1 Review: Project 11

1.1 Strength

The authors undertake a thorough time series analysis of NVIDIA's stock price, highlighting both conceptual and technical strengths in their report. From a conceptual perspective, the topic has strong links to the economy and financial markets, making it an area worth exploring. The analysis uncovers an ARMA(0, 0) + GARCH(1, 1) model with t-distribution errors, providing valuable insights for stock trading decisions. The stochastic volatility model constructed by the authors offers a possible explanation for underlying stock price dynamics. On the technical side, their data analysis employs key techniques covered in class, such as ARMA, GARCH, and stochastic volatility models. Moreover, the report discusses the combination of ARMA and GARCH, which adds depth to the discussion. The report is also well-written and logically structured, making it easy to follow.

1.2 Weakness

In "Introduction" section, authors claim that they will "improve predictive accuracy regarding future stock prices", but in their subsequent analysis, the authors don't prepare a hold-out test set and we also can't find the results of model prediction.

In "Introduction" section and "Exploratory Data Analysis" section, the data and data source are not well explained. For instance, authors miss the definition of "adjusted close price", and it's hard to see how the "adjusted close price" will "provide a nuanced reflection of NVIDIA's valuation, accounting for corporate actions". More importantly, authors don't provide the source of their data, and we can't find their CSV file in their submission. Due to this, we can't run their code and reproduce their results. By the way, the coding style needs to be improved. Authors hard-code the working directory to "/Users/huanglingqi/Desktop/Stats 531 Final Project", which further prevents us from reproducing their paper.

In "Exploratory Data Analysis" section, authors don't explain the huge jump on 2023/05/25 and its potential influence on their model fitting. As can be seen in their data plot, the close price on 2023/05/24 is 305 and the next day close price increases to 379, which is a considerable surge. This may be the reason of low ESS in their POMP model at time 340 (roughly). In section "POMP Model - Local Search", we can find a great decrease in ESS, probably suggesting the bad fit on 2023/05/25.

In section "Definition of Daily Log-return and Summary Statistics", some notations can be corrected and improved. For instance, they define the log return as $\log(X_t/X_{t-1})$, but usually we use the term "return" rather than "log return" [9]. Authors define "k-period log-return" but don't use it in the following parts. Besides, the definition of "k-period log-return" may be wrong. It should be $r_t(k) = r_t + r_{t-1} + \cdots + r_{t-k+1} = \log(X_t/X_{t-k})$.

In "ADF Test for Stationary" section, the ADF test is not correctly used and interpreted. According to [5]'s first comment, the ADF alternative hypothesis "data is stationary" is wrong, since the data may come from a non-stationary process that does not have a unit root. Besides, if p-value is less than 0.01, we should reject null hypothesis.

In "Likelihood Ratio Test" section, the LRT is not well defined and reported. The test statistics approximate the $\frac{1}{2}\mathcal{X}_d^2$ rather than "$\mathcal{X}_d$" [8]. Authors may forget to report the test statistics and its p-values.

In section "GARCH Model Selection", the model selection is not conducted, model building may be problematic, and model diagnosis is oversimplified. We expect authors to use an AIC table to select p and q for GARCH model, but they directly choose GARCH(1, 1) without justification. When checking their codes, we find they directly feed the return into GARCH model without demeaning. This violates the practice in [9] and may negatively influence the GARCH fit. The model diagnosis lacks some details. For instance, they don't show the p-values of coefficients, the test statistics of Shapiro-Wilk test and Jarque-Bera test, QQ plot and ACF plot.

In "POMP Model" section, authors don't leverage what they learn in GARCH analysis. In their GARCH analysis, they find the residuals following a t-distribution are more reasonable, but in their POMP model, they still build the model using normal distribution. (This suggestion comes from the second comment in [6]).

In "POMP Model - Local Search" section, model initialization is not described and the analysis of local search results is not in-depth. From their report, we can't find the initial values in model and how the model performs with these initial values. Besides, the local search results are not well discussed. For instance, the blue line in the loglik plot significantly outperforms other models, but authors don't mention it in their analysis. The blue line's ϕ value differs from the ϕ values of other models, which may contradict authors' statement that ϕ can converge. Another noticeable thing is that the σ_η may not converge. As shown in the "MIF2 convergence diagnostics" plot, some models finally have σ_η values larger than 10. As for the analysis of ESS, the authors ignore the great decrease at time 340. This may be the sign of outlier data points.

In "POMP Model - Global Search" section, search box is not well designed and the analysis miss some important points. The parameter box doesn't match the findings in local search. For instance, the start values of μ_h is set to be in range $(-1, 0)$, but if we look at the "loglik-mu_h" pair plot in local search, we know μ_h may be in range $(-6, 0)$. The authors showcase a summary of log likelihood but don't analyze it. The difference between maximal log likelihood and 3rd Qu. log likelihood is 9 log units, and this deserves discussion. We can see that the values of ϕ spread over a wide range, and it's better to have a discussion based on its poor man's confidence interval or profile likelihood confidence interval.

