# Continuous-time Markov-switching GARCH Process with Robust State Path Identification and Volatility Estimation

Yinan Li[1] and Fang Liu[1][✉][0000−0003−3028−5927]

University of Notre Dame, Notre Dame, IN 46530, U.S.A.
Yinan.Li.267@alumni.nd.edu
fang.liu.131@nd.edu

**Abstract.** We propose a continuous-time Markov-switching generalized autoregressive conditional heteroskedasticity (COMS-GARCH) process for handling irregularly spaced time series with multiple volatility states. We employ a Gibbs sampler in the Bayesian framework to estimate the COMS-GARCH model parameters, the latent state path and volatilities. To improve the computational efficiency and robustness of the identified state path and estimated volatilities, we propose a multi-path sampling scheme and incorporate the Bernoulli noise injection in the computational procedure. We provide theoretical justifications for the improved stability and robustness with the Bernoulli noise injection through the concept of ensemble learning and the low sensitivity of the objective function to external perturbation in the time series. The experiment results demonstrate that our proposed COMS-GARCH process and computational procedure are able to predict volatility regimes and volatilities in a time series with satisfactory accuracy.

**Keywords:** Bernoulli noise injection · ensemble learning · Gibbs sampler · irregularly (unevenly) spaced time series · maximum a posterior (MAP) · stability and robustness

## 1 Introduction

### 1.1 Motivation and Problem

Heteroskedasticity is a common issue in time series (TS) data. The generalized autoregressive conditional heteroskedasticity (GARCH) model is a popular discrete-time TS model that accommodates heteroskeasticity and estimates the underlying stochastic volatility. When there is variation in the presence of regime changes in the volatility dynamics, the Markov-switching GARCH (MS-GARCH) model can be employed and when collected TS data are irregularly spaced in time, continuous-time GARCH (CO-GARCH) can be applied.

In practice, TS data may exhibit heteroskedasticity and multiple regimes and are collected in irregularly spaced timepoints or on a near-continuous time scale. For example, heart rate variability TS are typically recorded on a millisecond scale and can have multiple volatility regimes corresponding to different activities or stress levels. Seismic waves TS, used for studying the earth's interior structure and predicting earthquakes, consist of wave types of different magnitudes and are recorded on a millisecond scale. Financial data are often irregularly spaced in time due to weekend and holiday effects and known to exhibit different volatility states, such as changing behavior drastically from steadily trending to extremely volatile after a major event or news.

To the best of our knowledge, there does not exist a MS-GARCH model for irregularly spaced TS nor a CO-GARCH model to handle multiple states. To fill the methodological gap and respond to the practical needs to handle irregularly-spaced and multi-state TS data, we propose a new COMS-GARCH process to analyze such data and develop a robust and efficient computational procedure to identify multiple states and estimate volatilities.

### 1.2 Related Work

The GARCH process has been extensively studied [13, 25, 26, 28, 32]. [27] derives the conditions under which the discretized-time GARCH model converges in distribution to a bivariate non-degenerate diffusion process as the length of the discrete-time intervals goes to zero. The fact that the limiting process consisting of two independent Brownian motions (that drives the underlying volatility process and the accumulated TS, respectively) contradicts the GARCH model's intuition that large volatilities are the feedback of large innovations. [10] applies different parameterizations as a function of the discrete-time interval to GARCH(1; 1) and obtained both degenerate and non-degenerate diffusion limits. [31] further shows the asymptotic non-equivalence between the GARCH model and the continuous-time bivariate diffusion limit except for the degenerate case. [18] proposes a COntinuous-time GARCH (CO-GARCH) model that replaces the Brownian motions by a single Lévy process and incorporates the feedback mechanism by modeling the squared innovation as the quadratic variation of the Lévy process. For the inference of the CO-GARCH process, there exist quasi-likelihood [7], method of moments (MoM) [18], pseudo-likelihood [20, 21] approaches, and Markov chain Monte Carlo (MCMC) procedures [24].

Volatility predictions by GARCH-type models may fail to capture the true variation in the presence of regime changes in the volatility dynamics [3, 19, 23]. The MS-GARCH model [14] solves this issue by employing a hidden discrete Markov chain to assign a state to each timepoint. For the inference in the MS-GARCH model, regular likelihood-based approaches require summing over exponentially many possible paths and can be computationally unfeasible. Several alternatives exist to deal with the problem. The collapsing procedures [11, 14, 16, 17] use simplified versions of the MS-GARCH model and incorporate recombination mechanisms of the state space. [15] proposes a new MS-GARCH model that is analytically tractable and allows the derivation of stationarity condition and the process properties. [2] employs a Markov Chain Expectation Maximization approach. [5] proposes a Bayesian MCMC method but it can be slow in convergence. Recent methods focus on the efficient sampling of state paths. [12] introduces a Viterbi-based technique to sample state paths; [4] proposes a particle MCMC algorithm; [6] uses a multi-point sampler in combination with the forward filtering backward sampling technique. Both the likelihood-based and MCMC estimations have been implemented in software (e.g., R package `MSGARCH` [1]).

For multi-state irregularly-spaced TS, simple and model-free approaches such as the realized volatility can be used to estimate the historical volatility, but they cannot systematically identify different volatility states. In addition, it is not always meaningful to aggregate measures across timepoints to calculate the

realized volatilities. Though the existing MS-GARCH model can also be used to identify different volatility regimens, it cannot analyze irregularly spaced TS. This methodological gap motives our work.

### 1.3 Our Contribution

We propose a COMS-GARCH process, employing the Lévy process to model volatility in each state and the continuous-time hidden Markov chain to model state switching. The estimate volatilities via the COMS-GARCH process are expected to be more robust than those obtained via the realized volatility, largely due to the computational procedure we design specifically for the state path and volatility estimation. Furthermore, the COMS-GARCH process can be used to forecast volatilities and states. Other contributions are listed below.

- We propose a Bayesian Gibbs sampler to obtain inference of the COMS-GARCH parameters and maximum a posterior (MAP) estimation for state path and volatilities.
- We develop a computational procedure with a multi-path sampling scheme and the Bernoulli noise injection (NI) to accelerate the optimization and improve the robustness of the predicted state path and volatilities.
- We provide theoretical justifications for the Bernoulli NI from the perspectives of ensemble learning of the state path and lowered sensitivity of the objective function to small random external perturbation in the TS.
- We run experiments in both simulated and real TS data to demonstrate the application of the COMS-GARCH procedures and the computational procedure and to show satisfactory accuracy in predicting states and volatilities.

## 2 COMS-GARCH Process

We develop the COMS-GARCH process to handle multiple states, extending the CO-GARCH(1,1) process [18, 20]. To the best of our knowledge, CO-GARCH$(1;1)$ is the only Lévy-process driven CO-GARCH model that has analytical solutions for the model parameters from the stochastic differential equations and is inference-capable in the context of pseudo-likelihood. [9] theoretically analyzes the CO-GARCH$(p;q)$ model driven by the Lévy process for general $p$ and $q$ values, but unable to obtain inferences for the model parameters.

