1. Introduction

In finance, volatility refers to the degree of variation or fluctuation in the price of a financial asset. Volatility is a key concept in finance because it reflects the level of risk associated with an investment. Investors and traders seek for measures of volatility to assess the potential for both gains and losses and to make financial and investment decisions, however, volatility is unobservable.

A statistical model–GARCH model is used to model and forecast volatility using financial time series return data. This is because GARCH captures known properties in stock returns under a time series dataset such as volatility clustering, fat-tail distribution, and weak autocorrelation on returns (Takaishi, 2018).

In this report, we work with the data set of daily return values of the Australian All Ordinaries Index (AORD) for the last ten years to model volatility. We examine the GARCH (1, 1) model and fit these using both Gaussian Variational Bayes (GVB) and Markov chain Monte Carlo (MCMC) with and without transformation. In the GARCH (1, 1) model, volatility is forecasted by the last observation of squared return and conditional variance. The aim of this report would be to compare the different methodology of estimating the posterior to see which works most optimally using this dataset.

2. EDA

The dataset was first centred such that $y_t = y_{t,\text{original}} - \bar{y}_{t,\text{original}}$.

Table 1: Statistical analysis of the centred AORD returns.

Standard Deviation	Minimum	Maximum	Skewness	Kurtosis
0.8805	−7.0784	4.4235	−0.9649	9.9521

Figure 1: Time series plot, Histogram and ACF plot of the centred AORD returns.

Financial returns often have fat-tailed distributions and high kurtosis, which for this sample is 9.9521 as seen in Table 1, with extreme price movements (both positive and negative) occurring more frequently than expected in a normal distribution. Return data often exhibit significant volatility implying sharp return movements, both up and down over short periods. Moreover, the returns show a slight negative skewness of -0.9649 which indicates that the tail on the left side is slightly longer which is supported by the extreme values in our stock returns where the minimum was -7.0784, and maximum was 4.4235. The centred AORD returns in our sample also seem to follow a stationary pattern, i.e., statistical properties (e.g., the mean) do not change significantly over time. Moreover, from the ACF where the correlation is extremely close to zero, the returns are behaving like a random walk, meaning past returns are not a reliable indicator of future returns. This supports the Efficient Market Hypothesis (EMH), which suggests that stock prices already reflect all available information, so past prices or returns can't be used to predict future prices or returns with any consistency (Fama, 1970).

3. Methodology

3.1. GARCH Model

3.1.1. Understanding the GARCH model

The GARCH model are widely employed to analyse volatility of financial time series data. In particular, GARCH model takes in values of the past squared observations and volatility to model the variance at time t. The GARCH (1,1) model is favoured for its relatively simple implementation which forecast volatility by fitting one autoregressive lag or ARCH term and one moving average lag (GARCH term).

$$y_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0,1), \qquad t = 1,2,\dots,T$$
$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \qquad t = 2,\dots,T$$

Since in finance, the historical volatility was empirically found to a good indicator for the future fluctuations, GARCH is considered as a good statistical model for volatility measure due to its ability to capture the time-varying nature of volatility.

3.1.2. Log-likelihood

Since $y_t = \sigma_t \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0,1)$, it follows that the return also follows a Gaussian distribution, albeit with different variance: $y_t \sim N(0, \sigma^2)$. Denoting $\theta = (\omega, \alpha, \beta)$, the likelihood function is:

$$\mathcal{L}(\theta|y) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi {\sigma_t}^2}} \exp\left(-\frac{{y_t}^2}{2{\sigma_t}^2}\right) \propto \prod_{t=1}^{T} \frac{1}{\sqrt{{\sigma_t}^2}} \exp\left(-\frac{{y_t}^2}{2{\sigma_t}^2}\right)$$

and the associated log-likelihood function would be:

$$\ell(\theta) = \sum_{t=1}^{T} \left(\log({\sigma_t}^2)^{-\frac{1}{2}} - \frac{{y_t}^2}{2{\sigma_t}^2}\right) = \sum_{t=1}^{T} \left(-\frac{1}{2}\log({\sigma_t}^2) - \frac{1}{2} \times \frac{{y_t}^2}{{\sigma_t}^2}\right) = -\frac{1}{2}\sum_{t=1}^{T} \left(\log({\sigma_t}^2) + \frac{{y_t}^2}{{\sigma_t}^2}\right) \quad (1)$$

3.2. Markov Chain Monte Carlo

3.2.1. Without transformation

Without transformation, the parameters are constrained to the stationary conditions where $\alpha + \beta < 1$ and positivity where $\omega,\ \alpha,\ \beta > 0$. The process starts with initialising the α, β, ω parameters to be used in the first iteration of the chain from the prior distributions:

$$\alpha \sim Beta(1.5,10), \qquad \beta \sim Beta(10,1.5), \qquad \omega \propto 1$$

The RWHM method was used with an adaptive sigma (Σ) as it can lead to faster convergence and more efficient sampling, reducing the computation time and producing better estimates as per Roberts & Rosenthal's (2009) recommendation. The sigma was then used to generate a proposal:

$$y = X_n + \epsilon, \qquad \epsilon \sim N(0, \Sigma)$$

The posterior has the form:

