

ESTIMATING GARCH(1, 1) IN THE PRESENCE OF MISSING DATA

BY DAMIEN C. H. WEE^a, FENG CHEN^b AND WILLIAM T. M. DUNSMUIR^c

*School of Mathematics and Statistics, UNSW Sydney, ^adchwee@gmail.com, ^bfeng.chen@unsw.edu.au,
^cw.dunsmuir@unsw.edu.au*

Maximum likelihood estimation of the famous GARCH(1, 1) model is generally straightforward, given the full observation series. However, when some observations are missing, the marginal likelihood of the observed data is intractable in most cases of interest, also intractable is the likelihood from temporally aggregated data. For both these problems, we propose to approximate the intractable likelihoods through sequential Monte Carlo (SMC). The SMC approximation is done in a smooth manner so that the resulting approximate likelihoods can be numerically optimized to obtain parameter estimates. In the case with data aggregation, the use of SMC is made possible by a novel state space representation of the aggregated GARCH series. Through extensive simulation experiments, the proposed method is found to be computationally feasible and produce more accurate estimators of the model parameters compared with other recently published methods, especially in the case with aggregated data. In addition, the Hessian matrix of the minus logarithm of the approximate likelihood can be inverted to produce fairly accurate standard error estimates. The proposed methodology is applied to the analysis of time series data on several exchange-traded funds on the Australian Stock Exchange with missing prices, due to interruptions such as scheduled trading holidays.

1. Introduction. This work is motivated by volatility modelling of returns of international exposure ETFs (exchange-traded funds) traded on the Australian Stock Exchange (ASX). On local public holidays, when the Australian exchange is closed, one does not typically receive pricing information for these ETFs, yet their underlying international holdings in fact remain subject to trading activity on offshore exchanges, so their prices continue to evolve despite no observation being available. While a natural model for financial return volatility is the celebrated GARCH(1, 1) model (Bollerslev (1986)), its application in situations like this is complicated by the presence of missing data.

Similar problems also appear in the analysis of economic time series and other applications. With respect to economic time series, quarterly economic indicators could be considered to have two-thirds the corresponding monthly time series missing. There has been considerable interest, by both academics and industry practitioners alike, to perform what is termed *Nowcasting* which are procedures to obtain higher frequency estimates of economic indicators that are published at lower frequencies (Higgins (2014)). Also, somewhat commonplace in economic time series are data currently published at monthly or quarterly frequencies but only recorded at lower frequencies in the past. Outside of financial applications, such as in wind speed volatility modelling, mechanical failures of measurement equipment can indeed lead to gaps in the data (Cripps and Dunsmuir (2003)).

This paper proposes a new methodology for fitting the widely used GARCH(1, 1) model when faced with missing data in time series. We deal with two types of missingness: *missing* observations and *aggregated* observations. To illustrate the distinction, consider a time series

Received January 2022; revised January 2023.

Key words and phrases. Hidden Markov model, incomplete data, intractable likelihood, sequential Monte Carlo, volatility modelling.

of daily log returns derived between the day's opening and closing prices. Bear in mind that the next day's opening price will not equal today's closing price due to continual shifts in investor's preferences post market close. When pricing information is not recorded for a particular day (be it a trading holiday or otherwise), observation of the day's open to close return is indeed *missing*. However, if the time series is of log returns calculated between daily closing prices and one day of pricing information is missed, then treating the situation as two one-day close to close returns being missing is rather wasteful. This is because a two-day close to close return is available and, therefore, can be treated as an *aggregated* observation of two one-day close to close log returns.

Fitting GARCH models, when the time series is fully observed, is straightforward because computation of the likelihood can be done using recursive calculations of conditional densities. However, when there are gaps in the observation series, integration is needed to compute the likelihood with the order of integration rendering the likelihood intractable in most cases of interest.

For fitting GARCH models in the presence of missing observations, an indirect inference method has been proposed by Blasques, Gorgi and Koopman (2021), while Cascone and Hotta (2019) tackle the problem of fitting GARCH models from aggregated observations using a quasi-likelihood estimator. Gábor, Sávai and Udvari (2017) in their empirical analysis, utilized expectation maximization (EM) to overcome missing data when fitting GARCH models, although they did not perform a simulation study to assess the accuracy of such an approach. However, a similar EM method was assessed in Sucarrat and Escribano (2018) for fitting log-GARCH models, treating zeros in the observation series as missing. It was found to produce biased estimators with the biases increasing with the percentage of zeros. A nonlinear state space formulation of GARCH models is presented in Ossandón and Bahamonde (2013) for which the authors suggest one could attempt to handle missing observations through their formulation using an extended Kalman filter, but performance of such an approach was not evaluated. For fitting the simpler ARCH model in the presence of missing observations, a least squares estimator was developed in Bondon and Bahamonde (2012). Note that most of these papers assume that the missingness is generated by a stationary and ergodic process of zeros and ones which is independent of the observed GARCH process. Others assume that the missingness process occurs in regular patterns. Our method applies to both of these cases and indeed to any method of generating the missing data as long as it is independent of the observed process. To the best of our knowledge, our method is the only one currently available that handles both data missingness and aggregation.

In this paper we pursue approximation of the intractable likelihood of GARCH(1, 1) models when challenged with either missing or aggregated observations through sequential Monte Carlo (SMC). Our SMC likelihood approximation is smooth as a function of the model parameters, and, therefore, it can be easily numerically optimized to obtain the (approximate) maximum likelihood estimator (MLE) of the parameters. Obtaining parameters this way is shown to provide improved estimation performance over the recent methods of Blasques, Gorgi and Koopman (2021) and Cascone and Hotta (2019). The method proves to deliver a computationally feasible and reliable estimator. Moreover, fairly accurate standard error estimates are obtained by inverting the Hessian matrix of the SMC approximated negative log-likelihood. For dealing with missing observations in the GARCH series, application of the SMC for smooth likelihood estimation is fairly straightforward; however, in the case of aggregated observations the use of SMC requires identification of a novel state space representation.

The remainder of this paper is organized as follows. Section 2.1 outlines a general SMC framework for smooth likelihood estimation of a hidden Markov system. Formulating the problem of likelihood estimation for GARCH models with missing observations under the

SMC framework is detailed in Section 2.2. A state space representation of temporally aggregated GARCH series to enable smooth SMC approximation of the likelihood is presented in Section 2.3. Estimation performance is assessed through simulation studies in Section 3 (Section 3.1 for missing observations and Section 3.2 for aggregated observations). In Section 4 we apply the proposed methodology to the analysis of the return time series of three ETFs traded on the ASX. Section 5 concludes with some remarks.

2. Methodology.

2.1. *SMC framework.* Consider the following hidden Markov system:

$$(1) \quad Y_i = f(X_i, Z_i),$$

$$(2) \quad X_i = g(X_{i-1}, Y_{i-1}, W_i),$$

for $i = 1, \dots, n$, where $(Y_i)_{i=1, \dots, n}$ is the observation series, $(X_i)_{i=1, \dots, n}$ is an unobserved discrete time Markov process, $(Z_i)_{i=1, \dots, n}$ is a serially independent noise series, $(W_i)_{i=1, \dots, n}$ is another serially independent noise series that is independent of $(Z_i)_{i=1, \dots, n}$ with f and g being suitably defined functions. The observations $(Y_i)_{i=1, \dots, n}$ are serially dependent, although conditionally independent given the hidden state X_i .

Suppose we wish to estimate a set of parameters θ that govern the system (1)–(2). The likelihood is given as follows:

$$(3) \quad \text{lik}(\theta) = p(Y_1) \prod_{i=2}^n p(Y_i|Y_{1:i-1}) = \int p(Y_1|X_1)\mathbb{P}(dX_1) \prod_{i=2}^n \int p(Y_i|X_i)\mathbb{P}(dX_i|Y_{1:i-1}).$$

In most cases of interest, however, the conditional distributions $\mathbb{P}(dX_i|Y_{1:i-1})$ are intractable which hinders direct evaluation of (3). In these cases one could resort to sequential Monte Carlo (SMC) to approximate the integrals in (3). The SMC involves simulating a sample of independent and identically distributed (i.i.d.) observations (called *particles*) from $\mathbb{P}(dX_i|Y_{1:i-1})$ or an approximation to it, $\tilde{\mathbb{P}}(dX_i|Y_{1:i-1})$, and then approximating the integral with respect to $\mathbb{P}(dX_i|Y_{1:i-1})$ in (3) by the empirical mean of the values of the integrand at the particles. Specifically, if we let $X_i^{(1:K)} = \{X_i^{(1)}, \dots, X_i^{(K)}\}$ denote the particles at time i and let $\hat{\mathbb{P}}(dX_i|Y_{1:i-1}) = \frac{1}{K} \sum_{k=1}^K \delta_{X_i^{(k)}}(dX_i)$, with $\delta_x(\cdot)$ the Dirac measure at x , denote their empirical distribution, then the approximation to the integral is

$$(4) \quad \int p(Y_i|X_i)\mathbb{P}(dX_i|Y_{1:i-1}) \approx \int p(Y_i|X_i)\hat{\mathbb{P}}(dX_i|Y_{1:i-1}) = \frac{1}{K} \sum_{k=1}^K p(Y_i|X_i^{(k)}).$$

The SMC approximation of the likelihood is then given by

$$(5) \quad \widehat{\text{lik}}(\theta) = \prod_{i=1}^n \left[\frac{1}{K} \sum_{k=1}^K p(Y_i|X_i^{(k)}) \right].$$

At this point we emphasize that two approximations to the conditional distribution $\mathbb{P}(dX_i|Y_{1:i-1})$ are involved: the “tilde” version $\tilde{\mathbb{P}}(dX_i|Y_{1:i-1})$ is used to generate particles, while the “hat” version $\hat{\mathbb{P}}(dX_i|Y_{1:i-1})$ is used both to approximate the integrals as in (4) and to produce the “tilde” version at the next time step, as we shall explain next.