Let $G_t$ for $t \ 2 \ (0;T)$ denote the observed TS, $L$ refer to the innovation that is modeled by a Lévy process, $s_t$ be the state $fs_tg$ and $\ ^2_t$ is the underlying volatility process at time $t$. Our proposed COMS-GARCH process on $(G; \ ^2; S) = (fG_tg; f \ ^2_t g; fs_tg)$ is the solution to the following set of stochastic differential equations

$$\begin{cases} dG_t = Y_t = \ _t dL_t(s_t) & (1) \\ d \ ^2_t = \ (s_t)dt \quad (s_t) \ ^2_{t-}dt + \ (s_t) \ ^2_{t-}d[L;L]_t & (2) \\ \Pr(s_t = jjs_{t-} = k) = \ _{jk}dt + o(dt) \text{ for } j \ 6= k & (3) \\ \Pr(s_t = kjs_{t-} = k) = 1 \quad \sum_{j \neq k} \ _{jk}dt + o(dt): & (4) \end{cases}$$

$t$ stands for $t \quad dt$ and $( \ ^2_{t-}; s_{t-})$ refers to the volatility and associated state at time $t \ . \ ; \ ,$ and in Eq (2) quantify how much the change in time $(dt)$,

the volatility at time $t$ , and the innovation at time $t$ affects the volatility, respectively. The increment of the Lévy process $dL_t(s_t)$ in Eq (1) is assumed standardized with mean 0 and variance 1, and $[L; L]_{t-}$ in Eq (2) is its quadratic variation process. Eqs (3) and (4) represent the hidden continuous-time Markov chain with discrete states and transition parameters $= f_{jk}g$ that model the regime switching in the TS for $j; k \; 2 \; f1;$ ; $g$ (note that $_{jk}$ in Eqs (7) and (8) is not a probability, and the parameter space for $_{jk}$ is $(0; 1)$ instead of $2$ $(0; 1)$).

Next, we define a family of discrete-time processes that approximates the above continuous-time process $(G; \ ^2; S)$, following the methodological framework in [20]. There are a couple of reasons for obtaining a discretized process. First, real-life observed TS data are often recorded in discrete-time, whether irregularly spaced or regardless of how fine the time scale is. Second, the discretization allows us to take advantage of the well-developed inferential approaches for discrete-time GARCH processes. We will show the discretized process converges to the COMS-GARCH process.

The discretization is defined over a finite time interval $[0; T]$ for $T > 0$. Let $0 = t_0 < t_1 <$ $< t_i <$ $< t_n = T$ be a deterministic sequence that divides $[0; T]$ into $n$ sub-intervals of lengths $t_i = t_i$ $t_{i-1}$ for integers $i = 1; : : : ; n$. Let $G_0 = 0$ and $_i$ be a first-jump approximation of the Lévy process [20]. A discretized COMS-GARCH process $(G_n; \ ^2_n; s_n) = (fG_ig; f \ ^2_ig; fs_ig)$ satisfies

$$\begin{cases} G_i \quad G_{i-1} = Y_i = \ _{i-1}( \ t_i)^{1=2} \ _i; & (5) \\ \ ^2_i = \ (s_i) \ t_i + \ ^2_{i-1} + \ (s_i)Y_i^2 \ \exp( \ (s_i) \ t_i); & (6) \\ \Pr(s_i = jjs_{i-1} = k) = 1 \ \exp( \ _{jk} \ t_i) \text{ for } j \ 6= k; & (7) \\ \Pr(s_i = kjs_{i-1} = k) = \ \sum_{j \neq k} \exp( \ _{jk} \ t_i): & (8) \end{cases}$$

Eqs (5) and (6) model the "CO" component in a similar manner to [20] and Eqs (7) and (8) model the "MS" part of the COMS-GARCH process. Since $Y_i$ is obtained by differencing the observed $G_i$, it is also observed. To ensure the positivity of Eq (6), we require $(k)$ and $(k)$ to be non-negative for all states $k = 1; : : : ;$ . To reflect the general belief that dependence between two quantities at two timepoints diminishes as the time gap increases, we also impose positivity on $(k)$ for all states. As $n \ ! \ 1$ , $t_i \ ! \ 0$ and the discretized COMS-GARCH process in Eqs (5) to (8) converges in probability to the COMS-GARCH process defined in Eqs (1) to (4), as stated in Lemma 1.

**Lemma 1** (Convergence of discretized COMS-GARCH process)**.** Let $(G; \ ^2; s)$ be the COMS-GARCH process on time interval $[0; T]$, and $(G_n; \ ^2_n; s_n)$ be its discretized process. As $n \ ! \ 1$ , $t_i \ ! \ 0$ for $i = 1; : : : ; n$ and $(G_n; \ ^2_n; s_n)$ converges in probability to $(G; \ ^2; s)$ in that the Skorokhod distance $D_S((G_n; \ ^2_n, S_n); (G; \ ^2; s)) \ \overset{p}{!} \ 0$ as $n \ ! \ 1$ .

The converges in probability in Lemma 1 also implies the convergence of $(G_n; \ ^2_n, s_n)$ in distribution to $(G; \ ^2; s)$. Lemma 1 is an extension of the theorem on the convergence of a discretized CO-GARCH process [20]. The added complexity in COMS-GARCH, that is, multiple states and state-dependent GARCH parameters, has no material impact on the discretization of the process and the underlying conditions that leads to the convergence. Therefore, the theoretical result of the

convergence of the discretized CO-GARCH process can be directly extended to the discretized COMS-GARCH process. In fact, the CO-GARCH model can be regarded as a special case of the COMS-GARCH process when the number of states is 1. Therefore, the inferential approaches and theoretical results for COMS-GARCH in the next section also apply to CO-GARCH.

## 3 Inference for COMS-GARCH Process

The parameters in the COMS-GARCH process include $= f\ (k);\ (k);\ (k)g\ 8\ k = 1;:::;$ and transition parameters . In addition, we are also interested in learning the latent state $s_i$ and volatility $^2_i$ for $i = 1;:::;n$ so to better understand an observed TS and to aid prediction of future states and volatilities. We propose a Bayesian Gibbs sampler coupled with the pseudo-likelihood that is defined as follows.

$$f(Y_i jY_1;:::;Y_{i-1};s_1;:::;s_i) = N(0;\ ^2_i);\ \text{where} \tag{9}$$

$$^2_i = \ ^2_{i-1}\ \frac{(s_i)}{(s_i)\quad(s_i)}\quad\frac{\exp((\ (s_i)\quad(s_i))\ \ t_i)\ \ 1}{(s_i)\quad(s_i)}\ +\frac{(s_i)\ \ t_i}{(s_i)\quad(s_i)} \tag{10}$$

$$^2_{i-1}\ \ t_i$$

$$= \ (s_{i-1})\ \ t_{i-1}\ \ t_i + \ \ t_i\ \ ^2_{i-2} + \ (s_{i-1})Y^2_{i-1}\ \ e^{-\ (s_i\ \ 1)\ \ t_i\ \ 1)}: \tag{11}$$