$p(\theta|y) \propto p(\theta)p(y|\theta) = p(\alpha)\ p(\beta)\ p(\omega)\ p(y|\alpha,\beta,\omega) = p(\alpha)\ p(\beta)\ \mathcal{L}(\theta|y)\ since\ \omega \propto 1$

$$= \frac{1}{B(1.5,10)}\alpha^{0.5}(1-\alpha)^9 \times \frac{1}{B(10,1.5)}\beta^9(1-\beta)^{0.5} \times \prod_{t=1}^{T}\frac{1}{\sqrt{{\sigma_t}^2}}\exp\left(-\frac{{y_t}^2}{2{\sigma_t}^2}\right)$$

$$\propto \alpha^{0.5}(1-\alpha)^9 \times \beta^9(1-\beta)^{0.5} \times \prod_{t=1}^{T}\frac{1}{\sqrt{{\sigma_t}^2}}\exp\left(-\frac{{y_t}^2}{2{\sigma_t}^2}\right) \; since\; B(\cdot) = const$$

$$= exp\left[0.5\log\alpha + 9\log(1-\alpha) + 9\log\beta + 0.5\log(1-\beta) - \frac{1}{2}\sum_{t=1}^{T}\left(\log({\sigma_t}^2) + \frac{{y_t}^2}{{\sigma_t}^2}\right)\right] (2)$$

The acceptance probability in the RWMH algorithm is:

$$\alpha = \min\left(\frac{k(\theta_{proposal})}{k(\theta_{current})}, 1\right) (3) \;\; where\; kernel\; k(\cdot)\; is\; the\; (2)\; term\; right\; above$$

Keeping in mind the positivity and stationary constraints on the parameters, the loglikelihood was calculated by first checking if the three parameters satisfy said constraints, else log posterior is set to $-\infty$ which causes the posterior to be zeroed out, thus rejecting the proposal:

$$k(\theta_{proposal}) = \exp\left(\log\left(k(\theta_{proposal})\right)\right) = e^{-\infty} = 0$$
$$\therefore \alpha = 0 \leq U, \;\; \forall\, U \sim \mathcal{U}(0,1)$$

In other words, the proposal always gets rejected if any of the proposed values for the parameters do not pass the constraints. Once the parameters satisfy the constraints, the GARCH (1,1) variances are calculated using:

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2, \quad t = 2, \ldots, T$$

Which is then plugged into formula (1) to get the corresponding log-likelihood, and into formula (2) and (3) to get the corresponding posterior and acceptance probability α. This computation can be found in the ***log_posterior.m*** file.

3.2.2. With transformation

Instead of the original constrained parameter θ, we adopt a transformed version of the parameters $\tilde{\theta} \coloneqq \left(\widetilde{\theta_1}, \widetilde{\theta_2}, \widetilde{\theta_3}\right)^T$ so that they are no longer constrained and therefore easier to estimate. The relationship between the original and transformed parameters is:

$$\omega = \exp\left(\widetilde{\theta_1}\right)$$

$$\alpha = \frac{\exp\left(\widetilde{\theta_2}\right)\exp\left(\widetilde{\theta_3}\right)}{1 + \exp\left(\widetilde{\theta_2}\right) + \exp\left(\widetilde{\theta_3}\right) + \exp\left(\widetilde{\theta_2}\right)\exp\left(\widetilde{\theta_3}\right)}$$

$$\beta = \frac{\exp\left(\widetilde{\theta_2}\right)}{1 + \exp\left(\widetilde{\theta_2}\right) + \exp\left(\widetilde{\theta_3}\right) + \exp\left(\widetilde{\theta_2}\right)\exp\left(\widetilde{\theta_3}\right)}$$

The process is the similar to that with the original parameters, in that a RWMH algorithm with adaptive Sigma is employed. However, since $\tilde{\theta}$ is unconstrained, we no longer adopt the Beta priors like in the previous approach. Instead, flat priors are chosen for all 3 of the transformed parameters:

$$\widetilde{\theta_1} \propto 1, \qquad \widetilde{\theta_2} \propto 1, \qquad \widetilde{\theta_3} \propto 1$$

The posterior therefore has the form:

$$p(\theta|y) \;\propto\; p(\theta)p(y|\theta) = p(\alpha)\, p(\beta)\, p(\omega)\, p(y|\alpha, \beta, \omega) = \mathcal{L}(\theta|y)$$

$$= \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\log({\sigma_t}^2)+\frac{{y_t}^2}{{\sigma_t}^2}\right)\right\} \quad (4)$$

The acceptance probability is:

$$\alpha = \min\left(\frac{k(\theta_{proposal})}{k(\theta_{current})}, 1\right) \; where\ kernel\ k(\cdot)\ is\ the\ (4)\ term\ right\ above$$

This computation can be found in the ***loglik_trans.m*** file. Note that the Marko chains generated in this MCMC implementation are for $\widetilde{\theta_1}, \widetilde{\theta_2}$ and $\widetilde{\theta_3}$. At each step, values are proposed for $\widetilde{\theta_1}, \widetilde{\theta_2}$ and $\widetilde{\theta_3}$, which are then transformed into values for the original ω, α and β. These are finally used to derive the conditional variances ${\sigma_t}^2$ that are needed to compute (4).