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

$$(6) \quad \mathbb{P}(dX_i|Y_{1:i-1}) = \int \mathbb{P}(dX_i|X_{i-1}, Y_{i-1})\mathbb{P}(dX_{i-1}|Y_{1:i-1}),$$

which motivates the “tilde” version approximation

$$(7) \quad \tilde{\mathbb{P}}(dX_i|Y_{1:i-1}) = \int \mathbb{P}(dX_i|X_{i-1}, Y_{i-1})\widehat{\mathbb{P}}(dX_{i-1}|Y_{1:i-1}),$$

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

$$(8) \quad \begin{aligned} \mathbb{P}(dX_{i-1}|Y_{1:i-1}) &= \frac{p(Y_{i-1}|X_{i-1})\mathbb{P}(dX_{i-1}|Y_{1:i-2})}{\int p(Y_{i-1}|X_{i-1})\mathbb{P}(dX_{i-1}|Y_{1:i-2})} \\ &\approx \frac{\sum_{k=1}^K p(Y_{i-1}|X_{i-1}^{(k)})\delta_{X_{i-1}^{(k)}}(dX_{i-1})}{\sum_{k=1}^K p(Y_{i-1}|X_{i-1}^{(k)})} =: \widehat{\mathbb{P}}(dX_{i-1}|Y_{1:i-1}). \end{aligned}$$

Define $w_i^k := p(Y_i|X_i^{(k)}) / \sum_{j=1}^K p(Y_i|X_i^{(j)})$ for $k = 1, \dots, K, i = 1, \dots, n$. Then combining (7) and (8) gives

$$(9) \quad \tilde{\mathbb{P}}(dX_i|Y_{1:i-1}) = \sum_{k=1}^K \mathbb{P}(dX_i|X_{i-1}^{(k)}, Y_{i-1})w_{i-1}^k.$$

To simulate a particle according to (9), one would first sample an index ν from the set $\{1, \dots, K\}$, according to the probabilities $\{w_{i-1}^1, \dots, w_{i-1}^K\}$, and then simulate a random variable W_i to obtain $X_i = g(X_{i-1}^{(\nu)}, Y_{i-1}, W_i)$. Equivalently, this process can be thought of as *bootstrap* resampling of the particles from the previous time step $\{X_{i-1}^{(1)}, \dots, X_{i-1}^{(K)}\}$ and then evolving this bootstrapped sample, rather than the original sample, to generate the X_i samples 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 depiction below summarizes how, starting from the predictive distribution of the hidden state at time i , $\tilde{\mathbb{P}}(dX_i|Y_{1:i-1})$, one obtains an approximation to the likelihood component $p(Y_i|Y_{1:i-1})$ and then updates the predictive distribution to that at time $i + 1$, $\widehat{\mathbb{P}}(dX_{i+1}|Y_{1:i})$,

$$\dots \rightarrow \tilde{\mathbb{P}}(dX_i|Y_{1:i-1}) \xrightarrow{\text{draw}} \widehat{\mathbb{P}}(dX_i|Y_{1:i-1}) \xrightarrow{(8)} \widehat{\mathbb{P}}(dX_i|Y_{1:i}) \xrightarrow{(9)} \tilde{\mathbb{P}}(dX_{i+1}|Y_{1:i}) \rightarrow \dots$$

(4)↓
 $\widehat{\text{lik}}(\theta)$ component $\widehat{p}(Y_i|Y_{1:i-1})$

The bootstrap resampling SMC procedure we have described was developed by Gordon, Salmond and Smith (1993). It was established in Del Moral (1996) that, even for a finite number of particles K , (5) is an unbiased estimator for the true likelihood function. Computation underflow, due to multiplication of many small terms, can hinder evaluation of (5). An alternative is to calculate $\log(\widehat{\text{lik}}(\theta))$ which is an asymptotically (as $K \rightarrow \infty$) unbiased estimator of the true log-likelihood (see Del Moral (1996)).

While bootstrap resampling leads to more samples in regions of substantial probability mass, it comes with the drawback that the SMC approximated likelihood surface, obtained in this way, is generally not amenable to numerical optimization. Even with the Monte Carlo randomness being fixed, the resulting likelihood surface approximation is typically not a smooth function of the model parameters. The reason is that the particles are effectively drawn from the discontinuous empirical cumulative distribution function (ECDF)

$$(10) \quad \widehat{F}_K(x) = \widehat{\mathbb{P}}(X_i \leq x|Y_{1:i}) = \sum_{k=1}^K \widehat{\mathbb{P}}(X_i^{(k)}|Y_{1:i})\mathbb{I}(X_i^{(k)} \leq x),$$

where $\mathbb{I}(\cdot)$ is the indicator function and $\widehat{\mathbb{P}}(X_i^{(k)} | Y_{1:i}) = w_i^k, k = 1, \dots, K$. Thus, even minute changes in the weights for the particles can cause sudden large changes in the resampled particles.

In the case when the support of X_i is some interval of the real line, [Malik and Pitt \(2011\)](#) provide a simple solution to remedy the discontinuity problem by constructing a continuous approximation $\widetilde{F}_K(x)$ of $\widehat{F}_K(x)$ and then generating particles by inverting uniforms based on $\widetilde{F}_K(x)$. Assuming $X_i^{(k)}, k = 1, \dots, K$ are sorted in ascending order; then

$$(11) \quad \widetilde{F}_K(x) = \pi_0 \mathbb{I}(x \geq X_i^{(1)}) + \sum_{k=1}^{K-1} \pi_k H\left(\frac{x - X_i^{(k)}}{X_i^{(k+1)} - X_i^{(k)}}\right) + \pi_K \mathbb{I}(x \geq X_i^{(K)}),$$

where $\pi_0 = \widehat{\mathbb{P}}(X_i^{(1)} | Y_{1:i})/2, \pi_K = \widehat{\mathbb{P}}(X_i^{(K)} | Y_{1:i})/2$ and $\pi_k = (\widehat{\mathbb{P}}(X_i^{(k+1)} | Y_{1:i}) + \widehat{\mathbb{P}}(X_i^{(k)} | Y_{1:i}))/2$ for $k = 1, \dots, K - 1$, with $H(z) := \max(0, \min(z, 1))$. It was shown in [Malik and Pitt \(2011\)](#) that the distance $\|\widehat{F}_K(x) - \widetilde{F}_K(x)\|_\infty$ is of order K^{-1} .

2.2. Missing observations. Formally, the GARCH(1, 1) is the following system of recurrence equations:

$$(12) \quad y_i = \sigma_i z_i,$$

$$(13) \quad \sigma_i^2 = \omega + \alpha y_{i-1}^2 + \beta \sigma_{i-1}^2,$$

for $i = 1, \dots, N$, where $\omega > 0, \alpha \geq 0, \beta \geq 0$ with $(z_i)_{i=1, \dots, N}$ an i.i.d. noise series with density h and finite second moment.

Assume that we observe a GARCH(1, 1) series on a set of $n < N$ time points $t_1 = 1 < t_2 < \dots < t_n = N$, where $t_j \in \{2, \dots, N - 1\}$ for $j = 2, \dots, n - 1$. We also assume an independent missingness process so that the observation time points t_1, \dots, t_n do not contain information on the underlying GARCH process. 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. The marginal likelihood of observations $y_{\mathbb{O}}$ is then given by

$$(14) \quad p_Y(y_{\mathbb{O}}; \theta) = \int_{\mathbb{R}^{N-n}} p_Y(y_1, \dots, y_N; \theta) \, dy_{\mathbb{M}}.$$

The likelihood defined in (14) requires evaluation of integrals of dimension $N - n$, which for even a modest amount of missing observations is intractable, making maximum likelihood infeasible.

Throughout we let $n_i = t_i - t_{i-1}, i = 2, \dots, n$ with $n_1 = t_1$. Fortunately, it is straightforward to put the problem of GARCH(1, 1) with missing observations in the hidden Markov model form of (1)–(2) by setting, for $i = 1, \dots, n, Y_i = y_{t_i}, Z_i = z_{t_i}, X_i = \sigma_{t_i}^2$ and taking $W_i = \{z_{t_i-1}, \dots, z_{t_{i-1}+1}\}$ (if $n_i = 1$ then $W_i \equiv \emptyset$). Then $Y_i = f(X_i, Z_i) = \sqrt{X_i} Z_i$ and

$$(15) \quad p(Y_i | X_i) = \frac{1}{\sqrt{X_i}} h\left(\frac{Y_i}{\sqrt{X_i}}\right).$$

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. With the resampling mechanism of [Malik and Pitt \(2011\)](#), a continuous log-likelihood surface can be constructed and, subsequently, maximized to obtain parameter estimates and their standard errors using the Hessian. Pseudocode detailing the methodology is presented in Algorithm 1.

Algorithm 1 Smooth SMC Log-Likelihood Estimation Missing GARCH

```

1: Initialization:
2:  $X_1^{1:K} \leftarrow \frac{\omega}{1-\alpha-\beta}$ 
3:  $\text{loglik} \leftarrow t_0 \leftarrow 0$ 
4: Set seed to fix randomness
5: Loop:
6: for  $i$  in  $1 : n$  do
7:    $w^{1:K} \leftarrow 0$ 
8:   for  $k$  in  $1 : K$  do
9:      $\sigma^{2(k)} = X_i^{(k)}$ 
10:    if  $n_i = t_i - t_{i-1} > 1$  then
11:      for  $j$  in  $1 : n_i - 1$  do
12:        Simulate noise variable  $z$ 
13:         $\sigma^{2(k)} \leftarrow \omega + (\alpha z^2 + \beta)\sigma^{2(k)}$ 
14:         $w^{(k)} = \frac{1}{\sqrt{\sigma^{2(k)}}} h\left(\frac{Y_i}{\sqrt{\sigma^{2(k)}}}\right)$ 
15:       $\text{loglik} \leftarrow \text{loglik} + \log\left(\frac{1}{K} \sum_{k=1}^K w^{(k)}\right)$ 
16:      Normalize  $w^{(k)} \leftarrow w^{(k)} / \sum_{j=1}^K w^{(j)}$ 
17:      Re-index  $\{\sigma^{2(k)}, w^{(k)}\}_{k=1, \dots, K}$  such that the new set of tuples is ascending in  $\sigma^{2(k)}$ 
18:      Simulate  $K$  i.i.d. standard uniforms  $U^{(1)} < \dots < U^{(K)}$ 
19:       $\pi^0 \leftarrow \frac{1}{2}w^{(1)}, \pi^K \leftarrow \frac{1}{2}w^{(K)}, s \leftarrow 1, r \leftarrow \pi^0$ 
20:      for  $j$  in  $1 : K$  do
21:        if  $U^{(j)} < \pi^0$  then
22:           $X_{i+1}^{(j)} = \sigma^{2(1)}$ 
23:        else if  $U^{(j)} > 1 - \pi^K$  then
24:           $X_{i+1}^{(j)} = \sigma^{2(K)}$ 
25:        else
26:          while  $U^{(j)} > r$  do
27:             $\pi^s = \frac{1}{2}(w^{(s)} + w^{(s+1)})$ 
28:             $r \leftarrow r + \pi^s$ 
29:             $s \leftarrow s + 1$ 
30:             $X_{i+1}^{(j)} = \sigma^{2(s-1)} + \frac{U^{(j)} - (r - \pi^{s-1})}{\pi^{s-1}}(\sigma^{2(s)} - \sigma^{2(s-1)})$ 
31:           $X_{i+1}^{(j)} \leftarrow \omega + \alpha Y_i^2 + \beta X_{i+1}^{(j)}$ 

```

2.3. *Aggregated observations.* Suppose one has as observations at time points $t_1 < \dots < t_n = N$, where $t_i \in \{1, \dots, N - 1\}$ for $i = 1, \dots, n - 1$, the aggregated value of GARCH elements between times $t_{i-1} + 1$ and t_i (with convention $t_0 \equiv 0$), that is, $\hat{y}_i = \sum_{k=1}^{n_i} y_{t_{i-1}+k}$, where $n_i = t_i - t_{i-1}$. Again, the observation times are assumed to be independent of the underlying GARCH process. Now, the problem with employing the conventional state space specification (as was used in Section 2.2) to formulate the situation in the form of (1) – (2) to approximate the likelihood of $(\hat{y}_1, \dots, \hat{y}_n)$, using SMC, is that $p(\sum_{k=1}^{n_i} y_{t_{i-1}+k} | \sigma_{t_{i-1}+1}^2)$ is not analytical for $n_i > 1$.