Eq (11) is obtained by taking the first-order Taylor expansion of Eq (10) around $t_i = 0$ and substituting $^2_{i-1}$ in Eq (6). Eqs (9) to (11) suggest that

$$\mathrm{E}(Y^2_i jY_1;:::;Y_{i-1};s_1;:::;s_i) = \mathrm{V}(Y_i jY_1;:::;Y_{i-1};s_1;:::;s_i) = \ ^2_i\quad ^2_{i-1}\ \ t_i: \tag{12}$$

### 3.1 Gibbs Sampler for Bayesian Inference on Model Parameters

Since we are interested in obtaining inferences for and given the pseudo-likelihood and predicting states and volatilities, methods that rely on integrating out the latent states, such as the EM and MC-EM algorithms (we provide their steps in the expanded paper for interested readers), do not work well. In contrast, the Bayesian framework provides a more convenient and straightforward approach to reach the inferential goal. Below we propose a Gibbs sampler to obtain Bayesian inferences for the COMS-GARCH process.

Define $t = (\ t_1;:::;\ t_n);\ Y = (Y_1;:::;Y_n);\ S = (s_1;:::;S_n)$, and $S$ is the set of all possible state paths. Denote the priors for and by $(\ ;\ )$ and assume $(\ ;\ ) = (\ )\ (\ )$. The conditional posterior distributions of $;$ , and the states are respectively

$$f(\ j\ ;Y;\ \ t;S) \propto (\ )L(\ ;\ jY;S) = (\ )\ \prod_{i=1}^{n}\ ^{-1}_i \exp\ \ Y^2_i=(2\ ^2_i)\ ; \tag{13}$$

$$f(\ _{1k};:::;\ \ _k j\ ;Y;\ \ t;S) = f(\ _{1k};:::;\ \ _k jS;\ \ t) \text{ for } k = 1;:::;\ ;$$

$$\propto (\ _{1k};:::;\ \ _k)\ \prod_{s_{i+1}=k,\ s_i=k}^{n-1}\Bigg(2\ +\ \sum_{j\neq k}\ e^{-\ jk\ \ t_{i+1}}\Bigg)\ \prod_{j\neq k}\ \prod_{s_{i+1}=j,\ s_i=k}^{n-1}(1\ \ e^{-\ jk\ \ t_{i+1}}); \tag{14}$$

$$f(s_i jS_{-i};\ \ ;\ ;Y;\ \ t) \propto \ _{s_i;s_{i\ \ 1}}\ \ _{s_{i+1};s_i}\ \prod_{t=i}^{n}\ ^{-1}_t \exp\ \ Y^2_t=(2\ ^2_t)\ ; \tag{15}$$

where $\sum_{j=1}\ \ _{jk} = 1$ in Eq (14), and

$$_{s_i;s_{i\ 1}} = \begin{cases} (\ 2 \quad + \sum_{k\neq s_{i\ 1}} \exp(\quad _{k;s_{i\ 1}} \quad t_i) \text{ when } s_i = s_{i-1} \\ 1 \quad \exp(\quad _{s_i;s_{i\ 1}} \quad t_i) \text{ when } s_i \neq s_{i-1} \end{cases};$$

similarly for $_{s_{i+1};s_i}$. When there are two states ( $= 2$), Eqs (14) and (15) can be simplified to

$$f(\ _{21}jS;Y; \quad t) \ / \ (\ _{21}) \prod_{s_{i+1}=1}^{n-1} e^{-\ _{21}\ t_i} \prod_{\substack{s_{i+1}=2 \\ s_i=1}}^{n-1} (1 \quad e^{-\ _{21}\ t_{i+1}}) \tag{16}$$

$$f(\ _{12}jS;Y; \quad t) \ / \ (\ _{12}) \prod_{\substack{s_{i+1}=2 \\ s_i=2}}^{n-1} e^{-\ _{12}\ t_i} \prod_{\substack{s_{i+1}=1 \\ s_i=2}}^{n-1} (1 \quad e^{-\ _{12}\ t_{i+1}}) \tag{17}$$

$$f(s_i jS_{-i}; \quad ; \ ;; \quad t) \ / \ _{1;s_{i\ 1}}^{2-s_i}\ _{2;s_{i\ 1}}^{s_i-1}\ _{1;s_i}^{2-s_{i+1}}\ _{2;s_i}^{s_{i+1}-1} \prod_{t=i}^{n} \ ^{-1}_{t} \exp\left(\frac{Y_t^2}{2\ ^2_t}\right); \tag{18}$$

where $_{1;s_{i\ 1}} = e^{-\ _{21}\ t_i}$ if $s_{i-1} = 1$, and $1 \quad e^{-\ _{21}\ t_i}$ if $s_{i-1} = 2$; $_{1;s_i} = e^{-\ _{21}\ t_{i+1}}$ if $s_i = 1$, and $1 \quad e^{-\ _{21}\ t_{i+1}}$ if $s_i = 2$.

The Gibbs sampler draws samples on ; and $s_i$ for $i = 1;:::;n$ alternatively from Eqs (13), (14), and (15). Upon convergence, after burning and thinning, we will have multiple, say $M$, sets of posterior samples of ; , based on which their posterior inferences can be obtained. We will also have $M$ sets of samples on state $s_i$ and can calculate the posterior volatility $^2_i$ at each timepoint via Eq (6). Connecting the states across the $n$ times points from each set of the state posterior samples leads to a state path. Due to the large sample space (totally $^n$ possible paths), it is difficult to identify the MAP estimate for the state path out of the $M$ paths with acceptable accuracy unless $M \gg \ ^n$ and a significant portion of paths have close-to-0 posterior probabilities with a few paths having significantly higher probabilities compared to the rest. To solve this issue, we design a new computational algorithm (reSAVE) as detailed next.

### 3.2 Estimation for State Path and Volatility

To deal with the computational challenge in obtaining the MAP estimates for state path via the Gibbs sampler in Sec 3.1, we propose an inferentially **R**obust and computationally **E**fficient iterative procedure for **S**tate path **A**nd **V**olatility **E**stimation (reSAVE). The steps of the procedure are listed in Alg. 1.

The inferential robustness for the MAP estimates from the reSAVE procedure is brought by the Bernoulli NI implemented in each iteration of the procedure, leading to both ensemble learning and improved stability of the object function from which the MAP estimates are obtained (Sec 3.4). The computational efficiency of the reSAVE can be attributed to a couple of factors: the Bernoulli NI that generates a sub-TS (smaller data size) in each iteration, of sampling of a small set of state $m$ to calculate MAP estimates for the state path, and the employment of a maximization-maximization scheme to obtain the MAP estimates of the parameters and and for the state path in each iteration.