Once the Markov chains for $\widetilde{\theta_1}, \widetilde{\theta_2}$ and $\widetilde{\theta_3}$ are obtained, the $(\widetilde{\theta_1}, \widetilde{\theta_2}, \widetilde{\theta_3})$ values at iteration i are transformed back into the corresponding (ω, α, β) to arrive at the Markov chain for the original parameters (ω, α, β).

3.3. Gaussian Variational Bayes

3.3.1. Transformation of parameters

In this model, transformed parameters $\tilde{\theta} \coloneqq (\widetilde{\theta_1}, \widetilde{\theta_2}, \widetilde{\theta_3})^T$ will be adopted which are unconstrained and easier to estimate and fit in the GVB. The transformation applied is the same as the previous 3.2.2 MCMC with transformation section.

3.3.2. The gradient of log-likelihood function

Flat priors are used for all parameters for simplicity, i.e., $p(\omega) \propto 1,\ p(\alpha) \propto 1,\ p(\beta) \propto 1$.

The gradient of the log-likelihood function for the Gaussian GARCH (1,1) model

$$\nabla_{\tilde{\theta}}\ell(\tilde{\theta}) = \left(\nabla_{\tilde{\theta}}\ell(\widetilde{\theta_1}), \nabla_{\tilde{\theta}}\ell(\widetilde{\theta_2}), \nabla_{\tilde{\theta}}\ell(\widetilde{\theta_3})\right)^T$$

$$\nabla_{\tilde{\theta}}\ell(\widetilde{\theta_1}) = \frac{\partial\ell(\tilde{\theta})}{\partial\widetilde{\theta_1}} = -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_1}} + \frac{1}{2}\sum_{t=1}^{T}\frac{{y_t}^2}{{\sigma_t}^4}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_1}} = \frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_1}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right)$$

$$\nabla_{\tilde{\theta}}\ell(\widetilde{\theta_2}) = \frac{\partial\ell(\tilde{\theta})}{\partial\widetilde{\theta_2}} = -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_2}} + \frac{1}{2}\sum_{t=1}^{T}\frac{{y_t}^2}{{\sigma_t}^4}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_2}} = \frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_2}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right)$$

$$\nabla_{\tilde{\theta}}\ell(\widetilde{\theta_3}) = \frac{\partial\ell(\tilde{\theta})}{\partial\widetilde{\theta_3}} = -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_3}} + \frac{1}{2}\sum_{t=1}^{T}\frac{{y_t}^2}{{\sigma_t}^4}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_3}} = \frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_3}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right)$$

With $\tilde{\theta} \coloneqq (\widetilde{\theta_1}, \widetilde{\theta_2}, \widetilde{\theta_3})^T$, the gradient of the log-likelihood function can be written in more compact form:

$$\nabla_{\tilde{\theta}}\ell(\tilde{\theta}) = \frac{\partial\ell(\tilde{\theta})}{\partial\tilde{\theta}} = \frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\tilde{\theta}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right)$$

**Note: the mathematical derivation of the derivatives $\frac{\partial{\sigma_t}^2}{\partial\tilde{\theta}}$ will be shown detailly in section 3.3.4*

3.3.3. The variational distribution and gradient of log-variational distribution

We take approximation distribution $q_\lambda(\tilde{\theta})$ to be a multivariate Gaussian distribution with mean μ and covariance matrix $\sum$. The Cholesky decomposition which inverse of covariance matrix is defined as $\sum^{-1} = LL^T$ with L is a lower triangular matrix. The variational parameters to be optimized are now $\lambda = (\mu^T, vech(L)^T)^T$, where μ is a $d - dimenstion$ vector ($d = 3$), and $vech(L)$ is the half- vectorization of L, i.e. a $d(d+1)/2 - dimension$ vector obtained by vectorizing the lower triangular part of L. The log-density of variational distribution $q_\lambda(\tilde{\theta})$ is:

$$\log q_\lambda(\tilde{\theta}) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}|\sum| - \frac{1}{2}(\tilde{\theta}-\mu)^T\sum^{-1}(\tilde{\theta}-\mu)$$
$$= -\frac{d}{2}\log(2\pi) - \frac{1}{2}|L| - \frac{1}{2}(\tilde{\theta}-\mu)^T LL^T(\tilde{\theta}-\mu)$$

and the gradient of the log-variational distribution with respect to parameter $\tilde{\theta}$

$$\nabla_{\tilde{\theta}}\log q_\lambda(\tilde{\theta}) = -LL^T(\tilde{\theta}-\mu)$$

3.3.4. Gradient function derivation

Given the log-likelihood function of the GARCH (1,1) model with Gaussian innovations and the log-variational approximation, we can drive the expression of $h_\lambda(\tilde{\theta})$

$$h_\lambda(\tilde{\theta}) = \log\frac{p(\tilde{\theta})\mathcal{L}(\tilde{\theta};y)}{q_\lambda(\tilde{\theta})} = \log p(\tilde{\theta}) + \ell(\tilde{\theta}) - \log q_\lambda(\tilde{\theta})$$
$$= -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log({\sigma_t}^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{{y_t}^2}{{\sigma_t}^2} + \frac{d}{2}\log(2\pi) + \frac{1}{2}|L| + \frac{1}{2}(\tilde{\theta}-\mu)^T LL^T(\tilde{\theta}-\mu)$$