To overcome this, consider then the following alternative specification with three-dimensional state process:

$$(16) \quad X_i = \left(\sigma_{t_{i-1}+1}^2, \mathbb{I}(n_i > 1) \sum_{k=1}^{n_i-1} z_{t_{i-1}+k} \sigma_{t_{i-1}+k}, \sigma_{t_i}^2 \right),$$

with $Y_i = \widehat{y}_i$, $Z_i = z_{t_i}$ and $W_i = \{z_{t_{i-1}+1}, \dots, z_{t_{i-1}+n_i-1}\}$ when $n_i > 1$ ($W_i \equiv 0$ when $n_i = 1$). In this way, denoting $X_{i,j}$, $j = 1, 2, 3$ as the j th component of (16), we have $Y_i = f(X_i, Z_i) = X_{i,2} + \sqrt{X_{i,3}}Z_i$ with the analytical quantity,

$$(17) \quad p(Y_i|X_i) = \frac{1}{\sqrt{X_{i,3}}}h\left(\frac{Y_i - X_{i,2}}{\sqrt{X_{i,3}}}\right).$$

As state variable (16) is three-dimensional, this impedes immediate application of the continuous resampling procedure of Malik and Pitt (2011). To remedy this, we approximate through kernel smoothing the ECDF of $\sigma_{t_i+1}^2|Y_{1:i}$ from the particles $X_i^{(k)} = (X_{i,1}^{(k)}, X_{i,2}^{(k)}, X_{i,3}^{(k)})$ as follows:

$$(18) \quad \widehat{\mathbb{P}}(\sigma_{t_i+1}^2 \leq x|Y_{1:i}) = \frac{\sum_{k=1}^K w_i^k \mathbb{I}(\chi_i^k \leq x)}{\sum_{k=1}^K w_i^k},$$

where

$$(19) \quad \chi_i^k = \omega + \alpha(Y_i - X_{i,2}^{(k)})^2 + \beta X_{i,3}^{(k)},$$

$$(20) \quad w_i^k = \frac{\sum_{j=1}^K p(Y_i|X_i = X_i^{(j)})\phi((\chi_i^j - \chi_i^k)/c)}{\sum_{j=1}^K \phi((\chi_i^j - \chi_i^k)/c)},$$

with $\phi(\cdot)$ the standard Gaussian density and some small number c the kernel radius. The one-dimensional continuous resampling procedure of Malik and Pitt (2011) may then be applied to (18). Pseudocode detailing the methodology is presented in Algorithm 2.

3. Simulation study.

3.1. Missing observations.

3.1.1. *SMC vs indirect inference and other methods.* In this section we test the performance of the SMC estimation method for the Gaussian innovation GARCH(1, 1) process at varying degrees of missingness, also assessing its relative performance against other methods in the literature. These include the naive method of data splicing (i.e., ignoring the missing data altogether), the pseudo maximum likelihood (PML) method (Creal, Schwaab and Koopman (2014)) and the indirect inference (II) method of Blasques, Gorgi and Koopman (2021). The naive method requires no further explanation. The PML estimator $\widehat{\theta}_{\text{PML}}$ in this case is defined as the maximizer of the pseudo-likelihood function

$$(21) \quad L_{\text{pseu}}(\theta) = \prod_{i=1}^n \frac{1}{\widetilde{\sigma}_i} \phi\left(\frac{y_{t_i}}{\widetilde{\sigma}_i}\right)$$

with $\widetilde{\sigma}_i^2 = \frac{\omega}{1 - \alpha - \beta} + \left[(\omega + \alpha y_{t_{i-1}}^2 + \beta \widetilde{\sigma}_{t_{i-1}}^2) - \frac{\omega}{1 - \alpha - \beta} \right] (\alpha + \beta)^{t_i - t_{i-1} - 1}$, $i \geq 1$,

where ϕ is the standard normal density function, $t_1 = 1$ and $t_0 = 0$. The II estimator is defined as

$$\widehat{\theta}_{\text{Ind}} = \underset{\bar{\theta}}{\operatorname{argmin}} \left\| \underset{\theta}{\operatorname{argmax}} \ell_{\text{Ind}}(\theta, \bar{\theta}; K) - \widehat{\theta}_{\text{PML}} \right\|$$

with $\ell_{\text{Ind}}(\theta, \bar{\theta}; K) = \frac{1}{K} \sum_{k=1}^K \log L_{\text{pseu}}^k(\theta, \bar{\theta})$,

Algorithm 2 Smooth SMC Log-Likelihood Estimation with Aggregated GARCH

```

1: Initialization:
2:  $X_{1,1}^{1:K} \leftarrow \frac{\omega}{1-\alpha-\beta}$ 
3:  $\text{loglik} \leftarrow t_0 \leftarrow 0$ 
4: Set seed to fix randomness
5: Loop:
6: for  $i$  in  $1 : n$  do
7:    $w^{1:K} \leftarrow p^{1:K} \leftarrow X_{i,2}^{1:K} \leftarrow 0$ 
8:   for  $k$  in  $1 : K$  do
9:      $\sigma^{2(k)} = X_{i,1}^{(k)}$ 
10:    if  $n_i = t_i - t_{i-1} > 1$  then
11:      for  $j$  in  $1 : n_i - 1$  do
12:        Simulate noise variable  $z$ 
13:         $X_{i,2}^{(k)} \leftarrow X_{i,2}^{(k)} + z\sqrt{\sigma^{2(k)}}$ 
14:         $\sigma^{2(k)} \leftarrow \omega + (\alpha z^2 + \beta)\sigma^{2(k)}$ 
15:         $X_{i,3}^{(k)} = \sigma^{2(k)}$ 
16:         $p^{(k)} = \frac{1}{\sqrt{X_{i,3}^{(k)}}} h\left(\frac{Y_i - X_{i,2}^{(k)}}{\sqrt{X_{i,3}^{(k)}}}\right)$ 
17:         $\sigma^{2(k)} \leftarrow \omega + \alpha(Y_i - X_{i,2}^{(k)})^2 + \beta\sigma^{2(k)}$ 
18:       $\text{loglik} \leftarrow \text{loglik} + \log\left(\frac{1}{K} \sum_{k=1}^K p^{(k)}\right)$ 
19:      for  $k$  in  $1 : K$  do
20:        if  $n_i > 1$  then
21:           $w^{(k)} \leftarrow \sum_{j=1}^K p^{(j)} \phi\left(\frac{\sigma^{2(j)} - \sigma^{2(k)}}{c}\right) / \sum_{j=1}^K \phi\left(\frac{\sigma^{2(j)} - \sigma^{2(k)}}{c}\right)$ 
22:        else
23:           $w^{(k)} \leftarrow p^{(k)}$ 
24:        Normalize  $w^{(k)} \leftarrow w^{(k)} / \sum_{j=1}^K w^{(j)}$ 
25:        Re-index  $\{\sigma^{2(k)}, w^{(k)}\}_{k=1, \dots, K}$  such that the new set of tuples is ascending in  $\sigma^{2(k)}$ 
26:        Simulate  $K$  i.i.d. standard uniforms  $U^{(1)} < \dots < U^{(K)}$ 
27:         $\pi^0 \leftarrow \frac{1}{2}w^{(1)}, \pi^K \leftarrow \frac{1}{2}w^{(K)}, s \leftarrow 1, r \leftarrow \pi^0$ 
28:        for  $j$  in  $1 : K$  do
29:          if  $U^{(j)} < \pi^0$  then
30:             $X_{i+1,1}^{(j)} = \sigma^{2(1)}$ 
31:          else if  $U^{(j)} > 1 - \pi^K$  then
32:             $X_{i+1,1}^{(j)} = \sigma^{2(K)}$ 
33:          else
34:            while  $U^{(j)} > r$  do
35:               $\pi^s = \frac{1}{2}(w^{(s)} + w^{(s+1)})$ 
36:               $r \leftarrow r + \pi^s$ 
37:               $s \leftarrow s + 1$ 
38:             $X_{i+1,1}^{(j)} = \sigma^{2(s-1)} + \frac{U^{(j)} - (r - \pi^{s-1})}{\pi^{s-1}}(\sigma^{2(s)} - \sigma^{2(s-1)})$ 

```

where $\|\cdot\|$ is the Euclidean norm, and $L_{\text{pseu}}^k(\theta, \bar{\theta})$ is the pseudo-likelihood of θ , given in (21), with the data $\{y_{t_j}, j = 1, 2, \dots, n\}$ replaced by artificial data $\{y_{t_j}^{*i}, j = 1, \dots, n\}$. The artificial data sets here are the values at times t_1, t_2, \dots, t_n of K independently simulated complete sample paths $y_1^{*i}, y_2^{*i}, \dots, y_{t_n}^{*i}$ of the GARCH(1, 1) process with parameters $\bar{\theta} = (\bar{\omega}, \bar{\alpha}, \bar{\beta})$.

In simulating the artificial data here, the seed of the pseudo-random number generator is fixed for the numerical optimizations to work.

To see the influence of sample size on the performance of the estimators, we choose two different numbers of observations, $n = 2500$ and $n = 10,000$. In order to study the impact of different levels of missingness on the performance of the methods, we vary the amount of missing data for each sample size n . The missing data percentages used in our experiments are $m_p = 10\%$, $m_p = 20\%$, and $m_p = 30\%$. For reference, the case with complete data ($m_p = 0\%$) is also considered. For each missingness scenario, we first simulated 1000 different complete GARCH(1, 1) times series of length $N = \lfloor \frac{n}{1-m_p} \rfloor$ with standard normal innovations and parameters $\omega = 0.1$, $\alpha = 0.2$ and $\beta = 0.75$. We then randomly selected $N - n$ different time points between 1 and N (noninclusive) and deleted the values of the 1000 complete time series at these time points. Note that the random configuration of missing occurrences is simulated once for each pair of n and m_p , then held fixed across the 1000 data sets. On each resultant incomplete data set, we calculated the parameter estimates using the SMC and the other three methods. In the case with complete data, we only used the exact maximum likelihood estimation method. With the various likelihood based estimators, we also calculated the estimated variance matrix by inverting the Hessian matrix, and the standard error estimates and the 95% confidence intervals accordingly. Variance estimation for the II estimator was not considered due to lacking of an easy-to-evaluate variance estimator.

The full results are provided in Table S.1 of the Supplementary Material (Wee, Chen and Dunsmuir (2023)) with Table 1 below displaying a subset. In these and other similar tables that follow, these are the definitions of the following summary statistics: *Est*, the average of the individual estimates; *Rel Bias*, the (empirical) bias divided by the true parameter, where the bias is calculated as Est minus the true parameter value; *SE*, the standard error, calculated as the standard deviation of the individual estimates; \widehat{SE} , the average of the standard error estimates; *RMSE*, the (empirical) root mean square error; *CP*, the (empirical) coverage probability of the 95% confidence interval (CI). The empirical densities of the different estimators for each of the three parameters at each combination of n and m_p , detailed above, are shown in Figure S.1 of the Supplementary Material, with a subset of plots provided below in Figure 1.