The number of iterations $N$ in the procedure can be prespecified or determined using a convergence criterion, such as the $l_1$ distances on the MAP estimates (e.g., $fj\hat{}^{(l+1)} \quad \hat{}^{(l)}j; j\hat{}^{(l+1)} \quad \hat{}^{(l)}; jS^{(l+1)} \quad S^{(l)}jg$) or the objective functions between two consecutive iterations. If the criterion reaches a prespecified threshold, the procedure stops. Regarding the number of sampled state paths $m > 1$, $m$ too small will not lead to stable MAP estimates; $m$ too large would increase

the computational costs. In the experiments presented in Sec 4, we used $m = 6$, which seems good enough. The ensemble size $b$ refers to the number of observations following a given timepoint $i$ that are used to update the conditional posterior distribution of $s_i$ (Proposition 1). The specification of $b$ is mainly for computational efficiency consideration and is optional.

**Algorithm 1:** The reSAVE Optimization Procedure

**input :** data $(\mathbf{Y}; \mathbf{t})$, initial values $\theta^{(0)}; \boldsymbol{\eta}^{(0)}; S^{(0)} = s_1^{(0)}, \ldots, s_n^{(0)}$, # of iterations $N$, # of sampled state paths $m$, ensemble size $b$

**output :** MAP estimates $\hat{S}_{\text{MAP}}; \hat{\sigma}^2_{\text{MAP}} = (\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2); \hat{\theta}_{\text{MAP}}; \hat{\boldsymbol{\eta}}_{\text{MAP}}$.

**for** $l = 1$ *to* $N$ **do**
  Apply Bernoulli NI in Alg. 2 to obtain a sub-TS $(\mathbf{Y}^{(l)}; \mathbf{t}^{(l)})$ of length $\tilde{n}^{(l)}$. Denote by $T^{(l)}$ the set of the original timepoints retained in the sub-TS;
  Calculate $\hat{\theta}^{(l)}_{\text{MAP}} = \arg\max f(\theta | S^{(l-1)}; \mathbf{Y}^{(l)}; \mathbf{t}^{(l)})$ and
  $$\hat{\boldsymbol{\eta}}^{(l)}_{\text{MAP}} = \arg\max_{\boldsymbol{\eta}} f(\boldsymbol{\eta} | S^{(l-1)}; \mathbf{Y}^{(l)}; \mathbf{t}^{(l)});$$
  **for** $j = 1$ *to* $m$ **do**
    **for** $i = 2$ *to* $n - 1$ **do**
      sample $s_i^{(j)}$ for $i \in T^{(l)}$ given $\theta^{(l)}_{\text{MAP}}; \boldsymbol{\eta}^{(l)}_{\text{MAP}}; \mathbf{Y}^{(l)}; \mathbf{t}^{(l)}; S^{*(l-1)}_{-i}$ per Eq (15);
    **end**
    Let $S^{(j)} = s_1^{(j)}, s_2^{(j)}, \ldots, s_{\tilde{n}^{(l)}}^{(j)}$.
  **end**
  Let $\mathcal{S} = S^{(1)}, \ldots, S^{(m)}$, $S^{(l)} = \arg\max_{S \in \mathcal{S}} f(S | \theta^{(l)}_{\text{MAP}}; \boldsymbol{\eta}^{(l)}_{\text{MAP}}; \mathbf{Y}^{(l)}; \mathbf{t}^{(l)})$, and $S^{(l)} = \{s_i^{(l)}\}_{i \in T^{(l)}} \cup \{s_i^{(l-1)}\}_{i \notin T^{(l)}}$;
**end**
Calculate MAP estimate for volatility $\{\sigma_i^2\}_{i=1,\ldots,n}$ given $S_{\text{MAP}} = S^{(N)}; \hat{\theta}_{\text{MAP}} = \hat{\theta}^{(N)}_{\text{MAP}}, \hat{\boldsymbol{\eta}}_{\text{MAP}} = \hat{\boldsymbol{\eta}}^{(N)}_{\text{MAP}}$ via Eq (6).

The MAP estimates of $\theta$ and $\boldsymbol{\eta}$ can be obtained either through direct optimization of their respective conditional posterior distributions or via MC approaches using samples from the conditional posterior distributions. The estimate of the state path in each iteration is defined as the path, out of the sampled $m$ paths, that maximizes the conditional posterior distribution of $S$, which is proportional to $\prod_{i=1}^{n} \sigma_i^{-1} \exp(-y_i^2/(2\sigma_i^2)) \, \pi_{i,i-1}$ [5], given the latest MAP estimates of $\theta$ and $\boldsymbol{\eta}$. Though the Bernoulli NI is designed more for achieving ensemble learning and improving the stability of the objective functions for the state path optimization (see Sec 3.4), we expect its usage also makes the inference for $\theta$ and $\boldsymbol{\eta}$ more robust, especially if $n$ is relatively large.

### 3.3 Bernoulli Noise Injection

A key step in Alg. 1 is generating sub-TS via Bernoulli NI. The rationale for sub-TS when estimating the state path and volatilities is that the estimation can be sensitive to the TS data for the COMS-GARCH process. The Bernoulli NI can create an ensemble of sub-TS' of considerable diversity among the ensemble members across iterations to reduce the sensitivity, to reduce the inference

sensitivity. Alg. 2 lists the steps of the Bernoulli NI. It outputs a sub-TS $\{\mathbf{G}, \mathbf{t}\}$, which is then fed to Alg. 1; only the states $\mathbf{S}$ at the retained timepoints are updated in each iteration, and the states of the dropped timepoints are kept at the values from the previous iteration, saving costs computationally. For the choice of Bernoulli rate $p$, a $k$-fold cross-validation (CV) procedure can be used, the detail of which are provided in the expanded paper.

**Algorithm 2:** Bernoulli Noise Injection

**input :** Original TS $\mathbf{G}$; Bernoulli NI rate $p$.
**output :** sub-TS $(\mathbf{Y}, \mathbf{t})$.

1 Draw $e_i$ from Bern$(1-p)$ for $i = 2, \ldots, n-1$. Set $e_0 = e_1 = e_n = 1$;
2 Let $\mathbf{G} = \{\mathbf{G} : \mathbf{e} \neq 0\}$, where $\mathbf{e} = \{e_i\}_{i=0}^{n}$, $n = \sum_{i=1}^{n} e_i$;
3 Obtain $\mathbf{Y} = \{Y_1, \ldots, Y_n\} = \text{diff}(\mathbf{G})$;
4 Set $\mathbf{t} = \mathbf{t}$. For $0 < i < n-1$, let $t_{i+1} = \begin{cases} i + t_{i+1} \text{ and } i = 0 \text{ if } e_i = 0 \\ t_{i+1} \quad t_{i+1} \text{ if } e_i = 1 \end{cases}$;
$\mathbf{t} = \{t : t \neq 0\}$.

**Claim 1.** The differenced $\mathbf{Y}$ in a sub-TS after Bernoulli NI is a summation of a sequence of differenced $\mathbf{Y}$ with the dropped observations in the original TS.