To implement an GVB algorithm, it is required to have gradient vector $\nabla_{\tilde{\theta}}h_\lambda(\tilde{\theta})$.

$$\nabla_{\tilde{\theta}}h_\lambda(\tilde{\theta}) = \nabla_{\tilde{\theta}}\log\frac{p(\tilde{\theta})\mathcal{L}(\tilde{\theta};y)}{q_\lambda(\tilde{\theta})} = \nabla_{\tilde{\theta}}\log p(\tilde{\theta}) + \nabla_{\tilde{\theta}}\ell(\tilde{\theta}) - \nabla_{\tilde{\theta}}\log q_\lambda(\tilde{\theta})$$

Combining all the above derived gradients of log-likelihood and log-variational approximation together, the mathematical expression for the gradient is

$$\nabla_{\tilde{\theta}}h_\lambda(\tilde{\theta}) = \frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial{\sigma_t}^2}{\partial\tilde{\theta}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right) + LL^T(\tilde{\theta}-\mu)$$

with

$$\frac{\partial{\sigma_t}^2}{\partial\tilde{\theta}} = \left(\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_1}}, \frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_2}}, \frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_3}}\right)^T$$

The derivatives of conditional variance ${\sigma_t}^2$ with respect to unconstrained parameters $\tilde{\theta} \coloneqq (\widetilde{\theta_1}, \widetilde{\theta_2}, \widetilde{\theta_3})^T$ can be derived further by using the chain rule, we have:

$$\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_1}} = \frac{\partial{\sigma_t}^2}{\partial\omega}\frac{\partial\omega}{\partial\widetilde{\theta_1}}$$
$$\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_2}} = \frac{\partial{\sigma_t}^2}{\partial\alpha}\frac{\partial\alpha}{\partial\widetilde{\theta_2}} + \frac{\partial{\sigma_t}^2}{\partial\beta}\frac{\partial\beta}{\partial\widetilde{\theta_2}}$$
$$\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_3}} = \frac{\partial{\sigma_t}^2}{\partial\alpha}\frac{\partial\alpha}{\partial\widetilde{\theta_3}} + \frac{\partial{\sigma_t}^2}{\partial\beta}\frac{\partial\beta}{\partial\widetilde{\theta_3}}$$

Taking the partial derivatives of conditional variance ${\sigma_t}^2$ with respect to $\theta = (\omega, \alpha, \beta)^T$, we can derive the following recursive formulas with $t = 2, \dots, T$

$$\frac{\partial{\sigma_t}^2}{\partial\omega} = 1 + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\omega}$$
$$\frac{\partial{\sigma_t}^2}{\partial\alpha} = {y_{t-1}}^2 + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\alpha}$$
$$\frac{\partial{\sigma_t}^2}{\partial\beta} = {\sigma_{t-1}}^2 + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\beta}$$

The derivatives of the original parameters θ with respect to the transformed parameters $\tilde{\theta} \coloneqq (\widetilde{\theta_1}, \widetilde{\theta_2}, \widetilde{\theta_3})^T$,

$$\frac{\partial\omega}{\partial\widetilde{\theta_1}} = \frac{\partial\exp(\widetilde{\theta_1})}{\partial\widetilde{\theta_1}} = \exp(\widetilde{\theta_1})$$

$$\frac{\partial \alpha}{\partial \widetilde{\theta_2}} = \frac{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3}) - \exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\left(\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)}{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\left(\exp(\widetilde{\theta_3})\right)^2}{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\left(1+\exp(\widetilde{\theta_3})\right)}{\left(\left(1+\exp(\widetilde{\theta_2})\right)\left(1+\exp(\widetilde{\theta_3})\right)\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})}{\left(1+\exp(\widetilde{\theta_2})\right)^2\left(1+\exp(\widetilde{\theta_3})\right)}$$

$$= \frac{\exp(\widetilde{\theta_2})^2\exp(-\widetilde{\theta_2})\exp(\widetilde{\theta_3})}{\left(1+\frac{1}{\exp(\widetilde{\theta_2})}\right)^2\exp(\widetilde{\theta_2})^2\left(1+\frac{1}{\exp(\widetilde{\theta_3})}\right)\exp(\widetilde{\theta_3})}$$

$$= \frac{\exp(-\widetilde{\theta_2})}{\left(1+\exp(-\widetilde{\theta_2})\right)^2\left(1+\exp(-\widetilde{\theta_3})\right)}$$

$$\frac{\partial \alpha}{\partial \widetilde{\theta_3}} = \frac{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3}) - \exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\left(\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)}{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})^2\exp(\widetilde{\theta_3})}{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\left(1+\exp(\widetilde{\theta_2})\right)}{\left(\left(1+\exp(\widetilde{\theta_2})\right)\left(1+\exp(\widetilde{\theta_3})\right)\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})}{\left(1+\exp(\widetilde{\theta_3})\right)^2\left(1+\exp(\widetilde{\theta_2})\right)}$$

