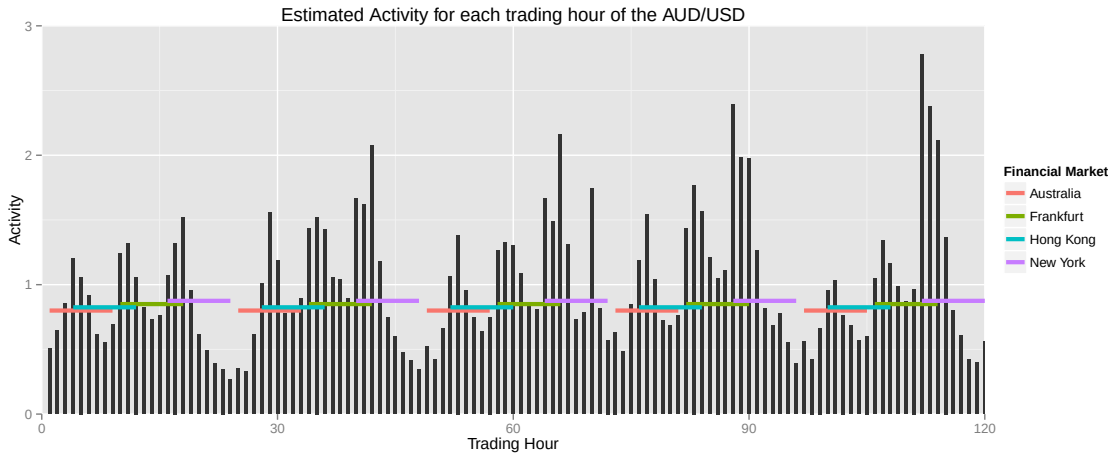


Figure A.1: Estimated *Activities* for each of the 120 trading hours (black vertical lines). The trading sessions of certain financial markets are indicated by the horizontal lines.

A.3 Resampling using Copula

A prototype method of utilising SMC with the "NORmal To Anything" (NORTA) method of Cario and Nelson [1997] is proposed in this section. The aim of this pairing is to yield, for hidden markov models that require two or more continuous variables to evolve the state equation, SMC likelihood surfaces that are amenable to numerical optimisation.

A.3.1 The NORTA method

To generate a realisation of a $d-$dimensional multivariate random variable $V = (V_1, \ldots, V_d)$, the NORTA method of random vector generation, as described by Channouf and L'Ecuyer [2009], is as follows. Denote the marginal distribution function for each component V_j, $j = 1, \ldots, d$ by $F_j(x) = \mathbb{P}(V_j \leq x)$. Calculate the *rank correlation* (a.k.a *Spearman's* ρ)

$$r^V_{i,j} = \mathbb{C}\text{orr}\big(F_i(V_i), F_j(V_j)\big) \tag{A.6}$$

for each $i, j = 1, \ldots, d$. Simulate a $d-$dimensional multivariate normal $S = (S_1, \ldots, S_d)$ with $d-$dimensional mean vector all zeroes and $d \times d-$dimensional covariance matrix Σ with each element given by

$$\Sigma_{i,j} = 2\sin(\pi r^V_{i,j}/6). \tag{A.7}$$

Then a realisation of V is given by

$$\Big(F_1^{-1}\big(\Phi(S_1)\big), \ldots, F_d^{-1}\big(\Phi(S_d)\big)\Big), \tag{A.8}$$

where $\Phi(.)$ is the standard normal cumulative distribution function and $F_j^{-1}(u) = \inf\{x : F_j(x) \geq u\}$, $j = 1, \ldots, d$.

A.3.2 SMC with NORTA resampling

Assume a hidden Markov model with $d-$dimensional hidden state X_i (with $d > 1$). Recall that at each time step $i = 1, \ldots, n$, the SMC algorithm produces a set of K particles $X_i^{(k)} = (X_{i,1}^{(k)}, \ldots, X_{i,d}^{(k)})$, $k = 1, \ldots, K$ with associated weights $w_i^{(k)} = p(Y_i|X_i = X_i^{(k)}) / \sum_{l=1}^{K} p(Y_i|X_i = X_i^{(l)})$. Instead of a bootstrap multinomial resample from the set $\{X_i^{(k)}\}_{k=1,\ldots,K}$ according to the probabilities $\{w_i^{(k)}\}_{k=1,\ldots,K}$, another approach is to use the NORTA method for resampling.