From Table 1 and Figure 1, we make the following observations:

- The behaviour of the SMC estimators in the presence of missing data is very similar to that of the exact MLE in the case with complete data. We observe negligible biases (relative to the standard error), approximately correct coverage probabilities of the 95% CIs, and RMSEs that decrease at the correct rate when the sample size increases (approximately halving when n quadruples).
- The performance of the SMC estimators does not seem to be affected by increasing the percentage of missing data when the amount of observed data remains the same, while the naive MLE and the PML estimator suffer from larger biases when m_p increases.
- Even with as few as 250 particles, the SMC estimator is superior in terms of RMSE to the other three methods, albeit marginally, with respect to the II estimator. The SMC method appears stabilized by 250 particles with summary statistics almost unchanged when the number of particles increases to 1000.
- The naive method, while extremely fast, suffers from significantly larger biases relative to their standard errors.
- The PML estimator is also extremely fast and can estimate the parameter ω with much smaller bias than the naive estimator, although its bias is still substantially larger than those of the SMC and the II estimators.

TABLE 1

Summary of the estimates of the parameters (true values: $\omega = 0.1, \alpha = 0.2, \beta = 0.75$) by the exact MLE, the naive method, the PML (pseudo maximum likelihood method), the II (indirect inference method with $K = 250$), the SMC (sequential Monte Carlo method with $K = 250$ or 1000) on 1000 simulated data sets of size $n = 2500$ or 10,000 and missing data percentage $m_p = 0\%, 10\%,$ or 30%

		$m_p = 0$		$m_p = 0.1$				$m_p = 0.3$					
		Exact			SMC	SMC	II	Naive	PML	SMC	SMC	II	
		MLE	Naive	PML	250	1000				250	1000		
$n = 2500$	ω	Rel Bias	0.039	0.143	0.038	0.032	0.030	0.047	0.452	0.046	0.024	0.022	0.020
		SE	0.020	0.022	0.020	0.020	0.020	0.020	0.026	0.020	0.019	0.019	0.019
		\widehat{SE}/SE	0.986	0.959	0.948	0.969	0.969		0.943	0.923	0.978	0.978	
		RMSE	0.020	0.026	0.020	0.020	0.020	0.021	0.052	0.020	0.019	0.019	0.019
		CP	0.956	0.937	0.942	0.946	0.941		0.591	0.953	0.958	0.955	
		Time (s)											
	α	Rel Bias	-0.001	0.021	0.023	-0.005	-0.005	0.004	0.092	0.104	0.002	0.002	0.012
		SE	0.021	0.022	0.022	0.021	0.021	0.022	0.025	0.024	0.022	0.021	0.023
		\widehat{SE}/SE	1.010	0.977	0.980	0.990	0.991		0.940	0.929	0.977	0.983	
		RMSE	0.021	0.023	0.022	0.021	0.021	0.022	0.031	0.032	0.022	0.021	0.023
		CP	0.947	0.941	0.944	0.942	0.945		0.880	0.852	0.945	0.944	
		Time (s)											
	β	Rel Bias	-0.004	-0.016	-0.010	-0.002	-0.002	-0.005	-0.055	-0.032	-0.002	-0.002	-0.004
		SE	0.025	0.026	0.026	0.025	0.025	0.026	0.030	0.028	0.026	0.026	0.027
		\widehat{SE}/SE	0.991	0.968	0.968	0.982	0.982		0.925	0.913	0.965	0.969	
		RMSE	0.025	0.029	0.027	0.025	0.025	0.026	0.051	0.037	0.026	0.026	0.027
		CP	0.953	0.929	0.936	0.933	0.936		0.677	0.849	0.945	0.939	
		Time (s)											
$n = 10,000$	ω	Rel Bias	0.011	0.119	0.015	0.009	0.008	0.013	0.438	0.034	0.013	0.012	0.019
		SE	0.010	0.010	0.010	0.009	0.009	0.010	0.014	0.010	0.010	0.010	0.010
		\widehat{SE}/SE	0.971	0.975	0.964	0.989	0.991		0.889	0.885	0.949	0.955	
		RMSE	0.010	0.016	0.010	0.009	0.009	0.010	0.046	0.011	0.010	0.010	0.010
		CP	0.949	0.821	0.939	0.943	0.945		0.045	0.912	0.933	0.945	
		Time (s)											
	α	Rel Bias	0.001	0.023	0.026	-0.004	-0.004	-0.004	0.088	0.102	0.002	0.001	0.001
		SE	0.011	0.011	0.011	0.011	0.011	0.011	0.012	0.012	0.011	0.011	0.011
		\widehat{SE}/SE	1.001	0.963	0.965	0.984	0.985		0.940	0.931	0.981	0.981	
		RMSE	0.011	0.012	0.012	0.011	0.011	0.011	0.021	0.024	0.011	0.011	0.011
		CP	0.948	0.927	0.927	0.946	0.947		0.670	0.567	0.953	0.951	
		Time (s)											
	β	Rel Bias	-0.001	-0.015	-0.009	-0.000	0.000	-0.001	-0.054	-0.031	-0.001	-0.001	-0.002
		SE	0.012	0.013	0.013	0.013	0.012	0.013	0.015	0.014	0.013	0.013	0.014
		\widehat{SE}/SE	0.981	0.955	0.952	0.976	0.977		0.909	0.913	0.965	0.966	
		RMSE	0.012	0.017	0.014	0.012	0.012	0.013	0.043	0.027	0.013	0.013	0.014
		CP	0.951	0.855	0.919	0.939	0.938		0.175	0.566	0.938	0.934	
		Time (s)											

- In terms of bias and variance, the II estimator is very similar to the SMC estimator, although its RMSE tends to be slightly larger. Furthermore, it is significantly slower to compute. For the same number of particles K , the II method is about 50 times slower than the SMC method, even though the SMC method also computes the Hessian matrix while the II method does not.
- While the empirical densities of the estimators by all four methods seem unimodal and symmetrical and have similar spreads, only those of the SMC estimators and the II consistently peak near the true values of the parameters.
- The computational time for the SMC method can be seen to increase roughly linearly with $N \approx n/(1 - m_p)$ as well as the number of particles K . Here and hereafter, the computa-

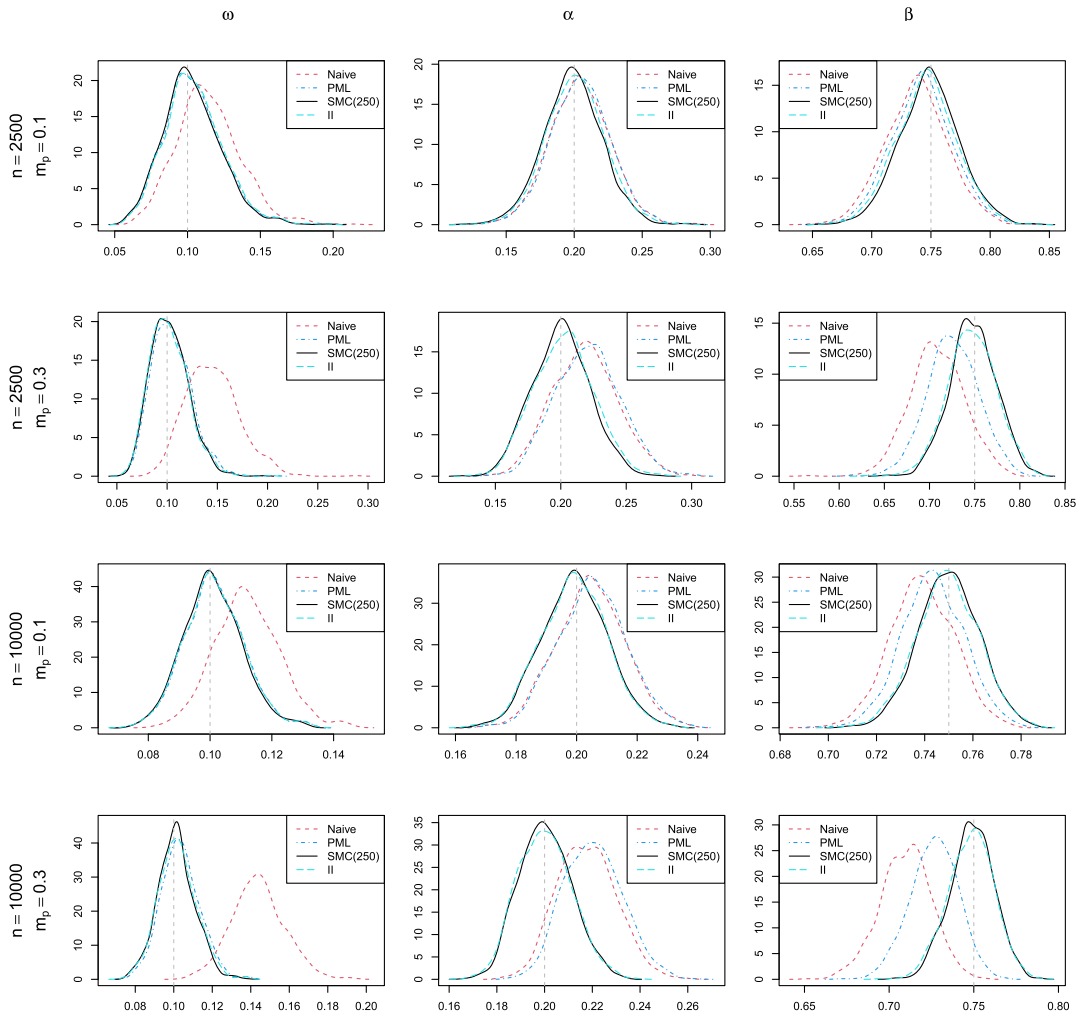


FIG. 1. Empirical densities of the naive, the PML, the SMC ($K = 250$), and the II estimators of the parameters ω , α , and β with different combinations of sample size $n \in \{2500, 10,000\}$ and missing data percentage $m_p \in \{0.1, 0.3\}$. The vertical dashed lines in the graphs indicate the respective true parameter values.

tional time of a method refers to the average running time of our code to obtain the parameter estimates on a single data set by that method on Intel Xeon Platinum 8274 3.2 GHz processors, with the objective function to be optimized in the method implemented in C++ and the numerical optimizations performed in R (R Core Team (2021)) using the `optim` function with the Nelder–Mead method (Nelder and Mead (1965)). The integration of the R and C++ code has been facilitated by the R package `Rcpp` (Eddelbuettel and François (2011)). The initial values used in the numerical optimizations are the same for all methods ($\alpha^{(0)} = \beta^{(0)} = 0.45$, $\omega^{(0)} = \frac{0.1}{n} \sum y_{it}^2$). The reported times for the likelihood based methods also include the time required to compute the Hessian matrix.

It is worth noting that, although in the reported simulation experiments we randomly generated one realization of the missingness process to use across all 1000 data sets, in unreported simulation experiments we have found that when the missingness process is regenerated for each data set, the SMC MLE works equally well and similarly outperforms the other methods considered.