Some small issues: (1) In section "Residual Analysis for ARMA(0, 0)", the result of Shapiro test is not shown; (2) In "GARCH Model Selection" section, "ϵ_n" is the white noise, but authors claim σ_n is the white noise. In GARCH Model's definition, the usage of p and q violates

tradition. For instance, it should be $\sum_{j=1}^{p} \alpha_j \tilde{X}_{n-j}^2$ rather than $\sum_{j=1}^{q} \alpha_j \tilde{X}_{n-j}^2$. (3) Authors' explanation of the difference between ARMA and GARCH, and why we should combine ARMA and GARCH is not very clear. We recommend the source [10] where a short and to the point answer is given. (4) In the "Reference" section, some references are incomplete. For instance, the reference "[5] Stats 509 Lecture Notes Chapter 10" lacks author and link. Reference "[9] https://ionides.github.io/531w24/midterm_project/project15/comments.html" lacks title and author.

1.3 Points for Improvement

(1) Authors should report the diagnosis result of GARCH models in details. For instance, they can follow the practice of [3] and [2]; (2) More stochastic volatility models can be tested, such as the simple stochastic volatility model in [3], simplified POMP model and force negative POMP model in [4]; (3) The interpretation and diagnosis of global/local search results for POMP model should be improved. For instance, authors can construct profile likelihood confidence interval according to the suggestion in [9]. (4) Authors should provide the CSV file of the data and RDS file of the experiment to improve the reproducibility.

2 Review: Project 14 (Revised)

2.1 Strength

The authors' report on tuberculosis (TB) in the US offers both conceptual and technical insights. While it doesn't reach a conclusion on the topic, the subject matter warrants further investigation and discussion. The report introduces an innovative SEIRS model with gamma noise and a declining β, which presents an intriguing approach that could lead to deeper exploration.

2.2 Weakness

In "Introduction" section, some necessary references are not included. Authors try to introduce tuberculosis (TB), how it spreads and its impact, but their statement lacks the support of evidence. We suggest that authors can quote the report from WHO [11] or the Wikipedia [12].

In "Dataset" section, it's not clear how authors obtain the data from the link [7]. In this link, only a website table is available and we can't find the download link of CSV file. It's possible that the CSV file in their submission is manually crafted. If this is true, authors may can't ensure the integrity of their data, because entering data by hand is error-prone.

In "Dataset" section, the data analysis may be problematic. We list some issues as followings: (1) Authors argue that "this proportionality suggests a consistent death ratio over time, which may indicate the effectiveness in treatments", but actually, the death rate is declining. (2) The statement "similar to the plot for number of TB cases and deaths, this may suggest that as fewer people get TB, fewer people die from it" may be misleading. The low death rate may be the result of advances in medical technology rather than the outcome of few infected patients. (3) It's meaningless to plot "number change" and "rate change" together. The rate change and number change are reported per year, and in such a short time slot, the population may not change much, so the rate change will be highly correlated to number change. We can also see this from their plot, where most rate change points overlap with the number change points. (4) In Figure 4, the percentage change ranges from -20% to +12% rather than "from -30% to +10%".

In "Explore Periods" section, authors conduct analysis on raw data but in the subsequent ARIMA model selection, they fit the model using differenced data. The period analysis based on raw data will be meaningless.

In "ARIMA model building and model diagnostics" section, they directly copy the code from STATS 531 2024 mid-term project 06 [1] without any reference. The function "model_selection_table" (line 223-377 in their blinded.Rmd) is the copy of function "model_selection_table" (line 278-432) from [1]. The function "build_and_diagnose_model" (line 390-450 in their blinded.Rmd) is the same as function
"build_and_diagnose_model" (line 117-183) in [1]. Interestingly, they forget to copy the functions "simulation_arima" (line 185-217) and "simulation_sarima" (line 219-276) in [1], and this will make the function "model_selection_table" crash if we set its parameter "simulation_time" to be larger than 0. These functions are written by Yicun Duan for his mid-term project. We notice that one of the authors is Yuxi Chen (they write their names in the blinded.Rmd file), the teammate of Yicun Duan in the mid-term project. However, this can't be the reason for them to copy the code without reference, especially when Yuxi Chen wrote less than 10 lines of codes and less than 100 words of report for mid-term project 06.