To this end, construct for $j = 1, \ldots, d$,

$$\hat{F}_j(x) = \sum_{k=1}^{K} \hat{w}_{i,j}^{(k)} \mathbb{I}(x \geq X_{i,j}^{(k)}) \tag{A.9}$$

whereby

$$\tilde{w}_{i,j}^{(k)} = \frac{\sum_{l=1}^{K} w_i^{(l)} \phi\Big((X_{i,j}^{(k)} - X_{i,j}^{(l)})/h\Big)}{\sum_{l=1}^{K} \phi\Big((X_{i,j}^{(k)} - X_{i,j}^{(l)})/h\Big)} \tag{A.10}$$

$$\hat{w}_{i,j}^{(k)} = \frac{\tilde{w}_{i,j}^{(k)}}{\sum_{l=1}^{K} \tilde{w}_{i,j}^{(l)}} \text{ for } k = 1, \ldots, K \tag{A.11}$$

where $\phi(.)$ is the standard normal density and h a very small number. Then construct for each $\hat{F}_j(x)$, $j = 1, \ldots, d$, its continuous approximation

$$\bar{F}_j(x) = \bar{w}_{i,j}^{(0)} \mathbb{I}(x \geq X_{i,j}^{(1)}) + \sum_{k=1}^{K-1} \bar{w}_{i,j}^{(k)} H\Big(\frac{x - X_{i,j}^{(k)}}{X_{i,j}^{(k+1)} - X_{i,j}^{(k)}}\Big) + \bar{w}_{i,j}^{(K)} \mathbb{I}(x \geq X_{i,j}^{(K)}) \tag{A.12}$$

where $\bar{w}_{i,j}^{(0)} = \hat{w}_{i,j}^{(1)}/2$, $\bar{w}_{i,j}^{(K)} = \hat{w}_{i,j}^{(K)}/2$ and $\bar{w}_{i,j}^{(k)} = (\hat{w}_{i,j}^{(k+1)} + \hat{w}_{i,j}^{(k)})/2$ for $k = 1, \ldots K-1$ and $H(z) := \max\Big(0, \min(z, 1)\Big)$.

Then using for each $j = 1, \ldots, d$, $\bar{F}_j(x)$ in place of $F_j(x)$ in (A.6) and $\bar{F}_j^{-1}(u)$ in place of $F_j^{-1}(u)$ in (A.8), a new realisation of X_i can be obtained by the NORTA method.

A.3.3 Pilot test case: ARMA-GARCH with missing data

The ARMA(1,1)-GARCH(1,1) is specified by

$$y_i - \varphi y_{i-1} = \epsilon_i + \vartheta \epsilon_{i-1} \tag{A.13}$$

$$\epsilon_i = \sigma_i z_i \tag{A.14}$$

$$\sigma_i^2 = \omega + \alpha \epsilon_{i-1}^2 + \beta \sigma_{i-1}^2 \tag{A.15}$$

where $|\varphi| < 1$, $|\vartheta| < 1$ and $z_i \sim$ i.i.d according to some zero mean, unit variance distribution $\mathbb{D}(0,1)$. Now we have,

$$y_i - \epsilon_i = \varphi y_{i-1} + \vartheta \epsilon_{i-1} \tag{A.16}$$

$$= \varphi(y_{i-1} - \epsilon_{i-1}) + (\varphi + \vartheta)\epsilon_{i-1}. \tag{A.17}$$

Let $\mu_i = y_i - \epsilon_i$, then

$$\mu_i = \varphi \mu_{i-1} + (\varphi + \vartheta)\epsilon_{i-1}. \tag{A.18}$$

Thus, $y_i = \mu_i + \epsilon_i$ and

$$y_i | \mu_i, \sigma_i^2 \sim \mathbb{D}(\mu_i, \sigma_i^2). \tag{A.19}$$

Assume one only has partial observation of the series $(y_i)_{i=1,\ldots,N}$ at a set of times $1 = t_1 < t_2 < \ldots < t_n = N$ where $t_i \in \{2, \ldots, N-1\}$ for $i = 2, \ldots, n-1$. The marginal likelihood of the partially observed ARMA(1,1)-GARCH(1,1) can be approximated using SMC by putting the problem in the form of (1.8) – (1.9) by setting, for $i = 1, \ldots, n$: $Y_i = y_{t_i}, Z_i = z_{t_i}, X_i = (X_{i,1}, X_{i,2}) = (\mu_{t_i}, \sigma_{t_i}^2)$,