3.1.2. *SMC at high levels of missingness.* We further assess the SMC method's performance at more extreme levels of missingness. Two cases are considered: only observing every third value of the series, that is, 66.7% missing, corresponding to a monthly series observed only on a quarterly basis and every fifth value, that is, 80% missing, corresponding to a daily series only observed once a trading week. The results are displayed in Table 2. We see that, in the case when only every third value is observed, even with $K = 250$ particles, the empirical biases of the estimators are all negligible relative to their respective SEs and the SE estimates are all fairly close to the empirical SEs, despite a slight downward bias. Furthermore, the empirical CPs of the 95% CIs are all 0.90 or higher and the RMSEs all halve approximately when n quadruples. Increasing the number of particles from $K = 250$ to $K = 1000$ (and quadrupling the computational cost) only sees a slight improvement in the performance of the estimators. However, in the case when only every fifth value of the series is observed, even with $K = 1000$ particles and $n = 10,000$ observations, the variances of the SMC estimators appear to be severely underestimated. Although this is largely due to a few extreme outliers among the 1000 estimates distorting the average, the tendency of the inverted Hessian to underestimate the variance seems persistent at both missingness levels, albeit less severe at the lower level.

To address this issue and to avoid the arbitrariness of the choice of the number of particles K , we experimented the following strategy with an adaptive K . Start with a relatively small number of particles K_0 (say $10 \leq K_0 \leq 100$) to get an initial SMC estimate quickly, and then keep increasing the value of K and reestimating the parameter until the estimate converges. In each reestimation step, the previous parameter estimate is used as the initial value in the optimization. The converged value is taken as the final SMC estimate $\hat{\theta}$. To estimate the Hessian matrix at $\hat{\theta}$, we first compute the Hessian matrix of the smooth SMC approximation to the negative log-likelihood at $\hat{\theta}$ with K set at the final value at convergence (using, e.g., the `optimHess` function in R) and then double the value of K and recompute the Hessian. The weighted average of these two numerical Hessians is taken as our final estimate of the Hessian, where the latter one's weight is twice that of the former. The final Hessian estimate is then inverted to estimate the variance of $\hat{\theta}$. For ease of reference, we call this method of Hessian matrix and variance estimation the *particle doubling* method.

The summary statistics of the estimation results using this strategy with an adaptive K are also included in Table 2. We see that the issue of the downward bias in variance estimation noted earlier has been largely overcome. The CPs of the 95% CIs have all improved, and the RMSEs approximately halve when n quadruples in both missingness scenarios. The average computation time has also shortened substantially for nearly all cases, even compared with the estimator with only $K = 250$ particles, suggesting in many cases using $K = 250$ particles is more than necessary. Overall, the SMC method for parameter estimation with an adaptive K value and the particle doubling method for variance estimation seem to work satisfactorily. Moreover, the deterministic mechanism of data missingness in these examples also suggests that the SMC method is not affected by nonstationarity of the missingness process.

3.2. Aggregated observations.

3.2.1. *SMC vs quasi-maximum likelihood.* This section analyzes the performance of the SMC estimation method at varying degrees of aggregation, evaluating also its relative performance against the quasi-maximum likelihood (QML) estimator of Cascone and Hotta (2019). The QML estimator constructs the quasi-likelihood by assuming the conditional distributions of the aggregated returns are Gaussian. For reference, the estimation results, using the naive MLE method (which ignores data aggregation altogether), are also included. Let us define the term *aggregation level* as $A_L = N/n$, that is, the average number of underlying GARCH observations per aggregation window. We again consider two sample sizes:

TABLE 2

Summary of the SMC parameter estimates for 1000 simulated data sets with $n = 2500$ or $n = 10,000$ observations and fairly extreme levels of missingness, where either every third ($m_p = 66.7\%$) or every fifth ($m_p = 80\%$) value of the Gaussian innovation GARCH(1, 1) series is observed. True values of the parameters: $\omega = 0.1$, $\alpha = 0.2$, $\beta = 0.75$

		Every third value observed ($m_p = 66.7\%$)									Every fifth value observed ($m_p = 80\%$)								
		$K = 250$			$K = 1000$			K adaptive			$K = 250$			$K = 1000$			K adaptive		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
$n = 2500$	Rel Bias	0.038	0.007	-0.004	0.031	0.005	-0.003	0.084	0.010	-0.007	0.021	0.011	-0.004	0.015	0.007	-0.003	0.045	0.018	-0.006
	SE	0.019	0.023	0.028	0.019	0.022	0.028	0.020	0.025	0.030	0.019	0.025	0.032	0.019	0.025	0.031	0.021	0.028	0.034
	\widehat{SE}/SE	0.890	0.946	0.919	0.922	0.985	0.964	1.101	1.071	1.146	0.808	0.851	0.811	0.864	0.910	0.877	0.980	1.003	1.015
	RMSE	0.019	0.023	0.028	0.019	0.022	0.028	0.022	0.025	0.030	0.020	0.025	0.032	0.019	0.025	0.031	0.022	0.028	0.035
	CP	0.924	0.915	0.896	0.938	0.923	0.925	0.946	0.926	0.935	0.874	0.892	0.870	0.893	0.908	0.901	0.912	0.920	0.915
	Time (s)	24.18			103.1			3.80			36.65			154.1			11.86		
$n = 10,000$	Rel Bias	0.012	0.004	-0.001	0.007	0.001	-0.001	0.020	0.004	-0.002	0.041	0.013	-0.005	0.040	0.011	-0.005	0.016	0.010	-0.003
	SE	0.009	0.012	0.014	0.009	0.012	0.014	0.009	0.012	0.014	0.054	0.015	0.037	0.055	0.021	0.042	0.009	0.013	0.016
	\widehat{SE}/SE	0.938	0.959	0.950	0.958	0.975	0.970	0.974	0.987	0.996	0.154	0.776	0.379	0.157	0.567	0.346	1.004	0.944	0.968
	RMSE	0.009	0.012	0.014	0.009	0.012	0.014	0.010	0.012	0.015	0.054	0.015	0.037	0.055	0.021	0.042	0.009	0.013	0.016
	CP	0.933	0.925	0.938	0.934	0.932	0.939	0.943	0.937	0.944	0.932	0.895	0.904	0.939	0.919	0.929	0.949	0.920	0.940
	Time (s)	93.91			400.7			54.4			143.4			609.4			158.7		

$n = 2500$ and $n = 10,000$. For each n we simulate for a given A_L , 1000 different data sets of a length $N = \lfloor nA_L \rfloor$ GARCH(1, 1) series with standard normal innovations and true parameters $\omega = 0.1$, $\alpha = 0.2$, and $\beta = 0.75$. Then for each data set, the end point of the last aggregation window is taken as $t_n = N$, with the remaining $n - 1$ aggregation window end points chosen with equal probability (without replacement) from the set $\{1, \dots, N - 1\}$. The configuration of aggregations is simulated once for each A_L and then held fixed across the 1000 data sets.

Dealing with weekday trading holidays in financial price series, one would expect to encounter $A_L < 1.05$. However, we shall analyze slightly more severe levels at 1.11, 1.25, and 1.43 (corresponding to $10/a$ for $a = 9, 8, 7$, respectively). The full results are provided in Table S.2 of the Supplementary Material with a subset displayed below in Table 3. From either table we observe the following:

- The SMC estimates are seen to be superior, even at just 250 particles, to the QML and the naive estimates in terms of negligible bias, smaller SE, and smaller RMSE. Performance of the naive and the QML estimators noticeably deteriorates as A_L increases, while the SMC performance is barely affected.
- Summary statistics between the SMC method utilizing 250 and 1000 particles are fairly similar, although they are in general marginally better utilizing 1000 particles.
- Variance estimation by simply inverting the Hessian of the SMC approximation of the negative log-likelihood or by inverting the Hessian of the negative quasi log-likelihood tends to underestimate the true variance of the respective estimator. Consequently, the coverage probabilities of the 95% CIs are noticeably lower than the nominal level of 0.95, especially when the aggregation level is higher. While this detriment is already seen to be less severe in the case of SMC, utilizing the SMC estimator with an adaptive number of particles K , together with the particle doubling method for variance estimation, improves the 95% CIs attained with SMC to roughly the correct coverage probabilities.
- When the number of observations changes from $n = 2500$ to $n = 10,000$, the RMSE of the SMC estimators roughly halve, while this is not observed for the QML estimator.
- The median computational time to obtain an estimate by a method and the corresponding standard error appears to scale approximately linearly in N and K . Note that here we report the median rather than the mean of the computation times as the computation times are highly skewed to the right, especially for the SMC method with adaptive number of particles. In practice, one can, as we have, apply a truncation in Line 21 of Algorithm 2 so that the sum is taken only over j 's such that $|\sigma^{2(j)} - \sigma^{2(k)}| \leq 3c$, reducing the otherwise K^2 operations to around linear.

Empirical densities for both the SMC (K adaptive) and QML parameter estimates are provided below in Figure 2 for $A_L \in \{1.11, 1.43\}$, with those for $A_L = 1.25$ available in Figure S.2 of the Supplementary Material. Across the board, we see that the densities of the SMC estimates peak very close to the respective true parameter values, whereas the peaks of the densities of the naive and the QML estimates are substantially farther away from the true values, especially when the level of data aggregation is higher.

3.2.2. SMC at high levels of aggregation. As mentioned, the norm encountered with financial price series is $A_L < 1.05$. However, when it comes to economic time series, one could be faced with a monthly series aggregated to quarterly ($A_L = 3$) or a daily series that is aggregated up to a trading week ($A_L = 5$). Thus, we also investigate use of our SMC method at these higher levels with the results summarized in Table 4. We see that, in terms of bias and RMSE, the estimation performance is still reasonably reliable with a trend, across all parameters, of improvements in both bias and standard error, as more particles are employed.

TABLE 3

Summary of SMC, naive MLE, and QML estimates for 1000 simulated data sets of $n = 2500$ or $n = 10,000$ observations each of an aggregated Gaussian innovation GARCH(1, 1) series at aggregation levels $A_L = 1.11$ or 1.43. True parameter values: $\omega = 0.1$, $\alpha = 0.2$, $\beta = 0.75$