In "POMP - Stochastic Model" and "POMP - Adding stochasticity to compartment transitions" sections, some formulas may be incorrect. We correct these formulas according to their code: (1) $\frac{\mathrm{d}S}{\mathrm{d}t} = \mu_{RS}R - dw(t) \times (\beta - \beta_t(t - 1952))\frac{I}{N} \times S$; (2) $\frac{\mathrm{d}E}{\mathrm{d}t} = dw(t) \times (\beta - \beta_t(t - 1952))\frac{I}{N} \times S - \mu_{EI}E$; (3) $\tilde{S}(t + \delta) = \tilde{S}(t) - \mathrm{Binomial}(\tilde{S}(t), 1 - \exp(-dw(t) \cdot \beta \cdot \frac{I(t)}{N(t)}\delta)) \rightarrow \Delta N_{SE}(t) = \mathrm{Binomial}(S, 1 - \exp\{-dw(t) \times (\beta - \beta_t(t - 1952))\frac{I}{N}\Delta t\})$; (4) $\tilde{E}(t + \delta)$, $\tilde{I}(t + \delta)$, $\tilde{R}(t + \delta)$ and $\tilde{H}(t + \delta)$ need similar modification.

In the "POMP - Model" section, we can see high variance in the simulation outcome, especially during the period 1990-2020. Is this the result of gamma noise? If so, authors may need to consider removing or reducing the gamma noise.

In the "POMP - Model" section, authors claim they control the overdispersion through parameter k, but in "POMP - Process" section, they state that the overdispersion is realized through introducing the gamma noise. These are confusing arguments.

In the "POMP - Model" section, authors fail to conduct local search, and their "best" parameters are the start values of the local search rather than the local search outcome. When seeing their code, we find that their local search is just one-time run of the MIF2 function which produces only one line in the result plot. Besides, they set the initial parameters in local search to their "best" parameters. If we see their result plot, we can observe that the local search finds a different set of parameters. For instance, the β goes to about 15, instead of 4.342608e+01 reported by the authors.

Some small issues: (1) In "POMP - Model Parameters", parameter β, μ_{EI}, and etc are not coded in Latex. (2) In "POMP - Process" section, the force of infection should be μ_{SE} rather than μ_{IR}. (3) In "POMP - Process" section, the definition of gamma noise is not given. We also don't know why we should use gamma noise rather than normal noise. (4) In "Remarks and Conclusion" section, the analysis of ARIMA model selection is wrong. We should choose model whose smallest root is far from 1 rather than "close to 1".

Other issues: (1) No global search; (2) The diagnosis of local/global search is not performed. (3) Some references are not used. For instance, reference [2], [3], [4] and [6] can't be found in their paper.

2.3   Points for Improvement

(1) Authors should finish the work by conducting and analyzing local/global search; (2) Author should try some simple models like SEIR and SEIRS without gamma noise. The data shows a simple declining trend without many peaks and troughs, and in this case, the data may be fitted using a simpler model; (3) The article is not well written and organized. It takes time to understand authors' idea by viewing their code. (4) Last but not least, authors should never plagiarize someone else's code and must always provide proper citations.

References

[1] ANONYMOUS AUTHORS. Mid-term Project 06 (2024). `https://ionides.github.io/531w24/midterm_project/project06/blinded.html`. Accessed: 2024-04-27.

[2] ANONYMOUS AUTHORS. Project 07 (2022). `https://ionides.github.io/531w22/final_project/project07/blinded.html`. Accessed: 2024-04-27.

[3] ANONYMOUS AUTHORS. Project 14 (2022). `https://ionides.github.io/531w22/final_project/project14/Blinded.html`. Accessed: 2024-04-27.

[4] ANONYMOUS AUTHORS. Project 22 (2022). `https://ionides.github.io/531w22/final_project/project22/blinded.html`. Accessed: 2024-04-27.

[5] ANONYMOUS AUTHORS. The Comments on Project 05 (2022). `https://ionides.github.io/531w22/midterm_project/project05/comments.html`. Accessed: 2024-02-29.

[6] ANONYMOUS AUTHORS. The Comments on Project 07 (2022). `https://ionides.github.io/531w22/final_project/project07/comments.html`. Accessed: 2024-04-27.

[7] CDC. Tuberculosis. `https://www.cdc.gov/tb/statistics/reports/2020/table1.htm`. Accessed: 2024-04-27.

[8] IONIDES, E. L. STATS 531 Lecture 05. `https://ionides.github.io/531w24/05/index.html`. Accessed: 2024-01-15.

[9] IONIDES, E. L. STATS 531 Lecture 16. `https://ionides.github.io/531w24/16/index.html`. Accessed: 2024-04-26.

[10] STACKEXCHANGE. What is the difference between GARCH and ARMA? `https://stats.stackexchange.com/questions/41509/what-is-the-difference-between-garch-and-arma`. Accessed: 2024-04-27.

[11] WHO. Tuberculosis. `https://www.who.int/health-topics/tuberculosis#tab=tab_1`. Accessed: 2024-04-27.

[12] WIKIPEDIA. Tuberculosis. `https://en.wikipedia.org/wiki/Tuberculosis`. Accessed: 2024-04-27.