Claim 1 is a simple but interesting fact. For example, if $G_{i+1}$ gets dropped from the sequence of $\ldots, G_i, G_{i+1}, G_{i+2}, \ldots$, then $Y_{i'} = G_{i+2} - G_i = (G_{i+2} - G_{i+1}) + (G_{i+1} - G_i) = Y_{i+2} + Y_{i+1}$; say $r$ observations are dropped between $G_i$ and $G_{i+r+1}$, then $Y_{i'} = G_{i+r+1} - G_i = (G_{i+r+1} - G_{i+r}) + (G_{i+r} - G_{i+r-1}) + \cdots + (G_{i+1} - G_i) = Y_{i+r+1} + \cdots + Y_{i+1}$. This fact is used in the proof of Proposition 2 in Sec 3.4. Since the NI rate $p$ is usually small and the timepoints are dropped from the original TS randomly, with the fine time scale on which the TS is collected, the COMS-GARCH process can "digests" these "missing" timepoints effortlessly, without needing an ad-hoc approach to handle these dropped timepoints. The full conditional distributions of $\cdots$, and states $\{s_i\}$ given the sub-TS in each iteration are given in Eqs (13) and (15) by replacing the original TS $(\mathbf{Y}, \mathbf{t})$ with the sub-TS $(\mathbf{Y}, \mathbf{t})$.

The Bernoulli NI for COMS-GARCH is inspired by the dropout technique for regularizing neural networks (NNs) [30], which injects Bernoulli noises to input and hidden nodes during training, leading to model regularization. The Bernoulli NI we propose here is different procedurally in that it is applied to the observed data and drops random timepoints in the original TS in each iteration, rather than generating sub-models; in other words, the COMS-GARCH model remains as is during training. The benefits of the Bernoulli NI here include reduced computational cost, its connection with ensemble learning and inferential stability and robustness. The Bernoulli NI also bears some similarity to bagging [8], a well-known ensemble learning algorithm, but differs from bagging in two aspects. First, the Bernoulli NI leads to a random sub-TS (without replacement) of the original TS in each iteration of Alg. 1 whereas bagging often generates a bootstrapped sample set with replacement that is of the same size as the training data. Second, bagging often generates multiple sets of samples, trains a model on each set in parallel, and then ensembles them into a meta-model, whereas

the ensemble learning brought by the Bernoulli NI to the MAP estimation for COMS-GARCH is implicit, iterative, and realized sequentially.

### 3.4 Theoretical Analysis on Inferential Benefits of Bernoulli NI

In this section, we investigate theoretically the reasons behind the inferential benefits (improved efficiency and robustness in the MAP estimates of state path and volatilities) of the Bernoulli NI in two aspects.

First, we show, through the iterative Alg. 1, that the Bernoulli NI leads to sequential and implicit ensemble learning of the parameters and states for the COMS-GARCH model. The formal results are given in Proposition 1.

**Proposition 1** (ensemble learning of state path)**.** Let $\mathbf{Y}_{j-1} = (Y_1, \ldots, Y_{j-1})$, $\mathbf{S}_1 = (s_1, \ldots, s_i = k_1, \ldots, s_j)$ and $\mathbf{S}_2 = (s_1, \ldots, s_i = k_2, \ldots, s_j)$ $\forall k_1 \neq k_2 \in \{1, \ldots, g\}; i \leq n - b$. Assume that for $\forall \varepsilon > 0$, $\exists b \in \mathbb{N}^+$ such that

$$\frac{\prod_{j=i}^{n} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_1)}{\prod_{j=i}^{n} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_2)} \Big/ \frac{\prod_{j=i}^{i+b} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_1)}{\prod_{j=i}^{i+b} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_2)} < \varepsilon. \tag{19}$$

There exist $C_{k-1}^{b-1}$ ways to yield a set of b observations from a sequence of $k \in [b, n - i]$ consecutive observations. Denote the ensemble of the resultant $C_{k-1}^{b-1}$ sub-TS' by $\tilde{\mathcal{Y}}$. Given a Bernoulli NI rate $p$, the conditional posterior distribution of $s_i$ given the ensemble $\tilde{\mathcal{Y}}$ is

$$\sum_{k=b}^{n-i} p^{k-b}(1-p)^{b-1} \sum_{\tilde{\mathbf{Y}} \in \tilde{\mathcal{Y}}} f(s_i | \mathbf{S}_{-i}; \ldots; \tilde{\mathbf{Y}}). \tag{20}$$

The proof of Proposition 1 is straightforward. Eq (6) implies that the conditional distribution of $Y_j$ depends only on its variance since its mean is fixed at 0. Eq (9) suggests that the impact of state $s_i$ on $\sigma_{j-1}^2$ (and thus $\sigma_j^2$) decreases as $i$ departs from $j$ given the recursive formula on $\sigma^2$. Taken together, it implies that the state at time $t_i$ has minimal effect on the distribution of $Y_j$ at a future timepoint $t_j$ once the distance $t_j - t_i$ surpasses a certain threshold, which we use $b$ to denote. Mathematically, it means the ratio between $\prod_{j=i+b+1}^{n} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_1)$ and $\prod_{j=i+b+1}^{n} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_2)$ is arbitrarily close to 1, or

$$\frac{\prod_{j=i}^{i+b} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_1)}{\prod_{j=i}^{i+b} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_2)} \Big/ \frac{\prod_{j=i+b+1}^{n} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_1)}{\prod_{j=i+b+1}^{n} f(Y_j | \mathbf{Y}_{j-1}; \mathbf{S}_2)} - 1 <$$

for any $> 0$, leading to Eq (19). $k$ given $b$ and $p$ follows a negative binomial distribution, leading directly to Eq (20).

Taken together with Eq (15), Eq (19) implies the posterior distribution of $s_i$ can be almost surely determined by the $b$ observations in TS $\mathbf{Y}$ that immediately follow $t_i$; that is, $f(s_i | \mathbf{S}_{-i}; \ldots; \mathbf{Y}) = f(s_i | \mathbf{S}_{-i}; \ldots; Y_i, Y_{i+1}, \ldots, Y_{i+b})$. This narrow focus on just $b$ observations is undesirable especially when $b$ is small because the inference about the state path can become unstable and highly sensitive to even insignificant fluctuation in the TS. The Bernoulli NI helps mitigate this concern by diversifying the set of the $b$ observations. After the Bernoulli NI, the posterior probability of $s_i$ is a weighted average of the posterior distributions over multiple sets of $b$ observations with different compositions across iterations, as suggested by Eq (20), leading to more robust state estimation.