		$A_L = 1.11$					$A_L = 1.43$					
		Naive	QML	SMC 250	SMC 1000	SMC Adapt	Naive	QML	SMC 250	SMC 1000	SMC Adapt	
$n = 2500$	ω Rel Bias	0.336	0.177	0.043	0.040	0.045	1.323	0.384	0.050	0.039	0.044	
		SE	0.029	0.023	0.020	0.020	0.020	0.057	0.028	0.020	0.020	0.020
	\widehat{SE}/SE	0.835	0.654	0.849	0.939	1.053	0.676	0.582	0.634	0.806	1.02	
	RMSE	0.044	0.029	0.020	0.020	0.020	0.144	0.048	0.021	0.020	0.020	
	CP	0.773	0.742	0.883	0.929	0.945	0.069	0.429	0.777	0.857	0.937	
	α	Rel Bias	0.006	0.004	0.001	0.000	0.002	0.014	0.087	0.019	0.012	0.013
		SE	0.025	0.022	0.021	0.021	0.021	0.036	0.028	0.024	0.023	0.023
		\widehat{SE}/SE	0.881	0.697	0.853	0.958	1.090	0.644	0.606	0.562	0.764	1.028
		RMSE	0.025	0.022	0.021	0.021	0.021	0.037	0.033	0.024	0.023	0.023
		CP	0.921	0.819	0.874	0.931	0.942	0.784	0.693	0.697	0.840	0.923
	β	Rel Bias	-0.014	-0.034	-0.004	-0.004	-0.004	-0.040	-0.092	-0.008	-0.006	-0.006
		SE	0.031	0.028	0.026	0.025	0.026	0.044	0.035	0.027	0.027	0.027
		\widehat{SE}/SE	0.839	0.663	0.789	0.917	1.076	0.659	0.602	0.517	0.738	1.044
		RMSE	0.033	0.038	0.026	0.026	0.026	0.053	0.077	0.028	0.027	0.027
		CP	0.896	0.666	0.823	0.920	0.937	0.734	0.192	0.619	0.805	0.923
	Time (s)	0.017	0.036	10.24	43.91	10.56	0.017	0.042	13.69	58.80	44.72	
	$n = 10,000$	ω Rel Bias	0.301	0.138	0.017	0.014	0.016	1.271	0.328	0.019	0.008	0.015
			SE	0.013	0.011	0.009	0.009	0.009	0.028	0.013	0.010	0.009
\widehat{SE}/SE		0.883	0.679	0.936	0.999	1.019	0.669	0.588	0.751	0.885	1.055	
RMSE		0.033	0.017	0.009	0.009	0.009	0.130	0.035	0.010	0.009	0.010	
CP		0.268	0.530	0.917	0.950	0.954	0.000	0.070	0.825	0.911	0.947	
α		Rel Bias	0.007	0.006	0.004	0.003	0.004	0.020	0.075	0.017	0.006	0.012
		SE	0.013	0.011	0.011	0.011	0.011	0.018	0.014	0.012	0.012	0.012
		\widehat{SE}/SE	0.845	0.674	0.896	0.969	0.995	0.649	0.606	0.669	0.809	1.009
		RMSE	0.013	0.011	0.011	0.011	0.011	0.019	0.020	0.013	0.012	0.012
		CP	0.902	0.802	0.908	0.930	0.939	0.794	0.540	0.739	0.878	0.937
β		Rel Bias	-0.011	-0.031	-0.002	-0.002	-0.002	-0.037	-0.086	-0.004	-0.001	-0.003
		SE	0.015	0.014	0.012	0.012	0.012	0.021	0.017	0.014	0.013	0.014
		\widehat{SE}/SE	0.856	0.677	0.887	0.976	1.006	0.654	0.599	0.634	0.800	1.033
		RMSE	0.017	0.027	0.012	0.012	0.012	0.035	0.066	0.014	0.013	0.014
		CP	0.859	0.334	0.900	0.937	0.943	0.505	0.000	0.705	0.863	0.930
Time (s)		0.040	0.059	41.14	169.7	66.99	0.040	0.080	55.08	227.9	120.0	

However, the downward bias in variance estimation by simply inverting the Hessian at the end of optimization is more severe. With an adaptive number of particles and the particle doubling method for variance estimation, the SMC estimator seems able to provide substantial correction for the bias in variance estimation, although the empirical coverage probabilities of the 95% CIs are still a little short of the nominal level. In practice, to estimate the variance of the SMC estimator in the presence of highly aggregated data, a bootstrap based approach might be a viable alternative.

4. Real data analysis. In this section we apply our methodology to model the daily volatility of three ETFs, listed on the Australian Stock Exchange (ASX), that provide global exposure. On ASX trading holidays, these ETFs are not priced, yet it is likely their underlying

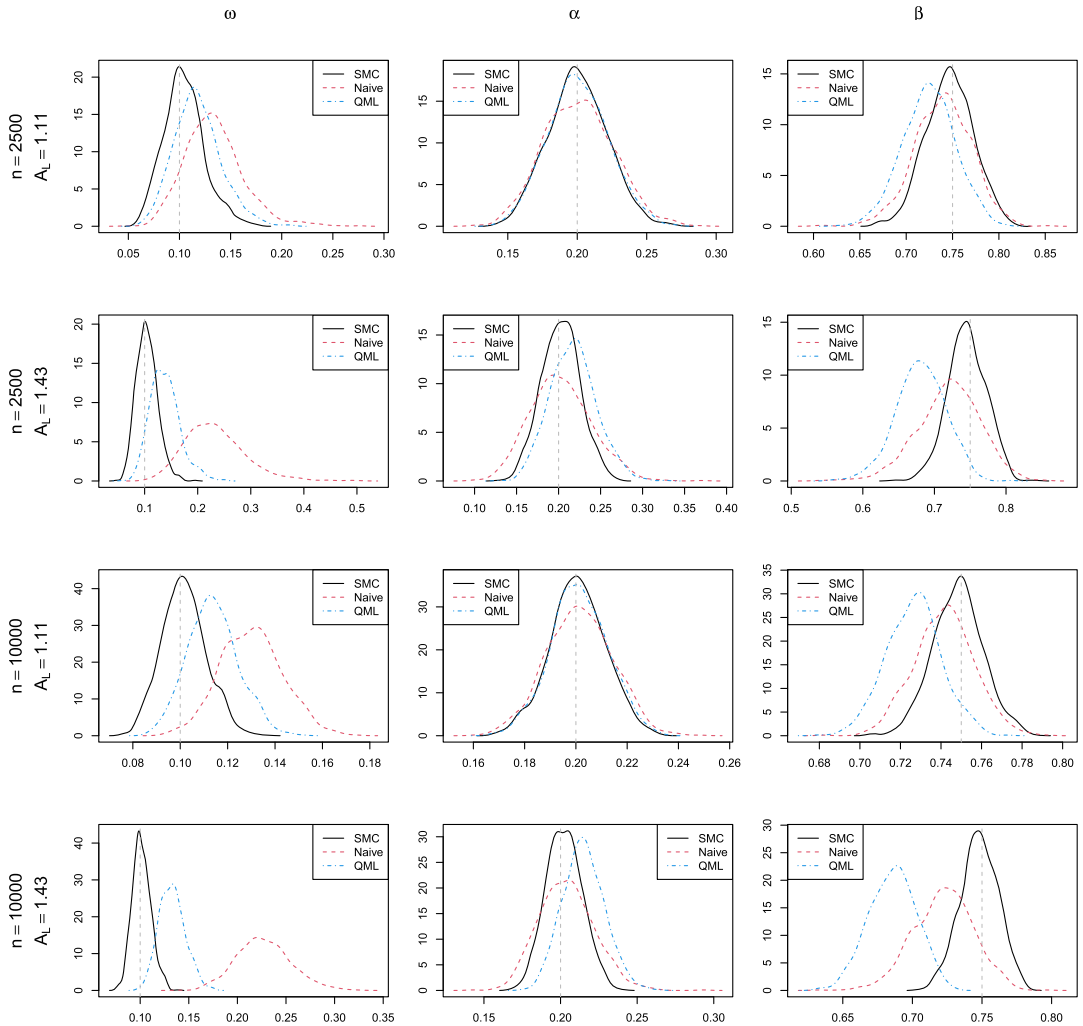


FIG. 2. Comparison between empirical densities of the SMC (K adaptive), naive ML, and QML estimates of the model parameters, based on aggregated data at different levels of aggregation. Vertical dashed lines indicate the true parameter values.

international holdings remain subject to trading activity on offshore exchanges. Ignoring this aspect, after a one-day ASX trading holiday one will not correctly account for the fact that, in reality, two days' worth of volatility has eventuated. The usual convention in studies on missing data in financial time series (cf. Bondon and Bahamonde (2012), Cascone and Hotta (2019)) is to ignore weekend gaps but recognize observations missing from trading holidays. This, therefore, is the convention we follow here.

For each ETF we examine two return series: log returns from the opening to closing price (OC returns) and log returns between successive closing prices (CC returns). When a day of pricing is missed, one has no choice but to treat OC returns as missing, whereas CC returns can be treated as aggregated. Thus, Algorithm 1 is used for the series of OC returns and Algorithm 2 employed for the series of CC returns. Bondon and Bahamonde (2012) in their application considered CC returns; however, a day of missing price information was treated by them as two consecutive CC returns missing, whereas our treatment fully incorporates all available information in the data set. Moreover the norm in empirical studies is the analysis of CC returns, although OC returns, also referred to as intraday returns, have been studied, for instance, in Kelly and Clark (2011) and Lou, Polk and Skouras (2019). Certain market

TABLE 4

Summary of the SMC parameter estimates for 1000 simulated data sets each of $n = 2500$ or $n = 10,000$ observations of a temporally aggregated Gaussian GARCH(1, 1) series at fairly high levels of aggregation ($A_L = 3$ or $A_L = 5$). True parameter values: $\omega = 0.1$, $\alpha = 0.2$, and $\beta = 0.75$

		$A_L = 3$									$A_L = 5$								
		$K = 250$			$K = 1000$			K adaptive			$K = 250$			$K = 1000$			K adaptive		
		ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
$n = 2500$	Rel Bias	0.132	0.111	-0.031	0.060	0.042	-0.013	0.045	0.026	-0.008	0.401	0.329	-0.098	0.128	0.106	-0.032	0.047	0.026	-0.009
	SE	0.025	0.033	0.038	0.021	0.027	0.033	0.021	0.027	0.032	0.047	0.058	0.072	0.029	0.037	0.046	0.021	0.027	0.034
	\widehat{SE}/SE	0.289	0.231	0.196	0.471	0.403	0.352	0.945	0.906	0.938	0.119	0.108	0.083	0.262	0.219	0.181	0.970	0.984	0.987
	RMSE	0.029	0.040	0.045	0.022	0.029	0.034	0.021	0.027	0.033	0.062	0.088	0.103	0.032	0.043	0.052	0.022	0.028	0.035
	CP	0.435	0.294	0.276	0.637	0.544	0.475	0.896	0.854	0.872	0.184	0.103	0.082	0.452	0.343	0.292	0.897	0.881	0.886
	Time (s)		27.15			117.3			338.7			38.04			167.1			2343	
$n = 10,000$	Rel Bias	0.116	0.120	-0.031	0.047	0.046	-0.013	0.043	0.041	-0.011	0.269	0.294	-0.080	0.090	0.106	-0.029	0.031	0.038	-0.010
	SE	0.012	0.016	0.019	0.010	0.013	0.016	0.010	0.015	0.017	0.027	0.035	0.043	0.013	0.020	0.023	0.011	0.016	0.019
	\widehat{SE}/SE	0.370	0.264	0.233	0.604	0.511	0.475	1.012	0.924	0.976	0.117	0.089	0.071	0.358	0.252	0.230	0.938	0.822	0.875
	RMSE	0.016	0.029	0.030	0.011	0.016	0.018	0.011	0.017	0.019	0.038	0.069	0.074	0.016	0.029	0.032	0.011	0.018	0.021
	CP	0.360	0.153	0.178	0.688	0.550	0.529	0.893	0.835	0.852	0.111	0.013	0.026	0.459	0.231	0.254	0.872	0.832	0.848
	Time (s)		109.1			464.1			951.7			153.6			662.9			6166	

TABLE 5
Frequency of interobservation times for the IVV.AX, IVE.AX, and IZZ.AX series between January 4, 2010 and December 31, 2019

Series	n_i				A_L	m_p
	1	2	3	4		
IVV.AX	2462	42	19	1	1.032	3.18%
IVE.AX	2440	53	19	1	1.037	3.61%
IZZ.AX	2447	48	20	1	1.036	3.49%

participants, for example day traders, who generally do not hold positions overnight, may find value in the analysis of OC returns.