$n_i = t_i - t_{i-1}$ (with $t_0 = 0$),

$$W_i = \begin{cases} \{z_{t_i-1}, \ldots, z_{t_{i-1}+1}\} & \text{when } n_i > 1, \\ \emptyset & \text{when } n_i = 1, \end{cases} \tag{A.20}$$

with

$$X_i = g(X_{i-1}, Y_{i-1}, W_i) = \Big(g_1(X_{i-1}, Y_{i-1}, W_i), g_2(X_{i-1}, Y_{i-1}, W_i, n_i)\Big)$$

where

$$g_1(X_{i-1}, Y_{i-1}, W_i) = \begin{cases} (\varphi + \vartheta)Y_{i-1} - \vartheta X_{i-1,1} & \text{when } n_i = 1, \\ \varphi^{n_i} X_{i-1,1} + \varphi^{n_i-1}(\varphi + \vartheta)(Y_{i-1} - X_{i-1,1}) & \\ + (\varphi + \vartheta) \displaystyle\sum_{k=1}^{n_i-1} \varphi^{k-1} z_{t_i-k} \sqrt{g_2(X_{i-1}, Y_{i-1}, W_i, n_i - k)} & \text{when } n_i > 1, \end{cases} \tag{A.21}$$

$$g_2(X_{i-1}, Y_{i-1}, W_i, u) = \begin{cases} \omega + \alpha(Y_{i-1} - X_{i-1,1})^2 + \beta X_{i-1,2} & \text{when } u = 1, \\ \omega + \omega\Big(\displaystyle\sum_{j=1}^{u-1} \prod_{k=1}^{j} (\alpha z_{t_i-k}^2 + \beta)\Big) & \\ + (\alpha(Y_{i-1} - X_{i-1,1})^2 + \beta X_{i-1,2}) \displaystyle\prod_{j=1}^{u-1} (\alpha z_{t_i-j}^2 + \beta) & \text{when } u > 1. \end{cases} \tag{A.22}$$

Then $Y_i = f(X_i, Z_i) = X_{i,1} + \sqrt{X_{i,2}} Z_i$, with

$$p(Y_i | X_i) = \frac{1}{\sqrt{X_{i,2}}} q\Big(\frac{Y_i - X_{i,1}}{\sqrt{X_{i,2}}}\Big), \tag{A.23}$$

whereby q is the density of z_i.

This application is an example of a situation when two continuous variables are required to evolve the state equation.

The left hand side of Figure A.2 displays approximated log-likelihood profile plots using SMC with bootstrap resampling (keeping randomness fixed across parameter changes), for a simulated, partially observed ARMA(1,1)-GARCH(1,1) data set. The right hand side of Figure A.2 displays approximated log-likelihood profile plots using SMC with NORTA resampling, for the same data set. As is seen, while SMC with NORTA resampling does not appear to lead to likelihood surfaces that are perfectly smooth, they are considerably smoother than their bootstrap resampled counterparts and in fact, demonstrated in the simulation study below, to be sufficiently smooth for numerical optimisation to yield reliable parameter estimates.

A.3.3.1 A simulation study

We simulate 1000 data sets of a lenght $N = 2000$, ARMA(1,1)-GARCH(1,1) series with standard normal innovations and true parameters $\omega = 0.1, \alpha = 0.08, \beta = 0.9, \varphi = 0.7$ and $\vartheta = -0.55$. Within each data set, each observation at time $i = 2, \ldots, N-1$ was with a 20% independent chance set to missing. The configuration of missing observations is different across data sets.

