QBUS3830 Group Assignment

October 9, 2023

1 Background

Volatility (σ_t), defined as the standard deviation of a financial return such as stock returns or exchange rates, is of high interest in finance and banking as volatility is often used to characterize financial risk. More specifically, let $\{P_t,\ t = 0, 1, ...\}$ be a time series of the prices of an asset at the end of each time interval t, the return at time t is defined as

$$y_{t,\text{original}} = 100 \log \frac{P_t}{P_{t-1}}, t = 1, 2, ...$$

The data set `AORD_2012_2022`, available on Canvas, contains the daily return values $y_{t,\text{original}}$ of the Australia All Ordinaries Index (AORD) from 03-Jan-2012 to 28-June-2022. You will be working on this data set for this assignment.

The conditional variance $\sigma_t^2 = \mathbb{V}(y_t|y_{1:t-1})$ is the square of volatility σ_t, given the information up to time $t-1$. As the volatility is unobserved, econometricians often use statistical models in order to understand the volatility dynamics and produce its forecast. One of the most widely-used volatility models is the Generalized Autoregressive Conditional Heteroscedastic (GARCH) model (Engle, 1982; Bollerslev, 1986), which formulates σ_t^2 as a linear combination of the previous squared return and conditional variance. Professor Rober Engle has won the 2003 Nobel price in economic science because of his work in this area. Here we are going to work with a GARCH(1,1) model:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad t = 1, 2, ..., T \\ \sigma_t^2 &= w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad t = 2, ..., T, \end{aligned}$$

where $w,\alpha,\beta > 0$ (to ensure positivity of σ_t^2) and $\alpha+\beta < 1$ (to impose the stationarity condition of the GARCH(1,1) process). Note the return error ϵ_t is assumed to follow a standardize Normal distribution. As can be seen, in the $y_t = \sigma_t \epsilon_t$ equation there is no mean of the return included/modelled, thus you need to **centre the return** before inputting the return into the GARCH model. More specifically, your model input $y_t = y_{t,\text{original}} - \bar{y}_{t,\text{original}}$, where $\bar{y}_{t,\text{original}} = \sum_{t=1}^{T} y_{t,\text{original}}/T$.

The initial σ_1^2 is often set to the sample variance of the time series y_t. The model parameters θ include w, α and β.

Bayesian inference is used in this assignment. Then some prior distribution for θ is potentially needed. As $0 < \alpha,\beta < 1$ and $w > 0$, suitable priors for $\theta = (\alpha,\beta,w)$ are

$$\alpha \sim \text{Beta}(1.5, 10), \quad \beta \sim \text{Beta}(10, 1.5), \quad w \propto 1.$$

You can try and select the hyper-parameters in the Beta distributions. The hyper-parameters in the Beta distributions are selected to encourage α and β to be close to 0 and 1, respectively, as this has been long observed in the financial volatility modelling literature.

The prior selection is up to you. You can use the above suggested prior, or you can use your own choice of the prior. Even, you can choose to not use any prior (uninformative) for all parameters, i.e., $\alpha \propto 1$, $\beta \propto 1$, $w \propto 1$. As long as you can make sure your Bayesian algorithm is working properly with correctly converged chains, then it is fine.

2 Assignment task (50 marks)

Your task is to estimate the GARCH(1,1) model using MCMC or Variational Bayes (VB) and analysis the volatility of the AORD returns.

- (3 marks) Introduction. Write a few paragraphs stating the business problem and summarising your works, etc. Use plain English and avoid technical language as much as possible in this section (it should be for the general audience).
- (5 marks) Data pre-processing and exploratory data analysis (EDA). You need to centre the return data as mentioned in the background section. Then you need to perform an exploratory data analysis for

the return data (calculate and report min, max, skewness, kurtosis, etc.). Comment on the findings. Financial time series data have some unique characteristics that do not share with other type of time series, you're encouraged to incorporate the domain knowledge into your discussion. Therefore, you should search online on the financial volatility and GARCH related literatures to help you prepare the Introduction, data pre-processing and EDA sections.

- (30 marks) Methodology. Estimate the posterior distribution of θ, $p(\theta|y)$ for the GARCH model. You have two options: use MCMC or VB. You can decide your specific choice of the MCMC or VB algorithm, e.g., RWMH, Gibbs sampling, FFVB, etc. Please note here estimating the model with VB is a more challenging task and you might need to do some self study and research on implementing VB, thus using VB is NOT compulsory. However, if your group correctly implements both MCMC and VB, and compare the two methods together for estimating the GARCH model, a **bonus mark of 5** (on top of the 50 marks) will be awarded.

  Please report the estimates of the posterior mean and posterior standard deviation of θ. In addiction to the posterior mean and posterior standard deviation results of θ, the methodology section of your report should include your implementation details, including: your general understanding of the GARCH model (e.g., what the model is doing and why it can help us generate the volatility forecast); how you build the (log) likelihood function and incorporate the prior; how you implement the MCMC or VB algorithms according to the GARCH framework; whether your MCMC or VB algorithm is converged; some interpretations of your results, and your findings.

  Hint: You must impose the the stationarity condition $\alpha+\beta<1$ and the positivity constraint (i.e., $w,\alpha,\beta>0$).

  (i) If you use MCMC, then a naive way to do is to reject the proposal if the stationarity condition and the positivity constraint are not satisfied. However, if you don't tune the algorithm carefully, your Markov chain might take a long time to converge.

