STSB6816 Test 1 of 2024

Mathematical Statistics and Actuarial Science; University of the Free State

2024/04/04

Time: 170 minutes; Marks: 45

Instructions

- Answer all questions in a single R Markdown document. Please knit to PDF or Word at the end and submit both the PDF/Word document and the ".Rmd" file for assessment, in that order.
- Label questions clearly, as it is done on this question paper.
- All results accurate to about 3 decimal places.
- Show all derivations, formulas, code, sources, and reasoning.
- Intervals should cover 95% probability unless stated otherwise.
- No communication software, devices, or communication capable websites may be accessed prior to submission. You may not (nor even appear to) attempt to communicate or pass information to another student.
- Use of AI tools must be disclosed and summarised.

Introduction

The data is provided at https://ufs.blackboard.com. It is the monthly log-returns on a company share price over time (assume regular time intervals).

Senior analysts have suggested that you model the log-returns using a normal distribution as trading was stable during this period of interest. However, they point out that the company experienced gradually deteriorating sentiment during this period, depressing returns; along with gradually increasing volatility due to company directors behaving more and more erratically on social media.

Question 1

1.1) Write down the density function and log density function of Y_i where $Y_i \sim N(\mu_i, \sigma_i^2)$. **[2]**

1.2) Now let $\mu_i = \beta_0 + \beta_1 x_i$ and $\sigma_i = \delta_0 + \delta_1 x_i$, where X is an explanatory variable (time in this case) and $\delta_0, \delta_1 > 0$. Derive the **log** likelihood of a sample $(y_1, y_2, \dots, y_n)$ after plugging in these expressions. **[2]**

[Note, for interest, that this is *not* a standard model specification. *It will work* here though, if implemented correctly.]

1.3) Import the data set into your statistical software and explore it visually. You could use a time series plot and discuss what you see. **[4]**

1.4) Fit the model using uniform priors and find posterior mode estimates, or fit the model via the maximum likelihood approach and provide parameter estimates. **[10]**

[Any valid method will receive full credit (here only). You could use Stan's *optimizing*, or simulate the posterior (*sampling*) and then use kernel density estimates of the modes (via *density* perhaps), or try *mle* or *optim*.]

Hint: the parameter estimates you obtain should be (very roughly) about: $\delta_0 = 0.07$, $\delta_1 = 0.002$, $\beta_0 = 0.12$, $\beta_1 = -0.003$. *If you are unable to obtain reasonable estimates in good time then use these values in further questions that rely on them.*

1.5) Obtain a reasonable prediction and 95% prediction interval for the log-return of the next month $(n + 1)$. **[9]**

[Note that the interval will require simulating the posterior distribution of the parameters, or otherwise accounting for their uncertainty, if not already done.]

1.6) Add your prediction and interval to the plot and state whether it seems like a reasonable extension of the observed patterns. **[3]**

1.7) If you were formulating the model yourself, what would you change? Why and how? Would you change the how the variance is modelled or the constraints implemented? Would you change the prior distributions? Would you fit the model differently? **[6]**

1.8) Simulate a new time series, over the same period as your current time series, by plugging the parameter estimates you obtained into the model. Does your series look like the original? What does that tell you about your model fit? **[4]**

1.9) Fit an ordinary linear regression model (ignoring heteroscedasticity). Simulate a new time series, over the same period as your current time series, by plugging the parameter estimates you obtained from the ordinary linear model into the ordinary linear model. Does your series look like the original? Would you say the fit is better or worse? **[5]**