The results from maximising the log-likelihood surface obtained utilising SMC with NORTA resampling, using 200 particles, for these data sets are summarised in Table A.1, with kernel density plots of the estimates obtained provided in Figure A.3. Based on these preliminary results, while likelihood surfaces produced by this method are not perfectly smooth, the improvement in *smoothness*, appears at least in this scenario, to be sufficent for numerical optimisation to yield fairly accurate parameter estimates.

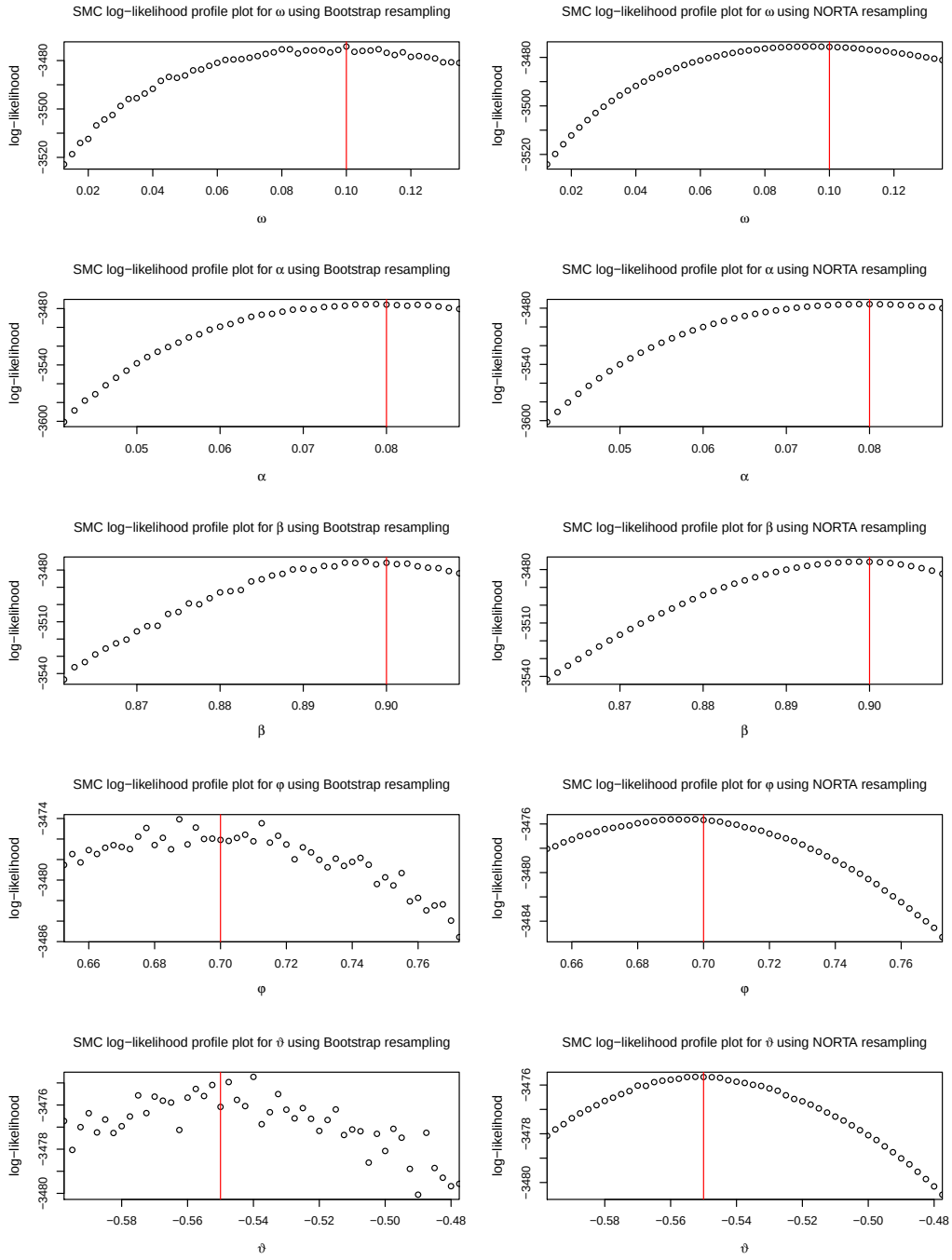


Figure A.2: SMC approximated log-likelihood profile plots, for each of the five ARMA(1,1)-GARCH(1,1) model parameters (LHS: using bootstrap resampling, RHS: using NORTA resampling), for a simulated, partially observed ARMA(1,1)-GARCH(1,1) data set ($N = 2000$ with approximately 20% of observations missing). The true parameter values are indicated by the red vertical lines.