$$= \frac{\exp(\widetilde{\theta_3})^2\exp(-\widetilde{\theta_3})\exp(\widetilde{\theta_2})}{\left(1+\frac{1}{\exp(\widetilde{\theta_3})}\right)^2\exp(\widetilde{\theta_3})^2\left(1+\frac{1}{\exp(\widetilde{\theta_2})}\right)\exp(\widetilde{\theta_2})}$$

$$= \frac{\exp(-\widetilde{\theta_3})}{\left(1+\exp(-\widetilde{\theta_3})\right)^2\left(1+\exp(-\widetilde{\theta_2})\right)}$$

$$\frac{\partial \beta}{\partial \widetilde{\theta_2}} = \frac{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)\exp(\widetilde{\theta_2}) - \exp(\widetilde{\theta_2})\left(\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)}{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})}{\left(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})\left(1+\exp(\widetilde{\theta_3})\right)}{\left(\left(1+\exp(\widetilde{\theta_2})\right)\left(1+\exp(\widetilde{\theta_3})\right)\right)^2}$$

$$= \frac{\exp(\widetilde{\theta_2})}{(1+\exp(\widetilde{\theta_2}))^2(1+\exp(\widetilde{\theta_3}))}$$
$$= \frac{\exp(\widetilde{\theta_2})^2\exp(-\widetilde{\theta_2})\exp(\widetilde{\theta_3})\exp(-\widetilde{\theta_3})}{\left(1+\frac{1}{\exp(\widetilde{\theta_2})}\right)^2\exp(\widetilde{\theta_2})^2\left(1+\frac{1}{\exp(\widetilde{\theta_3})}\right)\exp(\widetilde{\theta_3})}$$
$$= \frac{\exp(-\widetilde{\theta_2})\exp(-\widetilde{\theta_3})}{(1+\exp(-\widetilde{\theta_2}))^2(1+\exp(-\widetilde{\theta_3}))}$$

$$\frac{\partial\beta}{\partial\widetilde{\theta_3}} = \frac{-\exp(\widetilde{\theta_2})\left(\exp(\widetilde{\theta_3}) + \exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})\right)}{(1+\exp(\widetilde{\theta_2})+\exp(\widetilde{\theta_3})+\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3}))^2}$$
$$= \frac{-\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})(1+\exp(\widetilde{\theta_2}))}{\left((1+\exp(\widetilde{\theta_2}))(1+\exp(\widetilde{\theta_3}))\right)^2}$$
$$= \frac{-\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})}{(1+\exp(\widetilde{\theta_3}))^2(1+\exp(\widetilde{\theta_2}))}$$
$$= \frac{-\exp(\widetilde{\theta_2})\exp(\widetilde{\theta_3})}{(1+\exp(\widetilde{\theta_3}))^2\left(1+\frac{1}{\exp(\widetilde{\theta_2})}\right)\exp(\widetilde{\theta_2})}$$
$$= -\frac{\exp(\widetilde{\theta_3})}{(1+\exp(\widetilde{\theta_3}))^2(1+\exp(-\widetilde{\theta_2}))}$$

$$\frac{\partial\omega}{\partial\widetilde{\theta_2}} = \frac{\partial\omega}{\partial\widetilde{\theta_3}} = \frac{\partial\alpha}{\partial\widetilde{\theta_1}} = \frac{\partial\beta}{\partial\widetilde{\theta_1}} = 0$$

Joining all the partial derivatives together, we can have:

$$\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_1}} = \exp(\widetilde{\theta_1}) + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\widetilde{\theta_1}} = \omega + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\widetilde{\theta_1}}$$

$$\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_2}} = {y_{t-1}}^2\frac{\exp(-\widetilde{\theta_2})}{(1+\exp(-\widetilde{\theta_2}))^2(1+\exp(-\widetilde{\theta_3}))} + {\sigma_{t-1}}^2\frac{\exp(-\widetilde{\theta_2})\exp(-\widetilde{\theta_3})}{(1+\exp(-\widetilde{\theta_2}))^2(1+\exp(-\widetilde{\theta_3}))} + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\widetilde{\theta_2}}$$
$$= {y_{t-1}}^2\alpha\frac{\exp(-\widetilde{\theta_2})}{1+\exp(-\widetilde{\theta_2})} + {\sigma_{t-1}}^2\beta\frac{\exp(-\widetilde{\theta_2})}{1+\exp(-\widetilde{\theta_2})} + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\alpha}$$

$$\frac{\partial{\sigma_t}^2}{\partial\widetilde{\theta_3}} = {y_{t-1}}^2\frac{\exp(-\widetilde{\theta_3})}{(1+\exp(-\widetilde{\theta_3}))^2(1+\exp(-\widetilde{\theta_2}))} + {\sigma_{t-1}}^2\frac{-\exp(\widetilde{\theta_3})}{(1+\exp(\widetilde{\theta_3}))^2(1+\exp(-\widetilde{\theta_2}))} + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\widetilde{\theta_3}}$$
$$= {y_{t-1}}^2\alpha\frac{\exp(-\widetilde{\theta_3})}{1+\exp(-\widetilde{\theta_3})} + {\sigma_{t-1}}^2\beta\frac{-\exp(\widetilde{\theta_3})}{1+\exp(\widetilde{\theta_3})} + \beta\frac{\partial{\sigma_{t-1}}^2}{\partial\widetilde{\theta_3}}$$