The three ETFs analyzed are the iShares S&P 500 ETF (IVV.AX), iShares MSCI EAFE ETF (IVE.AX), and iShares China Large-Cap ETF (IZZ.AX). The study period is from January 4, 2010 to December 31, 2019 ($N = 2607$), with pricing data obtained from *Yahoo! Finance*. In this period there are a total of 81 ASX trading holidays. Additionally, there are a further two, 13, and 10 missing data points encountered, respectively, in IVV.AX, IVE.AX, and IZZ.AX on days when no volume was traded. On these days price discovery is inhibited. While no trades on the ETF were successfully executed on the day, it is quite likely that some of its underlying holdings (for which ETFs generally hold many) were subject to trading activity that day, moving the ETF's net asset value. Market mechanisms are in place to ensure an ETF's price does not deviate too much from the net asset value of its underlying holdings. Table 5 summarizes the interobservation times for each ETF series. The time series of daily close prices for the ETFs are displayed in Figure 3 with vertical bars indicating occurrence of missing price points.

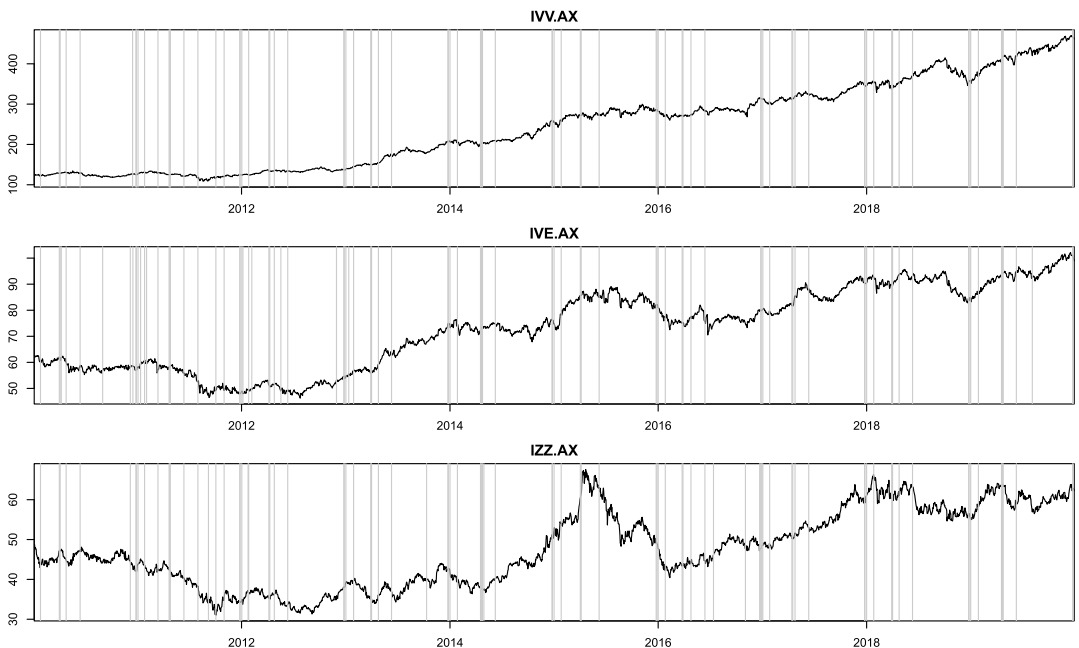


FIG. 3. *Time series of daily closes for IVV.AX, IVE.AX, and IZZ.AX from January 4, 2010 to December 31, 2019. Vertical bars indicate occurrence of missing prices.*

TABLE 6

The fitted GARCH models, along with estimated 95% confidence intervals, for the three CC returns data sets

Series	ω	α	β
IVV.AX	0.0462 (0.0257, 0.0667)	0.1081 (0.0760, 0.1402)	0.8264 (0.7737, 0.8790)
IVE.AX	0.0439 (0.0257, 0.0622)	0.1240 (0.0909, 0.1572)	0.8266 (0.7822, 0.8710)
IZZ.AX	0.0269 (0.0083, 0.0455)	0.0414 (0.0255, 0.0574)	0.9409 (0.9160, 0.9657)

For all data sets, we found utilizing the SMC algorithms with 250 and 1000 particles yielded similar parameter estimates and confidence intervals. To avoid getting caught in a local optimum solution, optimizations were initialized from 10 scattered starting points.

4.1. *CC returns.* Turning attention first to the CC returns, the parameters along with 95% confidence intervals of fitting a GARCH(1, 1) with standard normal noise to the CC returns are presented in Table 6.

As the SMC procedure does not yield point estimates of (σ_i^2) , standardized residuals cannot be calculated and investigated. As an alternative, a diagnostic to assess the fitted model, proposed in Pitt, Malik and Doucet (2014), is to consider the Rosenblatt (1952) residuals, that is, the fitted conditional distribution functions evaluated at the observed data,

$$(22) \quad F(Y_i|Y_{1:i-1}) = \int F(Y_i|X_i)\mathbb{P}(dX_i|Y_{1:i-1}) \approx \frac{1}{K} \sum_{k=1}^K F(Y_i|X_i^{(k)}) =: \hat{u}_i,$$

for $i = 1, \dots, n$. If the parameters and model are appropriate, then $\hat{u}_i \sim \text{i.i.d. } U(0, 1)$ as $K \rightarrow \infty$.

For all three CC returns series, Figure 4 displays QQ-plots of the estimated \hat{u}_i against the $U(0, 1)$ distribution. Based on these QQ-plots, the model with normal innovations seems a reasonable choice.

4.2. *OC returns.* Two GARCH(1, 1) models are fit to each OC returns series, one with Normal(0, 1) innovations and the other using Laplace(0, 1). The fits along with 95% confidence intervals are presented in Table 7. Based on the QQ-plots in Figure 5, in all series the model with Laplacian noise appears the better suited.

REMARK. We find little benefit, in our context, to provide commentary to contrast the difference in fitted volatility models between an ETF’s OC, and CC returns. Empirical studies

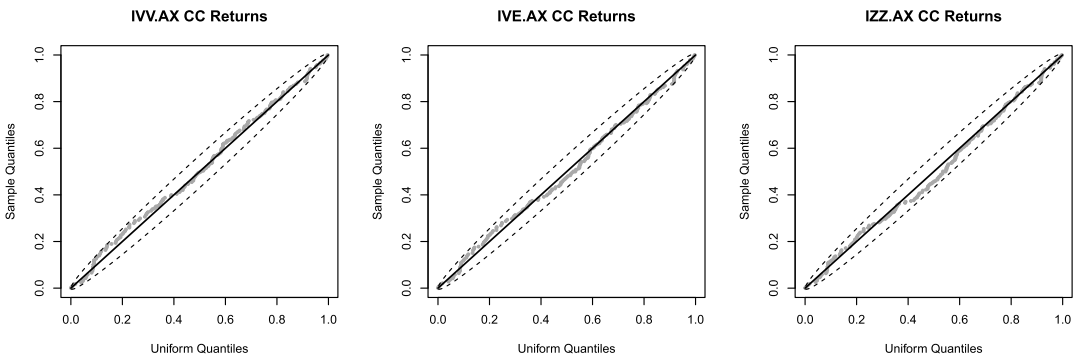


FIG. 4. QQ-plots of the estimated \hat{u}_i against the $U(0, 1)$ distribution for each CC returns series. Dashed lines indicate 95% pointwise confidence bands.

TABLE 7

The fitted GARCH models, along with estimated 95% confidence intervals, for the three OC returns data sets

Series	Noise	ω	α	β
IVV.AX	Normal	0.0409 (0.0185, 0.0634)	0.1609 (0.0975, 0.2244)	0.6794 (0.5413, 0.8174)
	Laplace	0.0227 (−0.0026, 0.0481)	0.0511 (0.0116, 0.0905)	0.6883 (0.3917, 0.9848)
IVE.AX	Normal	0.0044 (0.0024, 0.0064)	0.0232 (0.0150, 0.0315)	0.9680 (0.9583, 0.9777)
	Laplace	0.0059 (−0.0031, 0.0149)	0.0114 (−0.0033, 0.0261)	0.9418 (0.8625, 1.0212)
IZZ.AX	Normal	0.0449 (0.0161, 0.0738)	0.0385 (0.0219, 0.0551)	0.9205 (0.8836, 0.9574)
	Laplace	0.0196 (0.0032, 0.0361)	0.0167 (0.0064, 0.0269)	0.9359 (0.8957, 0.9762)

(cf. Kelly and Clark (2011), Lou, Polk and Skouras (2019)) have found behavioural differences in a financial instrument's intraday and overnight (close to open) returns; thus, different volatility dynamics would be unsurprising.

5. Concluding remarks. Fitting the GARCH(1, 1) model to time series with missing or aggregated observations is challenging due to the intractable likelihood in such cases. This paper proposes a sequential Monte Carlo method to approximate the intractable likelihood whereby the approximated likelihood as a function of the model parameters is sufficiently smooth for the purpose of numerical optimization. Our numerical experiments reveal that the resulting parameter estimates, obtained from maximizing the SMC approximated likelihood (SMC MLE), behave just as well in the presence of missing data as those from standard MLE when one has complete data. The SMC MLE method has been shown to outperform in terms

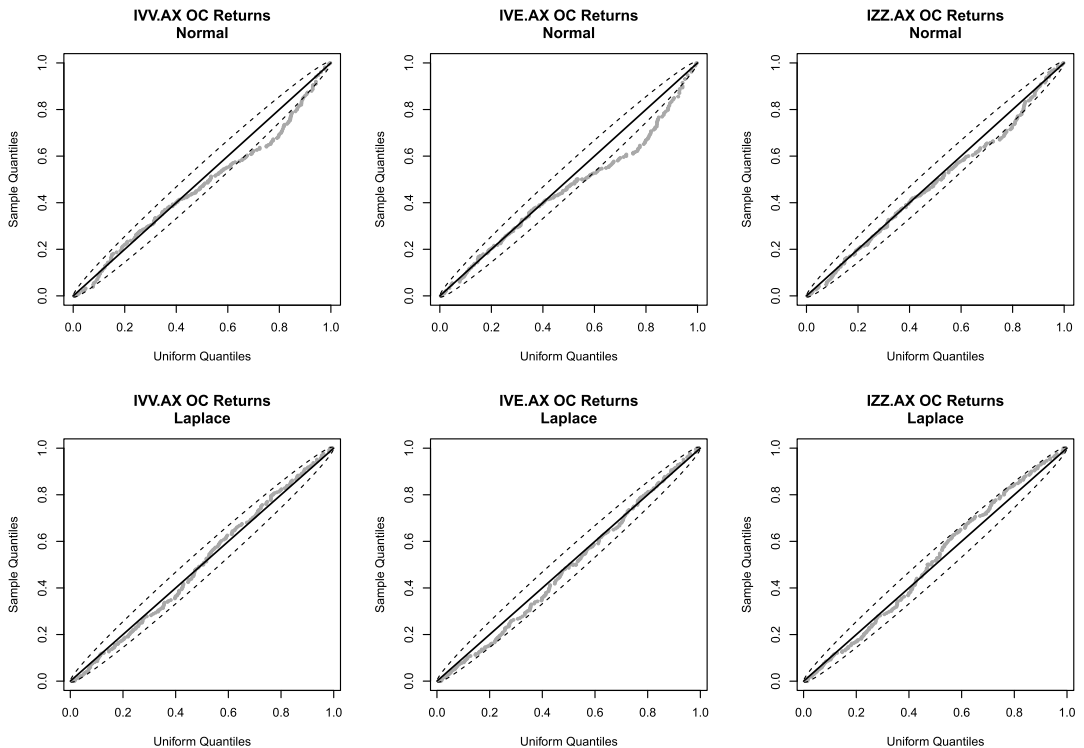


FIG. 5. *QQ-plots of the estimated \hat{u}_i against the $U(0, 1)$ distribution for each OC returns series, using both Normal and Laplace innovations. Dashed lines indicate 95% pointwise confidence bands. In all series the model with Laplacian noise appears the better suited.*

of RMSE the existing methods to handle missing or aggregated data. While this outperformance is seen to be marginal relative to the indirect inference method of Blasques, Gorgi and Koopman (2021), the SMC MLE method has the added advantage of being significantly faster to compute with easy to obtain standard error estimates.