Table A.1: Summary statistics of the parameter estimates obtained, across the 1000 simulated, partially observed ARMA(1,1)-GARCH(1,1) data sets, by utilising SMC with NORTA resampling using 200 particles. Refer to Section 3.2.1 for definitions of the summary statistics **Mean**, **Bias**, **SD** and **RMSE**.

	ω	α	β	φ	ϑ
Mean	0.1271	0.0825	0.8913	0.6746	-0.5240
Bias	0.0271	0.0025	-0.0087	-0.0254	0.0260
SD	0.0529	0.0149	0.0217	0.0974	0.1093
RMSE	0.0594	0.0151	0.0234	0.1006	0.1124

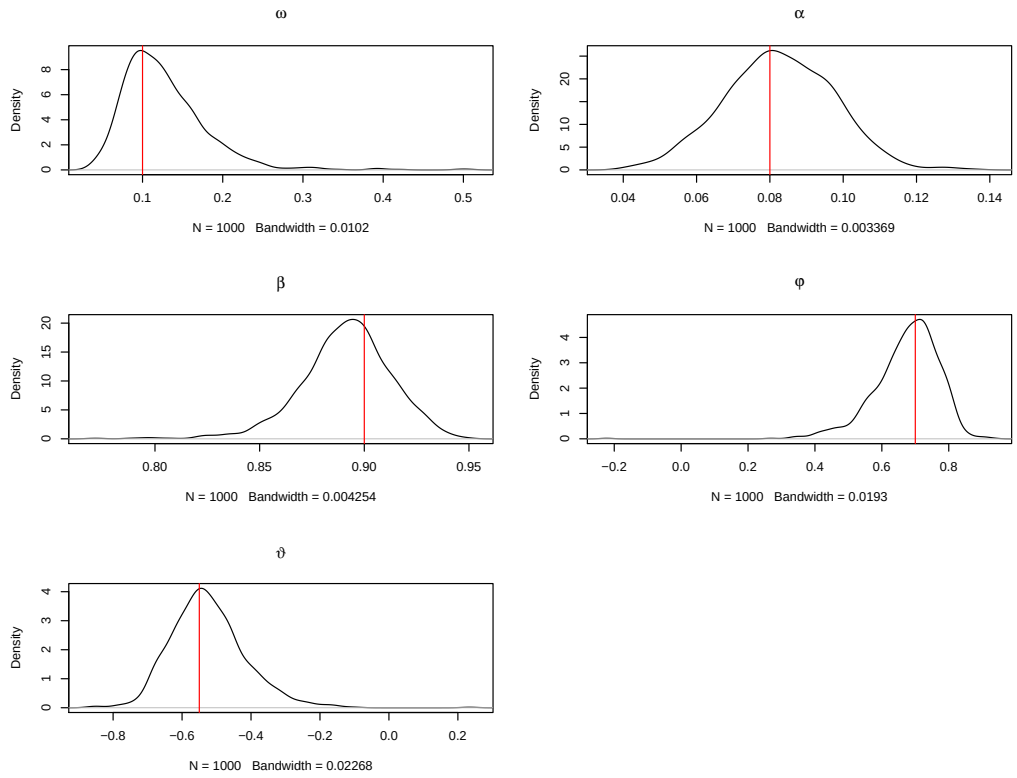


Figure A.3: Kernel density plots of the parameter estimates obtained from the 1000 simulated, partially observed ARMA(1,1)-GARCH(1,1) data sets, utilising SMC with NORTA resampling employing 200 particles. The true parameter values are indicated by the red vertical lines.