3.3.5. The gradient of lower bound

The gradient of lower bound with respect to variational parameters λ is $\widehat{\nabla_\lambda}\mathcal{L}(\lambda) = \left(\widehat{\nabla_\mu}\mathcal{L}(\lambda)^T, \widehat{\nabla_{vech(L)}}\mathcal{L}(\lambda)^T\right)^T$

$$\widehat{\nabla_\mu}\mathcal{L}(\lambda) = \frac{1}{S}\sum_{s=1}^{S}\nabla_{\tilde{\theta}}h_\lambda(\tilde{\theta}_s)$$

$$= \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial {\sigma_t}^2}{\partial\tilde{\theta}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right)+ LL^T(\tilde{\theta}-\mu)\right)$$

$$\widehat{\nabla_{vech(L)}}\mathcal{L}(\lambda) = \frac{1}{S}\sum_{s=1}^{S}vech\left(\nabla_{\tilde{\theta}}h_\lambda(\tilde{\theta}_s)\varepsilon_s^T\right)$$

$$= \frac{1}{S}\sum_{s=1}^{S}vech\left(\left(\frac{1}{2}\sum_{t=1}^{T}\frac{1}{{\sigma_t}^2}\frac{\partial {\sigma_t}^2}{\partial\tilde{\theta}}\left(\frac{{y_t}^2}{{\sigma_t}^2}-1\right)+ LL^T(\tilde{\theta}-\mu)\right)\varepsilon_s^T\right)$$

3.3.6. The initialization of GVB algorithm

The algorithm parameters were set to the default values employed in Tran, Nguyen and Dao (2021), with slight adjustment. Specifically**,** the initial inputs of the GVB for initialization step are

- Initial 3-vector $\mu^{(0)}$ which is randomly generated from uniform distribution.
- $L^{(0)}$ is 3x3 matrix with the diagonal equals to 0.1. This simple covariance matrix seems to work well in the model.
- Initial variational parameters $\lambda^{(0)} := \left(\mu^{(0)^T}, vech\left(L^{(0)}\right)^T\right)^T$
- Adaptive learning weights $\beta_1, \beta_2 = 0.9$
- Fixed learning rate $\epsilon_0 = 0.002$
- Maximum numbers of iteration 5000
- Threshold $\tau = 2500$
- Rolling window size $t_W = 10$
- Maximum patience parameter $P = 20$
- Monte Carlo Samples $S = 50$
- Mathematical expression of function $h(\theta)\ and\ \nabla_\theta h(\theta)$ are shown in above section 3.3.4. The implementation of these functions can be found in the ***grad_h_function.m*** file.

4. Results

4.1. Parameter estimates for GARCH (1,1)

Table 2: Posterior mean and standard deviation from the MLE, GVB methods and MCMC estimates.

	ω	α	β
MCMC (no transformation) posterior mean	0.0243	0.1078	0.8590
MCMC (no transformation) posterior standard deviation	0.0062	0.0156	0.0215
MCMC (with transformation) posterior mean	0.0212	0.1048	0.8671
MCMC (with transformation) posterior standard deviation	0.0058	0.0149	0.0198
GVB Posterior mean	0.0225	0.1047	0.8642

MLE	0.0221	0.1038	0.8657

It can be seen that all of the methods provide a very similar estimation to the MLE which is the posterior means from the MATLAB GARCH toolbox.

4.2. Convergence and Lower Bound

For MCMC, as will be seen in the following trace plots (Figure 2 to 5), the Markov chains begin converging around the 1000th iteration and start to stabilise around the 2000th. The convergence values are also reasonable, as can be seen from the comparison with estimates using the GARCH toolbox that employs the MLE approach.

For VB, figure 6 shows the lower bound values with respect to the number of iterations in implementations. Interestingly, the lower bound appears to reach the convergence and stay the same after around 2100th iteration.

4.2.1. MCMC without transformation

Figure 2: Trace plots without removing any burn-ins to show convergence.