Fig 1 provides a visual illustration of the ensemble effect achieved through the Bernoulli NI. The dashed line in each plot corresponds to the right $y$-axis that presents the size of an ensemble. The solid lines correspond to the left $y$-axis, representing the weights assigned to ensembles of different sizes. When there is no Bernoulli NI ($p = 0$ and $k = b$), the ensemble is of size 1 (the first point on the dashed line in each plot). For $p > 0$, we have more than one way of generating the set of $b$ observations; and the actual ensemble size depends on $p$ and $k$. In brief, for a fixed $b$, as $k$ increases, the size of the ensemble set $\Upsilon$, $C_{k-1}^{b-1}$, increases dramatically (the dashed line within each plot), implying more sub-TS' are involved to obtain the posterior distribution of $s_i$. In addition, the ensemble set also increases dramatically with $b$ for a fixed $k$ $b$ value (the trend of the dashed lines across the 3 plots). The separated lines for different $p$ suggest that ensembles of different sizes are not weighted equally toward the conditional posterior distribution of $s_i$: the larger an ensemble, the smaller the weight it carries, especially for small $p$. Fig 1 implies that $p$ as small as $O(0.01)$ can create an ensemble of sub-TS' of enough diversity among the ensemble members to bring more robustness to the inference. A large $p$ leads to a larger ensemble, but also a higher computational cost which could overshadow the improved diversity. In addition, a large $p$ may drop too many timepoints and lead to too much fluctuation in the sub-TS' from iteration to iteration, resulting in possibly large bias or large variance in the estimation.

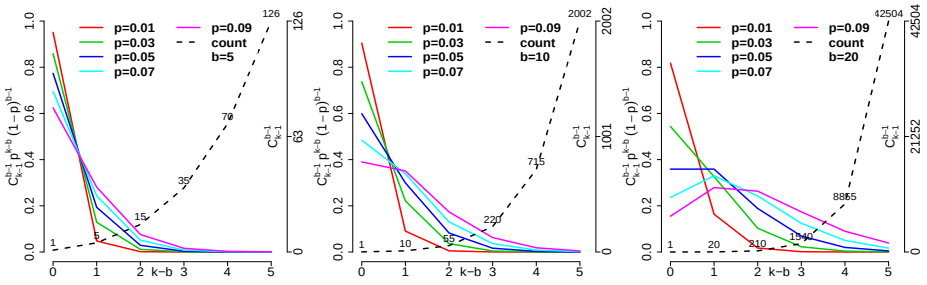



**Fig. 1.** Size of ensemble $\Upsilon$ (right $y$-axis) and weights $w(k; p; b)$ assigned to ensembles of different sizes (left $y$-axis) for different $p$ and $b$

Second, we establish that the Bernoulli NI also improves the stability of the objective function from which the MAP estimation is obtained, in the presence of random external perturbation in the TS. With a more stable objective function, the MAP estimates of ( ; ; $S$; $^2$) are also expected to be more stable. The formal result is presented in Proposition 2 (the proof is in the expanded paper).

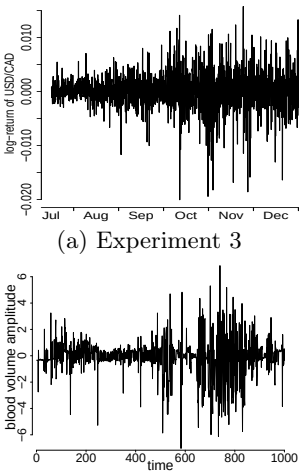
**Proposition 2** (improved stability of objective function)**.** Let $Y_i' = Y_i + z_i$, with $z_i$ $N(0; "^2)$ independently for $i = 1; : : : ; n$, be an externally perturbed observation to the original observation $Y_i$ from TS $\Upsilon$; and $Y_i'$ comprises the perturbed TS $\Upsilon'$. Let $\Upsilon$ and $\Upsilon'$ denote a sub-TS of $\Upsilon$ and $\Upsilon'$, respectively, after implementing the Bernoulli NI in an iteration of Alg. 1. The difference in the objection function (negative log-likelihood function or negative log-posterior distributions of ( ; ); $S$ and $^2$) given $\Upsilon'$ vs. that given $\Upsilon$ after the Bernoulli NI is on average smaller than the difference obtained without the Bernoulli NI.

## 4 Experiments

We run 4 experiments to demonstrate the applications of the COMS-GARCH process and the reSAVE algorithm. Experiment 1 shows the improved robustness of the MAP estimates of volatilities brought by the Bernoulli NI in the reSAVE procedure in a one-state COMS-GARCH process. Experiment 2 demonstrates the inferential robustness and computational efficiency of the reSAVE procedure in a two-state COMS-GARCH process and to compare with the MSGARCH(1,1) process. Experiments 3 and 4 apply the COMS-GARCH process and the reSAVE procedure to a real exchange rate TS (`https://www.histdata.com`) and a real blood volume amplitude (BVA) TS (`https://archive.ics.uci.edu/ml/datasets/PPG-DaLiA`) – to identify multiple states and estimate volatility.

### 4.1 Experiment Setting

In experiment 1, the TS data $\mathsf{Y}$ is simulated from a single-state CO-GARCH model and has $500$ timepoints. The time gap $t$ between two consecutive observations is characterized by a Poisson process with rate , i.e., $\mathrm{E}(t) = {}^{-1}$. We examined 4 values (2.5, 5, 10, 20). In experiment 2, the simulated TS $\mathsf{Y}$ from the COMS-GARCH process has 2 states and $1;000$ timepoints. The transitions among the states between two adjacent timepoints are modeled by a hidden Markov process transition parameters. We examined 2 set of : ${}_{12} = {}_{21} = 0{:}1$ and $0{:}25$ ( ${}_{11} = {}_{22} = 0{:}9$ and $0{:}75$). A similar Poisson process as in experiment 1 was used to simulate the TS data in the two states at $= 10;40$, respectively. In experiment 3, to keep the data at a manageable size, we took every $90$-th observation and performed a log transformation on the exchange rate TS between US dollar and Canadian dollar. The final TS $\mathsf{G}$ contains $1501$ times points over half a year. In experiment 4, we extracted the measurements from the photoplethysmograph of the blood volume pulse (64 Hz; i.e., 64 times per second) by taking the valley and peak pulse values in each cycle and scaling them by $0{:}01$. We then took the first $1;000$ measurements from a random patient as the input $\mathsf{Y}$. The TS' in experiments 3 and 4 have multiple regimens and the timepoints are irregularly spaced.