References

C. Alexander. *Market Risk Analysis: Pricing, Hedging and Trading Financial Instruments*. John Wiley and Sons Ltd, 2008. 65

S. Alterman. Natural gas price volatility in the UK and North America. *Oxford Institute for Energy Studies*, NG, 60, 2012. 143

M. Augustyniak. Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics and Data Analysis*, 76:61–75, 2014. 106, 107

M. Augustyniak, M. Boudreault, and M. Morales. Maximum Likelihood Estimation of the Markov-Switching GARCH Model Based on a General Collapsing Procedure. *Methodology and Computing in Applied Probability*, 2017. doi: 10.1007/s11009-016-9541-4. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2365763. 106, 107, 110, 112, 115, 145

L. Bauwens, A. Preminger, and J. Rombouts. Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13:218–244, 2010. vi, 1, 104, 105, 106, 130

L. Bauwens, A. Dufays, and J. Rombouts. Marginal likelihood for Markov-switching and change-point GARCH. *Journal of Econometrics*, 178:508–522, 2014. 106

A. Behme, C. Klppelberg, and K. Mayr. Asymmetric COGARCH processes. *Journal of Applied Probability*, 51(A):161–173, 2014. 102

E. Bibbona and I. Negri. Higher moments and prediction based estimation for the COGARCH(1,1) model. *Scandinavian Journal of Statistics*, 42(4):891–910, 2015. 78, 101

M. Bildirici and O. Ersin. Modeling Markov Switching ARMA-GARCH Neural Networks Models and an Application to Forecasting Stock Returns. *The Scientific World Journal*, 2014. 106

M. Billio, R. Casarin, and A. Osuntuyi. Efficient Gibbs sampling for Markov switching GARCH models. *Computational Statistics and Data Analysis*, 100: 37–57, 2016. 106

T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986. v, 1, 2

P. Bondon and N. Bahamonde. Least squares estimation of ARCH models with missing observations. *Journal of Time Series Analysis*, 33(6):880–891, 2012. 4, 34, 51, 52

O. Cappe. Online sequential Monte Carlo EM algorithm. *In Proceedings of the IEEE Workshop on Statistical Signal Processing, Cardiff, Wales, U.K.*, pages 37–40, 2009. 150

O. Cappe. Online EM Algorithm for Hidden Markov Models. *Journal of Computational and Graphical Statistics*, 20(3):728–749, 2011. 150

M.C. Cario and B.L. Nelson. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. *Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston*, 1997. 158

N. Channouf and P. L'Ecuyer. Fitting a normal copula for a multivariate distribution with both discrete and continuous marginals. *In Proceedings of the 2009 Winter Simulation Conference*, pages 352–358, 2009. 158

E. Cripps and W.T.M Dunsmuir. Modeling the Variability of Sydney Harbor Wind Measurements. *Journal of Applied Meteorology*, 42(8):1131–1138, 2003. 2, 37

M. Cunningham. Weather, mood, and helping behavior: Quasi experiments with the sunshine samaritan. *Journal of Personality and Social Psychology*, 37(11): 1947–1956, 1979. 40

M. Dacorogna, U. Muller, R. Nagler, R. Olsen, and O. Pictet. A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market. *Journal of International Money and Finance*, 12:413–438, 1993. 155

R.A. Davis and G. Rodriguez-Yam. Estimation for State-Space Models: an Approximate Likelihood Approach. *Statistica Sinica*, 15:381–406, 2005. 7

T.A. Dean, S.S. Singh, A. Jasra, and G.W. Peters. Parameter Estimation for Hidden Markov Models with Intractable Likelihoods. *Scandinavian Journal of Statistics*, 41:970–987, 2014. 150

P. Del Moral. Non Linear Filtering: Interacting Particle Solution. *Markov Processes and Related Fields*, 2(4):555–580, 1996. 8, 16, 106

P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli*, 18(1):252–278, 2012. 149

F.X. Dieobold. Modeling The Persistence Of Conditional Variances: A Comment. *Econometric Reviews*, 5(1):51–56, 1986. 103

A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *In: The Oxford Handbook of Nonlinear Filtering (D. Crisan and B. Rozovsky, eds.), Oxford University Press*, pages 656–705, 2011. 149

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000. 149

F.C. Drost and T.E Nijman. Temporal Aggregation of GARCH Processes. *Econometrica*, 61:909–927, 1993. 20, 29, 31, 48, 53, 62, 63, 76

M. J. Dueker. Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business and Economic Statistics*, 15(1):26–34, 1997. 105, 106, 113