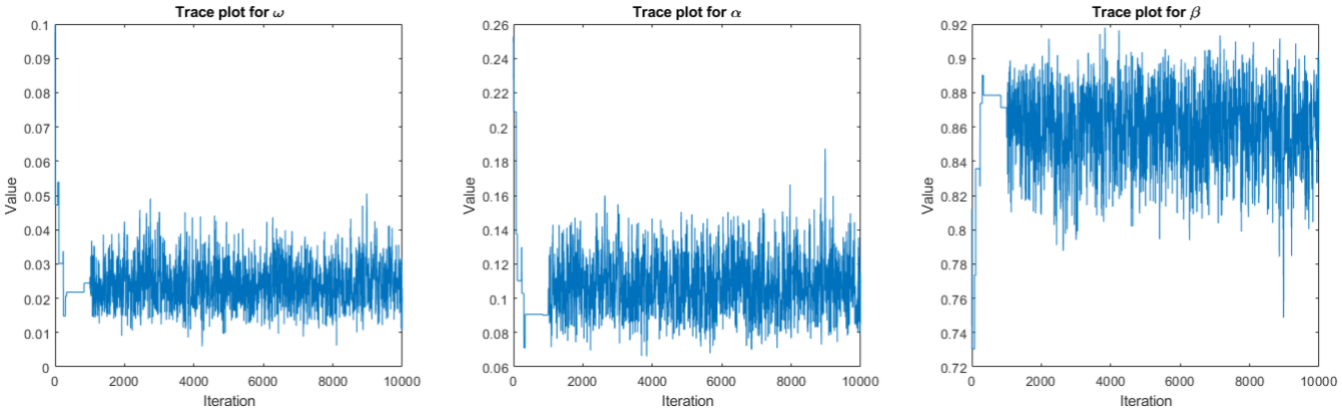
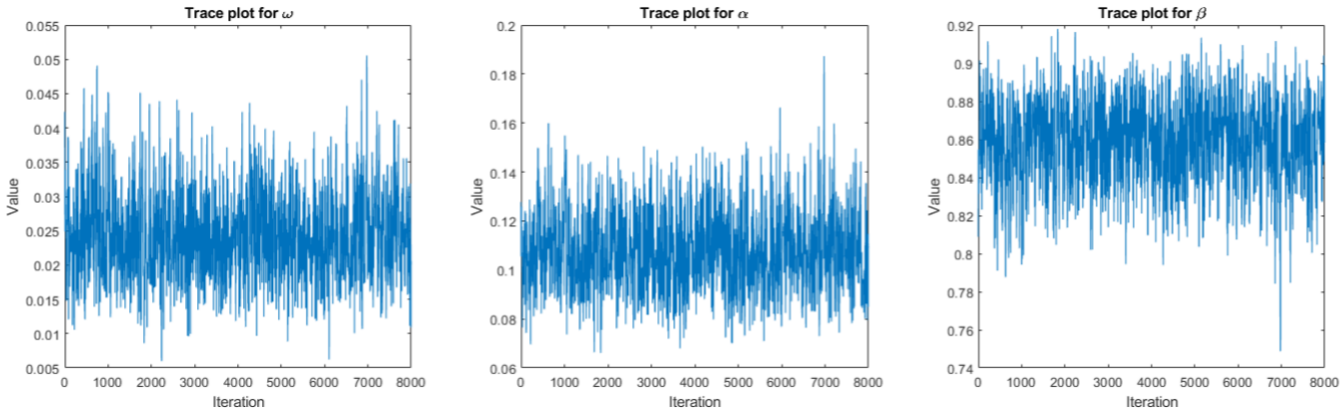


Figure 3: Trace plot after removing the first 2000 burn-ins iterations.



4.2.2. MCMC with transformation

Figure 4: Trace plots without removing any burn-ins to show convergence.