  (ii) An alternative is to use the following transformation., while it is up to you on using the following transformation or not for your MCMC. If you use VB, then you must use the transformation.

We have the following transformed parameters $\widetilde{\theta} := (\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3)^\top \in \mathbb{R}^3$ which are unconstrained. The mathematics behind the derivation of the transformation is omitted.

$$\omega = exp(\tilde{\theta}_1), \tag{1}$$

$$\alpha = \frac{exp(\tilde{\theta}_2)\exp(\tilde{\theta}_3)}{1 + exp(\tilde{\theta}_2) + exp(\tilde{\theta}_3) + exp(\tilde{\theta}_2)exp(\tilde{\theta}_3)}, \tag{2}$$

$$\beta = \frac{exp(\tilde{\theta}_2)}{1 + exp(\tilde{\theta}_2) + exp(\tilde{\theta}_3) + exp(\tilde{\theta}_2)exp(\tilde{\theta}_3)}, \tag{3}$$

$$\tag{4}$$

As can be seen, for any values of the $(\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3)$, the resulting w, α, β always satisfy the stationarity condition $\alpha + \beta < 1$ and the positivity constraint (i.e., $w, \alpha, \beta > 0$).

The unconstrained transformed parameters $\widetilde{\theta} := (\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3)^\top \in \mathbb{R}^3$ are easier to estimate the posterior. However, you must report the results in the original scale of θ, i.e. transform back from $\widetilde{\theta}$ to $\theta = (w, \alpha, \beta)$.

- (5 marks) Show the trace plot for each parameter respectively. Justify that your MCMC or VB chains for each parameter have been converged correctly.

  To double check whether you algorithm has reached a reasonable solution, you might compare you parameter posterior means with the parameter estimates from the Matlab GARCH toolbox which uses MLE, see `https://au.mathworks.com/help/econ/garch.html` and `https://au.mathworks.com/help/econ/maximum-likelihood-estimation-for-conditional-variance-models.html`. This is only a suggestion and is not a compulsory part to be included in your report. The Matlab GARCH toolbox pages also include some useful information on understanding the GARCH model.

- (5 marks) Then generate the forecast of the AORD volatility for 29-June-2022 with your estimated model.

- (2 marks) Final analysis, conclusion, limitations and future steps (non-technical).

- Appendix. In the appendix section, you MUST include **three meeting minutes** using the provided Minutes Template on Canvas. More detailed instructions are also given below. You can also put any other materials that you see appropriate into the Appendix section. The Appendix will NOT be counted into the length of the main report and there is no page limit for the Appendix.

3 Submission instructions

1. Due date/time: **Thursday 2-Nov-2023, 23:59 pm**.
2. Weight: 30% of the total mark of the unit.
3. Late submission: Deduction of 5% of the maximum mark for each calendar day after the due date. After ten calendar days late, a mark of zero will be awarded.
4. For the submission, each group should pick up a **group representative** who needs to submit the report and code files.
5. To facilitate your report writing process, a **Report_Instruction** file is also provided on Canvas. The required submissions are:
    - ONE written report (word or pdf format, through Canvas- Assignments-Group assignment report submission). The file should be named as **Group_xxx_report.pdf**, that reports your data analysis procedure and results. You should replace the xxx in the file name with your group ID. In your report, you should also include **the student IDs and names of all group members**.
    - One or multiple code files (Matlab m-file, through Canvas- Assignments-Group assignment code submission). Name the main code as **Group_xxx_code.m**, that implement your data analysis procedure. You should replace the xxx in the file name with your group ID. If you submit more than one code files, the main function of the code files should be named as "**Group_xxx_code.m**". The other code files should be named according to the actual function names, so that **the marker can directly run your code and replicate your results**.

6. The report file **Group_xxx_report.pdf** is a report that describes and summarises your data analysis problem:

    - Describe your data analysis procedure and report the results in detail. The report should be detailed enough so that other business analysts, who are supposed to have background in your field, understand and are able to implement the task.
    - There is no unique or best way of presentation. It's your opportunity to be creative.
    - Clearly and appropriately present any relevant graphs and tables.
    - The page limit is **15 pages** excluding appendix.
    - All the numerical results are reported up to **four decimal places**.
    - You do not need to have the cover page and table of content.

7. The **random seed** in **Group_xxx_code.m** must be fixed, so that the marker expects to have the same (or at least very close) results as you had when re-running your code.

8. Your group is required to submit **three meeting minutes** which are to be included in the appendix section of the your report. You should use the templates provided for preparing agendas and meeting minutes. The meeting minutes template is on Canvas. The more detailed the meeting minutes, the better (who does what next, what has been done by whom, etc.). In case of a dispute within a group, I will use the meeting minutes and/or request for more information to make adjustment to the individual marks. **Should a dispute occurs, please treat each other in a professional and respectful manner**.

9. Later, the unit coordinator will collect **peer feedback** on the performance of each group member. Therefore, it is crucial that each group member is contributing genuinely to the group assignment.

4 Marking Criteria

1. Marks will be allocated based on appropriateness and correctness of introduction, data pre-processing, exploratory data analysis, methodology (modelling and estimation methods), accuracy of your estimation results, your forecast, appropriate discussion and conclusions.

2. As the technical nature of this unit, additional marks might be awarded to projects that have in-depth technicality.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.