(a) Experiment 3

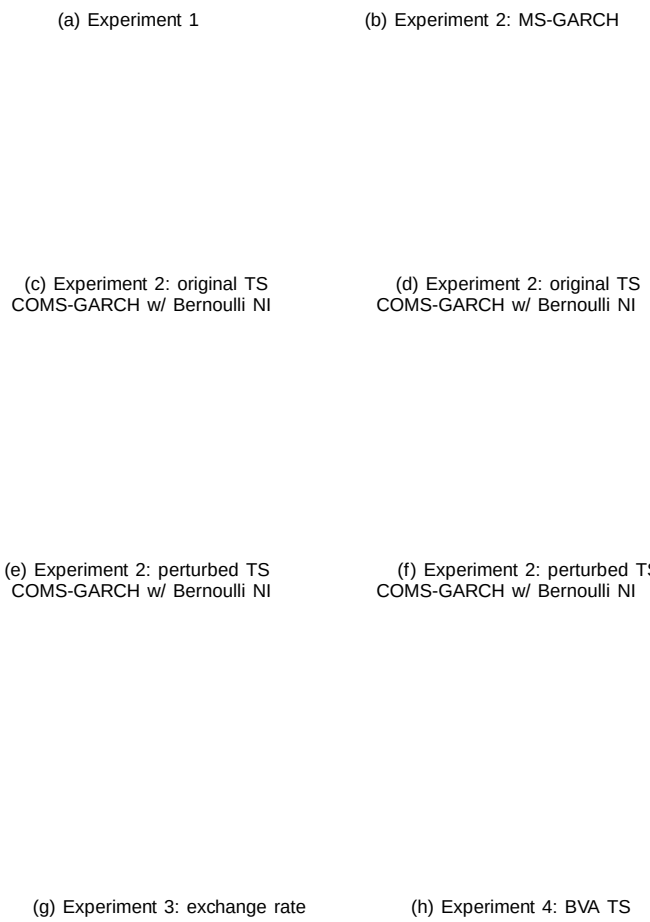
(b) Experiment 4

**Fig. 2.** Observed TS $\mathsf{Y}$

In all 4 experiments, we imposed non-informative priors on $= ( \ ; \ ; \ )$, set $b = 20$ and $m = 6$ for multiple path sampling. We examined a range of the Bernoulli NI rates in in experiments 1 and 2 and set $p$ at 0.02 in experiments 3 and 4. The convergence of the reSAVE procedure was examined by visual inspection of the trace plots of the MAP estimates for $( \ ; \ )$ and $\log( \ ; \ ; \ {}^{2}; Sj\mathsf{Y}; \ \mathrm{t})$. MS-GARCH(1, 1) in experiment 2 was implemented using R package `MSGARCH`. Since the MS-GARCH model assumes evenly-space TS data, we first applied linear interpolation to each simulated TS to obtain the equally spaced TS data before fitting the model.

### 4.2 Results

Due to space limitation, we present selected main results; more results can be found in the expanded paper. Fig 3 presents examples on the estimated volatilities and states (a randomly chosen repeat out of 50 is shown in experiments 1 and 2). The main observations are as follows. In Figs 3(a), (c) to (f)), the estimates of volatilities and states almost completely overlap with their true values at all timepoints, suggesting high accuracy in the estimation. The results in experiment



(a) Experiment 1 (b) Experiment 2: MS-GARCH

(c) Experiment 2: original TS COMS-GARCH w/ Bernoulli NI (d) Experiment 2: original TS COMS-GARCH w/ Bernoulli NI

(e) Experiment 2: perturbed TS COMS-GARCH w/ Bernoulli NI (f) Experiment 2: perturbed TS COMS-GARCH w/ Bernoulli NI

(g) Experiment 3: exchange rate (h) Experiment 4: BVA TS

Fig. 3. Estimated volatilities and states via COMS-GARCH. Experiment 1 has only one state so there is no state path estimation; (b) shows the results from MS-GARCH, a comparison method to COMS-GARCH

2 further suggest that when there is no external perturbation in the TS, the Bernoulli NI does not negatively impact the volatility and state estimation but improves the accuracy in the estimation when there is, implying that the Bernoulli NI is an intelligent technique, only doing its tricks when needed and is silent otherwise. By contract, MS-GARCH tend to over-estimate volatilities and there are more mis-predicted states (Fig 3(b)). In experiments 3 and 4 (Fig 3(g) and (h)), the estimated states and volatilities via COMS-GARCH and the reSAVE procedure re ect the two expected states in each case: a change in the state somewhere between September and October in experiment 3, and di erent BVA states corresponding to di erent physical conditions or emotional episodes of subjects in experiment 4.

Fig 4 summarizes the relative %|bias| of the estimated volatilities ($\ell_1$-distance scaled by the true volatility) and the state mis-prediction rate, both averaged across the time points in each TS and across the 50 repeats in experiments 1 and 2. Fig 4(a) suggests the volatility estimation can be sensitive to even mild uctuation in the TS data as the bias with externally perturbed TS (crosses at $p=0$) is larger than that at no external perturbation (circles at $p=0$). Bernoulli NI helps bring the bias down at all the examined values but there is not much of a di erence across $p$. In Fig 4(b) to 4(e), the accuracy of the state identi cation and volatility estimation is signi cantly improved with a proper Bernoulli NI rate $p$ than without NI. The smallest mis-prediction rate (8% 17%) and volatility estimation bias (12% 25%) are achieved around $p$ at 0:01 0:02 in most scenarios; and further increasing $p$ does not seem to improve the prediction accuracy. In addition, how much the Bernoulli NI helps in reducing the prediction bias relates to , , and .

(a) (b) (c) (d)

(a) volatility estimation from original TS in Experiment 2
(b) volatility estimation from perturbed TS in Experiment 2
(c) state prediction from original TS in Experiment 2
(d) state prediction from perturbed TS Experiment 2
(e) volatility estimation in Experiment 1

(e)

Fig. 4. Volatility estimation bias and state mis-prediction rate in Experiments 1 and 2

Fig 5 presents some trace plots from the iterative reSAVE procedure in experiment 2. The setting ($m = 6; p = 0.02$) converges with the least iterations, followed by ($m = 6; p = 0$). The setting ($m = 1; p = 0$) needs the most iterations to converge. This observation suggests both Bernoulli NI and the multiple path sampling scheme (m = 6) can accelerate the convergence of the reSAVE procedure in estimating the state path and volatilities.

In the expanded paper, we also present the biases and root mean squared errors for the MAP estimates of and . The estimates for $(k)$ and $(k)$ are generally accurate but there is noticeable estimation bias for and $(k)$ in some simulation scenarios. The relatively large bias for can be at least partially attributed to the low transition probabilities between different states, leading to data sparsity in estimating . Estimation bias of is rather a common problem and exists in the GARCH, CO-GARCH, and MS-GARCH estimation, rather than something unique to the Gibbs sampler or the reSAVE procedure we propose for the COMS-GARCH process. [7] suggests that the MLEs for the parameters from the GARCH model are biased; the estimation bias for the GARCH parameters for both the quasi-MLE and the constrained M-estimators (more robust) can be as large as 20% [22] and as large as 30% for the parameters of the CO-GARCH process [20].

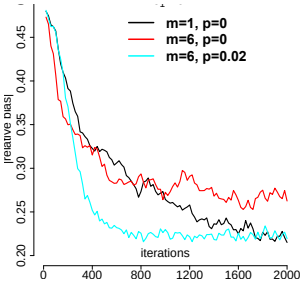



**Fig. 5.** Trace plots of state mis-prediction rate (top) and |relative bias| of estimated volatility (bottom) in a single TS repetition in Experiment 2

## 5 Discussion

We propose the COMS-GARCH process for handling irregularly spaced TS data with multiple volatility states. We also introduce the reSAVE procedure with the Bernoulli NI for obtaining the MAP estimates for model parameters, state path, and volatilities. The computational efficiency and inferential robustness of the reSAVE procedure are established and illustrated theoretically or empirically. Foresting is often of major interest in TS analysis as they provide insights into future trends and are useful for decision making (e.g., developing option trading strategies in financial markets, predicting earthquakes). We present in the expanded paper an algorithm for the $h^{th}$-step-ahead prediction of future volatilities and states through a trained COMS-GARCH process.