Furthermore, it is straightforward to adapt the methodology for use with other GARCH variants such as those of Nelson (1991) and Glosten, Jagannathan and Runkle (1992). In the case of a time series suffering both from periods of missing observations and from other periods of data aggregation, a SMC solution can still be implemented. One just need to switch between using lines 7–31 in Algorithm 1, when data is missing, and lines 7–38 in Algorithm 2 when data is aggregated. While for ease of exposition we have presented the algorithms assuming independence with respect to the missingness process, there certainly is flexibility to augment the algorithms to relax this. For example, as a means to gain insights into different volatility behaviours between trading and nontrading days, one could modify the algorithms here to fit a GARCH(1, 1) series with parameters switching between one set of parameters used on trading days and another employed on nontrading days. For the SMC framework proposed here to be extended to estimate GARCH(p, q) models for either $p > 1$ or $q > 1$ (or both) with missing or aggregated observations, the development of a high-dimensional sampling method to produce a smooth likelihood surface is required which is not straightforward (Malik and Pitt (2011)). However, “*the overwhelmingly most popular GARCH models in applications has been the GARCH(1, 1) model*” (Terasvirta (2009)); see also, for instance, Hansen and Lunde (2005). Therefore, methods for maximum likelihood estimation for this model are certainly of considerable practical value.

In the complete data case, the Gaussian quasi-maximum likelihood estimator (QMLE) for the GARCH model, which simply assumes the innovations z_t to be i.i.d. standard normal irrespective of their true distribution, is known to be consistent and asymptotically normal under certain regularity conditions (see, e.g., Han and Kristensen (2014), Francq and Thieu (2019)). This suggests that, in the case with complete data, the Gaussian MLE is robust to misspecification of the innovation distribution. When faced with missing data, however, such robustness seems to be compromised. For example, when the innovations are i.i.d. standardized skew- t distributed and 10% of data is missing, our simulation experiments (results available in the Supplementary Material) show that the Gaussian MLE, obtained using the SMC method by assuming Gaussian innovations, suffers from a persistent bias when the sample size increases. This suggests that, in the case with missing data, the Gaussian MLE is not consistent when the distribution of the innovations is misspecified. However, this should be viewed as a feature of the Gaussian MLE rather than a problem of the SMC method of obtaining the estimator.

A classical approach for likelihood maximization in the presence of missing data is the EM algorithm (Dempster, Laird and Rubin (1977)). In the problem considered in this paper, however, it is not straightforward to implement the EM algorithm. The reason is that the Q -function in the E-step of the algorithm or the expectation of the complete data log-likelihood of the parameters, conditional on the observed data and the current estimate of the parameters, is also an intractable integral per se. Thus, evaluating the Q -function (exactly or approximately) and maximizing it in the M-step for the updated parameter estimate is basically as hard as evaluating the marginal log-likelihood and maximizing it for the MLE directly. Given that the EM cycle typically needs to be repeated multiple times until convergence, we see no appeal in the EM algorithm in the current context. Some EM-like algorithms, such as those considered by Gábor, Sávai and Udvari (2017), Sucarrat and Escribano (2018), and Sucarrat and Grønneberg (2022), replace certain missing variables by their estimated values and then treat them as known in calculating the likelihood contributions of the observed data. The resulting likelihood is then maximized to obtain the estimates of the parameters. Such algorithms are intrinsically different from the EM algorithm of Dempster, Laird and Rubin (1977)

in that they typically do not produce the MLE as the EM algorithm does. If they are applied to the GARCH(1, 1) model with missing data, the resultant likelihoods to be maximized are similar to the pseudo-likelihood given in (21), and the resultant estimators are expected to be inconsistent with their biases increasing with the missing data percentage, similar to the PML estimator. Therefore, we do not consider the EM algorithm or the EM-like algorithms further in this work.

The two main methods introduced in this paper, Algorithm 1 to handle missing data as arises in the OC (intraday) returns and Algorithm 2 to handle aggregated observations as arises in the interday (CC) returns, have been successfully applied to model volatility in ETFs with a large number of observations typical of financial applications. By these methods accurate estimation is achieved while properly accounting for missing or aggregated data. We also illustrated the use of (approximate) Rosenblatt residuals to assess the goodness-of-fit of the model and found that, for OC returns, the GARCH(1, 1) model with Gaussian innovations is not adequate while it is with the heavier-tailed Laplacian innovations, which we speculate is due to the phenomenon of price gapping at market open.

Acknowledgements. The comments from the editor and the reviewers have led to improved presentation of this paper for which we are grateful. This research was supported by the Australian Government through the Australian Postgraduate Award. FC is also affiliated with the UNSW Data Science Hub (uDASH). This research includes computations using resources from the National Computational Infrastructure (NCI Australia), an NCRIS enabled capability supported by the Australian Government.

SUPPLEMENTARY MATERIAL

Supplement to “Estimating GARCH(1, 1) in the presence of missing data” (DOI: [10.1214/23-AOAS1734SUPP](https://doi.org/10.1214/23-AOAS1734SUPP); .pdf). The supplementary file contains extended tables and figures giving more detailed summary of the results of our simulation experiments.

REFERENCES

- BLASQUES, F., GORGI, P. and KOOPMAN, S. J. (2021). Missing observations in observation-driven time series models. *J. Econometrics* **221** 542–568. [MR4215038 https://doi.org/10.1016/j.jeconom.2020.07.043](https://doi.org/10.1016/j.jeconom.2020.07.043)
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. [MR0853051 https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- BONDON, P. and BAHAMONDE, N. (2012). Least squares estimation of ARCH models with missing observations. *J. Time Series Anal.* **33** 880–891. [MR2991906 https://doi.org/10.1111/j.1467-9892.2012.00803.x](https://doi.org/10.1111/j.1467-9892.2012.00803.x)
- CASCONI, M. H. and HOTTA, L. K. (2019). Quasi-maximum likelihood estimation of GARCH models in the presence of missing values. *J. Stat. Comput. Simul.* **89** 292–314. [MR3882041 https://doi.org/10.1080/00949655.2018.1546860](https://doi.org/10.1080/00949655.2018.1546860)
- CREAL, D., SCHWAAB, B. and KOOPMAN, S. J. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Rev. Econ. Stat.* **96** 898–915.
- CRIPPS, E. and DUNSMUIR, W. (2003). Modeling the variability of Sydney harbor wind measurements. *J. Appl. Meteorol.* **42** 1131–1138.
- DEL MORAL, P. (1996). Nonlinear filtering: Interacting particle solution. *Markov Process. Related Fields* **2** 555–579. [MR1431187](https://doi.org/10.1080/1080009949655)
- DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. [MR0501537](https://doi.org/10.2307/2346178)
- EDDELBUETEL, D. and FRANÇOIS, R. (2011). Rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40** 1–18.
- FRANCO, C. and THIEU, L. Q. (2019). QML inference for volatility models with covariates. *Econometric Theory* **35** 37–72. [MR3904171 https://doi.org/10.1017/S0266466617000512](https://doi.org/10.1017/S0266466617000512)
- GÁBOR, D. K., SÁVAI, M. and UDVARI, B. (2017). Missing data bias on a selective hedging strategy. *J. Compet.* **9** 15–19.
- GLOSTEN, L., JAGANNATHAN, R. and RUNKLE, D. (1992). On the relation between the expected value and the volatility of nominal excess return on stocks. *J. Finance* **46** 1779–1801.

- GORDON, N. J., SALMOND, D. J. and SMITH, A. F. (1993). A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE Proc. F* **140** 107–113.
- HAN, H. and KRISTENSEN, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *J. Bus. Econom. Statist.* **32** 416–429. MR3238595 <https://doi.org/10.1080/07350015.2014.897954>
- HANSEN, P. R. and LUNDE, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1, 1)? *J. Appl. Econometrics* **20** 873–889. MR2223415 <https://doi.org/10.1002/jae.800>
- HIGGINS, P. (2014). GDPNow: A model for GDP “nowcasting”. Federal Reserve Bank of Atlanta: Working Paper Series.
- KANTAS, N., DOUCET, A., SINGH, S. and MACIEJOWSKI, J. (2009). An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. *IFAC Proc. Vol.* **42** 774–785.
- KELLY, M. and CLARK, S. (2011). Returns in trading versus non-trading hours: The difference is day and night. *J. Asset Manag.* **12** 132–145.
- LOU, D., POLK, C. and SKOURAS, S. (2019). A tug of war: Overnight versus intraday expected returns. *J. Financ. Econ.* **134** 192–213.
- MALIK, S. and PITT, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *J. Econometrics* **165** 190–209. MR2846644 <https://doi.org/10.1016/j.jeconom.2011.07.006>
- NELDER, J. A. and MEAD, R. (1965). A simplex method for function minimization. *Comput. J.* **7** 308–313. MR3363409 <https://doi.org/10.1093/comjnl/7.4.308>
- NELSON, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59** 347–370. MR1097532 <https://doi.org/10.2307/2938260>
- OSSANDÓN, S. and BAHAMONDE, N. (2013). A new nonlinear formulation for GARCH models. *C. R. Math. Acad. Sci. Paris* **351** 235–239. MR3089685 <https://doi.org/10.1016/j.crma.2013.02.014>
- PITT, M. K., MALIK, S. and DOUCET, A. (2014). Simulated likelihood inference for stochastic volatility models using continuous particle filtering. *Ann. Inst. Statist. Math.* **66** 527–552. MR3211873 <https://doi.org/10.1007/s10463-014-0456-y>
- ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.* **23** 470–472. MR0049525 <https://doi.org/10.1214/aoms/1177729394>
- SUCARRAT, G. and ESCRIBANO, A. (2018). Estimation of log-GARCH models in the presence of zero returns. *Eur. J. Finance* **24** 809–827.
- SUCARRAT, G. and GRØNNEBERG, S. (2022). Risk estimation with a time-varying probability of zero returns. *J. Financ. Econom.* **20** 278–309.
- R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- TERASVIRTA, T. (2009). Introduction to univariate GARCH models. In *Handbook of Financial Time Series* 17–42. Springer, Berlin.
- WEE, D. C., CHEN, F. and DUNSMUIR, W. T. (2023). Supplement to “Estimating GARCH(1, 1) in the presence of missing data.” <https://doi.org/10.1214/23-AOAS1734SUPP>