W.T.M Dunsmuir. Estimation for stationary time series when data are irregularly spaced or missing. In Findley DF (ed.), Applied time series analysis II : proceedings of the Second Applied Time Series Symposium held in Tulsa, Oklahoma, March 3-5, 1980, New York : Academic Press, Tulsa, OK, pp. 609 - 650, 1981. 23

R.J. Elliott, J.W. Lau, H. Miao, and T.K. Siu. Viterbi-based estimation for Markov switching GARCH models. *Applied Mathematical Finance*, 19(3):219–231, 2012. 106

R. F. Engle and V. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48:1749–1778, 1993. 48

R.F. Engle and A.J. Patton. What good is a volatility model? *Quantitative Finance*, 1:237–245, 2001. 2

I. Fisher. What is Capital? *The Economic Journal*, 6(24):509–534, 1896. 52

P. Fortune. Are Stock Returns Different Over Weekends? A Jump Diffusion Analysis of the Weekend Effect. *New England Economic Review*, pages 3–19, 1999. 34

C. Francq and J.M. Zakoian. Estimating Weak GARCH Representations. *Econometric Theory*, 16:692–728, 2000. 63, 64

C. Francq and J.M. Zakoian. Deriving the autocovariances of powers of Markov-

switching GARCH models, with applications to statistical inference. *Computational Statistics and Data Analysis*, 52(6):3027–3046, 2008. 104, 106, 108

C. Francq, M. Roussignol, and J.M. Zakoian. Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22:197–220, 2001. 104

L. Glosten, R. Jagannathan, and D. Runkle. On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance*, 46:1779–1801, 1992. 1, 48

N. J. Gordon, D. J. Salmond, and A. F. Smith. A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F*, 140:107–113, 1993. 7, 15, 79, 116

C. Gourieroux. *ARCH Models and Financial Applications*. Springer-Verlag, New York, 1997. 74

S.F. Gray. Modeling the conditional distribution of interest-rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996. 105, 106

M. Haas, S. Mittnik, and M.S. Paolella. A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530, 2004. 105

J.D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989. 104

J.D Hamilton and G. Lin. Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593, 1996. 103

L. Harris. A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, 16(1):99–117, 1986. 43

J.M. Harrison and S.R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260, 1991. 77

S. Haug, C. Kluppelberg, A. Lindner, and M Zapp. Method of Moment Estimation in the COGARCH(1,1) Model. *The Econometrics Journal*, 10(2):320–341, 2007. 78, 81

C. Heyde. *Quasi-likelihood and its applications*. Springer, 1997. 95

D. Hirshleifer and T. Shumway. Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance*, 58(3):1009–1032, 2003. 40

E. Howarth and M. S. Hoffman. A multidimensional approach to the relationship between mood and weather. *British Journal of Psychology*, 75(1):15–23, 1984. 40

S. Huag and C. Czado. An exponential continuous time GARCH process. *Journal of Applied Probability*, 44:960–976, 2007. 102

J. Jeon and J.W. Taylor. Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation. *International Journal of Forecasting*, 32(3):991–1004, 2016. 37

N. Kantas, A. Doucet, S.S. Singh, and J.M. Maciejowski. An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-Space Models. *IFAC Proceedings Volumes*, 42(10):774–785, 2009. 15

M.J Kay, N. Cutler, A. Micolich, I. MacGill, and H. Outhred. Emerging challenges in wind energy forecasting for Australia. *Australian Meteorological and Oceanographic Journal*, 58:99–106, 2009. 37

M. Kim and S. Lee. On the maximum likelihood estimation for irregularly observed

data from COGARCH(1,1) models. *REVSTAT - Statistical Journal*, 11(2):135–168, 2013. 78

F.J.G.M. Klaassen. Improving GARCH Volatility Forecasts with Regime-Switching GARCH. *Empirical Economics*, 27:363–394, 2002. 105, 106, 113

C. Kluppelberg, A. Lindner, and R. Maller. A Continous-Time GARCH Process Driven by a Levy Process: Stationarity and Second Order Behaviour. *Journal of Applied Probability*, 41(3):601–622, 2004. vi, 1, 9, 11, 77, 80, 81, 91