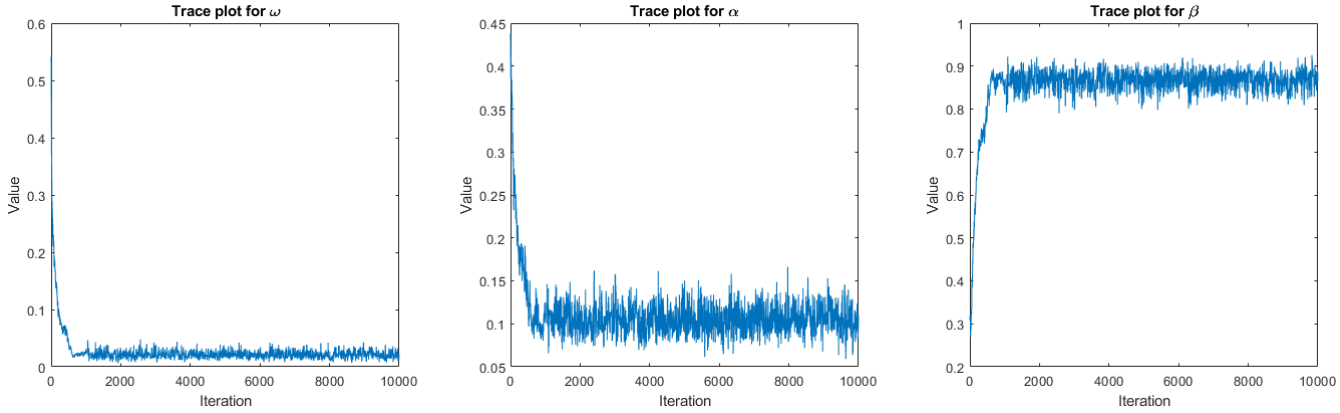
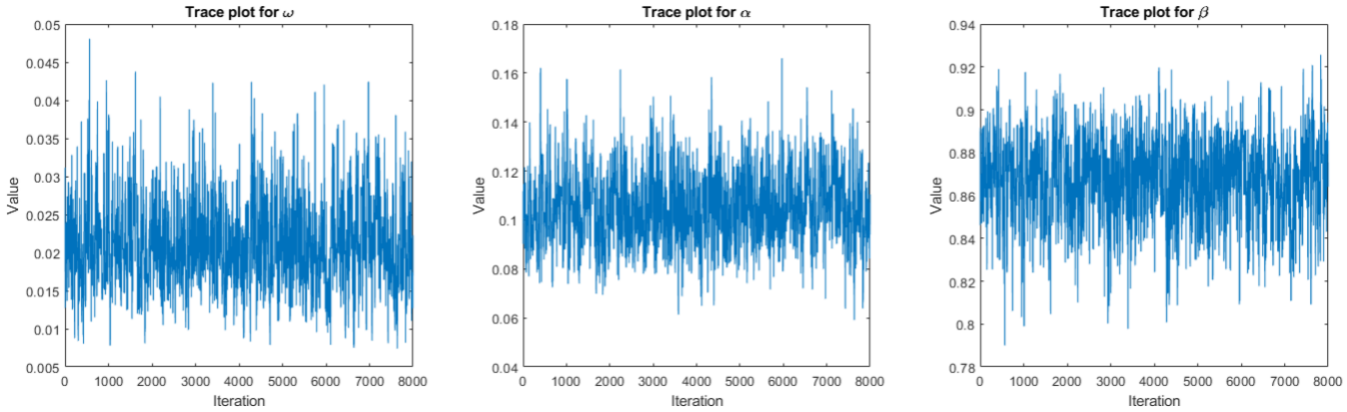
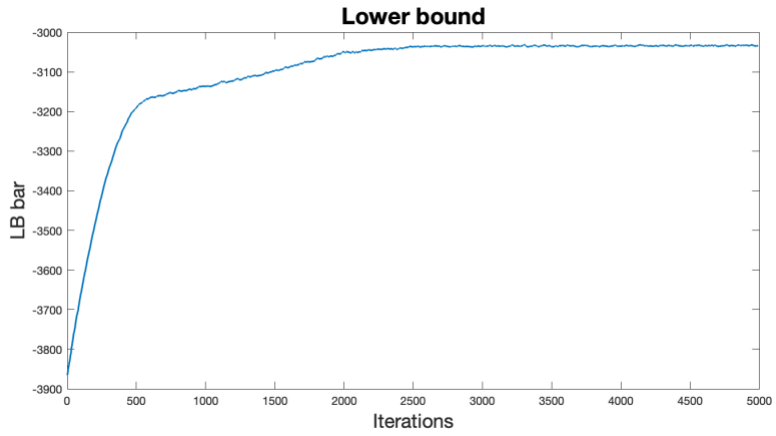


Figure 5: Trace plot after removing the first 2000 burn-ins iterations.



4.2.3. VB lower bound

Figure 6: Moving average lower bounds over iterations for GVB method.



4.3. Forecasting AORD volatility on June 29, 2022

Given that the GARCH model has the conditional variance of:

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

Thus, the next day conditional variance can be rewritten as:

$$\sigma_{t+1}^2 = \omega + \alpha y_t^2 + \beta \sigma_t^2$$

The term σ_t^2, which is instrumental in computing σ_{t+1}^2, represents the conditional variance at time t and depends on all preceding returns and their conditional variances. Hence, we calculated σ_t^2 iteratively from $t = 2$ to n, which provided the most recent conditional variance. This latest variance, when combined

with the current returns and estimated parameters, are used to forecast the next day volatility. The volatility forecasts for June 29, 2022, under each of the 3 implemented approaches are in Table 3.

Table 3: Volatility forecast under MCMC and VB methods.

	MCMC without Transformation	**MCMC with transformation**	**GVB**
Forecasted Volatility	1.2062	1.2215	1.2105

5. Final Analysis and Conclusion

To conclude, it was found that all three parameters have roughly the same estimates for ω, α, β. This in turn provided a similar forecast for volatility on June 29, 2022. However, there are some limitations within these process as MCMC with RWMH is more computationally expensive as each iteration is used to evaluate the model to provide more accurate estimations compared to VB. However, choosing whether to go with higher accuracy or with quicker computation depends on the tasks at hand, as if this was used with high frequency trading, then VB should be chosen as it allows quicker estimations. Future non-technical work includes regularly updating the model to account for new data and the dynamic financial market.

6. References

Fama, E.F. (1970). Efficient Capital Markets: a Review of Theory and Empirical Work. *The Journal of Finance*, [online] 25(2), pp.383–417. doi:https://doi.org/10.2307/2325486.

Roberts, G.O. and Rosenthal, J.S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2), pp.349–367. doi:https://doi.org/10.1198/jcgs.2009.06134.

Takaishi, T. (2018). GARCH Parameter Estimation by Machine Learning. *International Journal of Engineering and Applied Sciences (IJEAS)*, 5(8). doi:https://doi.org/10.31873/ijeas.5.8.05.

Tran, M.N., Nguyen, T.N. and Dao, V.H. (2021). A practical tutorial on Variational Bayes. *arXiv (Cornell University)*.