We conjecture that the reSAVE procedure is applicable not only to the COMS-GARCH and CO-GARCH processes but also to other solvable CO-*-GARCH processes. For example, it will make an interesting future topic to develop the COMS-Exponential-GARCH and COMS-Integrated-GARCH processes and examine the performance of the reSAVE procedure in these settings. We also expect that the reSAVE procedure can be used in the COMS-ARMA process

for trend estimation, yielding some types of weighted $l_2$ regularization on the ARMA parameters. More work is needed to prove conjectures formally.

Regarding when to use the COMS-GARCH process, TS' that are of high-frequency or irregularly spaced are potential candidates (the "CO" component); we may plot the TS and visually examine whether there is any sign for multiple states (the "MS" component) or there may exist domain or prior knowledge suggesting state multiplicity (e.g., experiments 3 and 4 in Section 4). For a more quantitative approach, one may fit the candidate models (e.g., COGARCH vs. COMS-GARCH, MS-GARCH and COMS-GARCH, $_1$ vs $_2$ states in COMS-GARCH) to the TS in the Bayesian framework, and compare the deviance information criterion (DIC) [29], a Bayesian measure of model fit and choose the model with the smaller (or) smallest DIC.

Finally, as discussed briefly in the experiments, there lacks in-depth theoretical investigation on the asymptotic properties of the MLE and MAP estimation for the MS-GARCH and CO-GARCH processes, and the estimation bias for *-GARCH parameters is well documented. We will look into the asymptotic properties of the COMS-GARCH process as $n \; ! \; 1$ and $T \; ! \; 1$ and develop more accurate inferential procedures for COMS-GARCH in the future.

## References


1. Ardia, D., Bluteau, K., Boudt, K., Catania, L., Trottier, D.A.: Markov-switching GARCH models in R: The MSGARCH package. Journal of Statistical Software **91**, 4 (2019)
2. Augustyniak, M.: Maximum likelihood estimation of the markov-switching GARCH model. Computational Statistics & Data Analysis **76**, 61–75 (2014)
3. Bauwens, L., De Backer, B., Dufays, A.: A bayesian method of change-point estimation with recurrent regimes: Application to GARCH models. Journal of Empirical Finance **29**, 207–229 (2014)
4. Bauwens, L., Dufaysa, A., V.K.Romboutsb, J.: Marginal likelihood for markov-switching and change-point GARCH models. Lecture Notes in Economics and Mathematical Systems **178(3)**, 508–522 (2014)
5. Bauwens, L., Preminger, A., Rombouts, J.V.K.: Theory and inference for a markov switching GARCH model. Econometrics Journal **13**, 218–244 (2010)
6. Billioa, M., Casarina, R., Osuntuyi, A.: Efficient gibbs sampling for markov switching GARCH models. Computational Statistics & Data Analysis **100**, 37–57 (2016)
7. Bollerslev, T., Engle, R.F., Nelson, D.B.: Arch models. Handbook of econometrics **4**, 2959–3038 (1994)
8. Breiman, L.: Bagging predictors. Machine Learning **24**(2), 123–140 (1996)
9. Brockwell, P., Chadraa, E., Lindner, A.: Continuous-time GARCH processes. The Annals of Applied Probability **16**, 790–826 (2006)
10. Corradi, V.: Reconsidering the continuous time limit of the GARCH (1, 1) process. Journal of econometrics **96**(1), 145–153 (2000)
11. Dueker, M.J.: Markov switching in GARCH processes and mean-reverting stock-market volatility. Journal of Business & Economic Statistics **15**(1), 26–34 (1997)
12. Elliott, R., Lau, J., Miao, H., Siu, T.: Viterbi-based estimation for markov switching GARCH models. Applied Mathematical Finance **19(3)**, 1–13 (2012)

13. Glosten, L.R., Jagannathan, R., Runkle, D.E.: On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance **48(5)**, 1779–1801 (1993)
14. Gray, S.F.: Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics **42(1)**, 27–62 (1996)
15. Haas, M., Mittnik, S., Paolella, M.S.: A new approach to markov-switching GARCH models. Journal of financial econometrics **2**(4), 493–530 (2004)
16. Kim, C.J.: Dynamic linear models with markov-switching. Journal of Econometrics **60**(1-2), 1–22 (1994)
17. Klaassen, F.: Improving GARCH volatility forecasts with regime-switching garch. In: Advances in Markov-switching models, pp. 223–254. Springer (2002)
18. Kluppelberg, C., Lindner, A., Maller, R.: A continuous-time GARCH process driven by a levy process: Stationarity and second-order behavior. Journal of Applied Probability **41**, 601–622 (2004)
19. Lamoureux, C.G., Lastrapes, W.D.: Persistence in variance, structural change, and the GARCH model. Journal of Business & Economic Statistics **8**(2), 225–234 (1990)
20. Maller, R.A., Müller, G., Szimayer, A.: GARCH modelling in continuous time for irregularly spaced time series data. Bernoulli **14(2)**, 519–542 (2008)
21. Marín, J.M., Rodríguez-Bernal, M.T., Romero, E.: Data cloning estimation of GARCH and COGARCH models. Journal of Statistical Computation and Simulation **85**(9), 1818–1831 (2015)
22. Mendes, M., De, B.V.: Assessing the bias of maximum likelihood estimates of contaminated GARCH models. Journal of Statistical Computation and Simulation **67**(4), 359–376 (2000)
23. Mikosch, T., Starica, C.: Nonstationarities in financial time series, the long-range dependence, and the igarch effects. Review of Economics and Statistics **86(1)**, 378–390 (2004)
24. Müller, G.: MCMC estimation of the COGARCH(1;1) model. Journal of Financial Econometrics **8(4)**, 481–510 (2010)
25. Nelson, D.B.: Stationarity and persistence in the garch(1,1) model. Econometric Theory **6**, 318–334 (1990)
26. Nelson, D.B.: Conditional heteroskedasticity in asset returns: A new approach. Econometrica **59**, 347–370 (1991)
27. Nelson, D.: Arch models as diffusion approximations. Journal of Econometrics **45**, 7–38 (1990)
28. Sentana, E.: Quadratic ARCH models. Review of Economic Studies **62(4)**, 639–661 (1995)
29. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A.: Bayesian measures of model complexity and fit. Journal of the royal statistical society: Series b (statistical methodology) **64**(4), 583–639 (2002)
30. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15**, 1929–1958 (2014)
31. Wang, Y.: Asymptotic nonequivalence of GARCH models and diffusions. The Annals of Statistics **30(3)**, 754–783 (2002)
32. Zakoian, J.M.: Threshold heteroskedastic models. Journal of Economic Dynamics and Control **18(5)**, 931–955 (1994)