C. Kuppelberg, A. Lindner, and R. Maller. Continous Time Volatility Modelling: COGARCH Versus Ornstein-Uhlenbeck Models. *In: The Shiryaev Festschrift: From Stochastic Calculus to Mathematical Finance (Y. Kabanov, R. Lipster and I. Stoyanov, eds.)*, pages 393–419, 2006. 91

C. Lamoureux and W. Lastrapes. Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. *Journal of Finance*, 45:221–229, 1990. 103

A. Lee. Towards Smooth Particle Filters for Likelihood Estimation with Multivariate Latent Variables. Master's thesis, University of British Columbia, Vancouver, Canada, 2008. 148

S.W. Lee and B.E. Hansen. Asymptotic theory for the GARCH(l.l) quasi-maximum likelihood estimator. *Econometric Theory*, 10:29–52, 1994. 31

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47:13–37, 1965. 43

J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York, 2001. 149

R. Maller, G. Muller, and A. Szimayer. GARCH Modelling in Continuous Time

for Irregularly Spaced Time Series Data. *Bernoulli*, 14:519–542, 2008. ix, 78, 81, 93, 95, 100, 101

J. M. Marin, M. T. Rodriguez-Bernal, and E. Romero. Data cloning estimation of GARCH and COGARCH models. *Journal of Statistical Computation and Simulation*, 85(9):1818–1831, 2015. 78

T. Mikosch and C. Starica. Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *Review of Economics and Statistics*, 86:378–390, 2004. 103

G. Muller. MCMC Estimation of the COGARCH(1,1) Model. *Journal of Financial Econometrics*, 8(4):481–510, 2010. 78, 79

U. Muller, M. Dacorogna, R. Olsen, O. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance*, 14:1189–1208, 1990. 90, 155

K. M. Murphy and R. H. Topel. Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, 3:370–379, 1985. 95

D. B. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990. 3, 105

D.B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991. 1, 48

W.K. Newey and K.D. West. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708, 1987. 40

T. Nijman and E. Sentana. Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics*, 71:71–80, 1996. 63

G.W. Oehlert. A Note on the Delta Method. *American Statistician*, 46(1):27–29, 1992. 64

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

M.A. Persinger. Lag responses in mood reports to changes in the weather matrix. *International Journal of Biometeorology*, 19(2):108–114, 1975. 40

M. Pitt and S. Malik. Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, 165:190–209, 2011. v, vi, 9, 10, 11, 13, 17, 18, 49, 58, 60, 79, 117, 120, 121, 147

M. Pitt, S. Malik, and A. Doucet. Simulated likelihood inference for stochastic volatility models using continuous particle filtering. *Annals of the Institute of Statistical Mathematics*, 66(3):527–552, 2014. 35

D.L. Poole and A.K. Mackworth. *Artificial Intelligence Foundations of Computational Agents*. Cambridge University Press, 2010. 7

P. Protter. *Stochastic Integration and Differential Equations*. Springer, New York, 2nd edition, 2005. 10, 80

M. Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23:470–472, 1952. 35

S. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976. 5

D.B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons Inc., 1987. 25

E.M. Saunders. Stock Prices and Wall Street Weather. *American Economic Review*, 83:1337–1345, 1993. 40

G.W. Schwert. Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, 44(5):1115–1153, 1989. 103

E. Sentana. Quadratic ARCH models. *Review of Economic Studies*, 62:639–661, 1995. 48

W.F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964. 43

R.J. Shiller. From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, 17(1):83–104, 2003. 40

T. Terasvirta. Introduction to univariate GARCH models. In *Handbook of Financial Time Series*, pages 17–42. Springer, 2009. 49

P.J Trombe, P. Pinson, and H. Madsen. A general probabilistic forecasting framework for offshore wind power fluctuations. *Energies*, 5(3):621–657, 2012. 37

R. van der Merwe, A. Doucet, J.F.G. de Freitas, and E. Wan. The Unscented Particle Filter. *Advances in Neural Information Processing Systems*, 8:351–357, 2000. 149

H. White. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817–838, 1980. 40

A. Worthington. An Empirical Note on Weather Effects in the Australian Stock Market. *Economic Papers*, 28(2):148–154, 2009. 40, 41, 42