Review of "Estimated Covid-19 burden in Spain: ARCH underreported non-stationary time series"

Reviewer 1

We thank the reviewer for their critical reading of our work and their comments and suggestions, responded below.

Reviewer Comment C 1 — Title, abstract, keywords weren't fitted together. The main approach ARCH wasn't interested within abstract in methods paragraph, while in title and keywords appear as main proposed approach.

Reply: We agree with the reviewer. The *Methods* paragraph in the abstract has been updated to:

"**Methods:** In this work, we explore the performance of Bayesian Synthetic Likelihood to estimate the parameters of a model based on AutoRegressive Conditional Heteroskedastic time series capable of dealing with misreported information and to reconstruct the most likely evolution of the phenomenon. The performance of the proposed methodology is evaluated through a comprehensive simulation study and illustrated by reconstructing the weekly Covid-19 incidence in each Spanish Autonomous Community."

The *keywords* have also been updated, replacing GARCH to ARCH, which is the main proposed approach as pointed out by the reviewer.

Reviewer Comment C 2 — The problems and objectives don't mentioned clearly in abstract.

Reply: To clarify, the following sentences have been added to the abstract:

"XXXXXXXXXXXXXXXXX"

Reviewer Comment C 3 — "Estimation" is not fitted with time series, while "Forecasting" or "prediction" is more appropriate.

Reply: We agree with the reviewer that "Estimation" is referred to parameters and "Forecasting" is usually used in relation to values of the series. We have carefully reviewed the manuscript and changed the wording accordingly. TODO: S'ha de fer!

Reviewer Comment C 4 — The measurements of error should be used such as MAPE, RMSE, ans others.

Reply: Although the main objective of this work is not in forecasting future values but in reconstructing the most likely values of the hidden process, error measures like those proposed by the reviewer can also be computed by taking the registered values as reference. Therefore, following the reviewer's suggestion, the following paragraph has been added to the manuscript (page XXXXXX):

"XXXXXXXXXXXXXXXXX"

Additionally, the table with specific RMSE and MAPE for each Spanish region has been added as Table SXXX in the Supplementary Material.

Reviewer Comment C 5 — Discussions and conclusions should include more details about ARCH and its results.

Reply: XXXXXXXXXXXXXXXXXXXXX

Reviewer 2

Reviewer Comment C 1 — Good manuscript but the statistics in paper were beyond my expertise to comment.

Reply: We thank the reviewer for their critical reading of our work and their positive comment.

Reviewer 3

Reviewer Comment C 1 — In this paper, the authors employ a method to deal with under-reported data and apply it to a dataset of COVID-19 cases in Spain. Although the method is of interest, the authors are not providing new methodological tools, but only applying existing ones. Besides, I have the following specific comments:

Reply: We thank the reviewer for their critical reading of our work and their comments and suggestions, responded below.

Reviewer Comment C 2 — The simulation study carried out by the authors is not motivated enough, in my opinion. The method has already been tested by their authors, so why performing this simulation study?

Reply: The method used to estimate the parameters of the model (Bayesian synthetic likelihood) is known and well-tested, but to the best of our knowledge it has never been used in the context of AutoRegressive Conditional Heteroskedasticity (ARCH) underreported models, and this is why its performance is assessed through the simulation study. To clarify, the following paragraph has been added to the manuscript (page XXX):

"Although the estimation method is already known and has been tested before, to the best of our knowledge it has never been used in the context of ARCH time series, and therefore A a thorough simulation study has been conducted..."

Reviewer Comment C 3 — In Figures 1 and 2 I feel that something is wrong with the plot. For instance, in Figure 2, I don't understand the values around 2022-01. Maybe the upper estimation for that time point is greater than the maximum value fixed for the y-axis?

Reply: As the reviewer points out, there is something curious about the values around 2022-01. The question is that in all simulations this particular value is detected as not being underreported ($Y_{2022-01} = X_{2022-01}$), maybe due to the breakout of a new variant with different characteristics around these dates. This is why the estimated and registered values and the 95% credible interval for this value are all equal, as represented in Figure 1 and Figure 2. To clarify, the following sentences has been added to the manuscript (page XXX):

""

Reviewer Comment C 4 — I also feel that there might be some inconsistency between the results highlighted in page 18 and those provided in Table 2 (estimates of q for Aragon and Extremadura).

Reply: We thank the reviewer for noticing these inconsistencies between the estimates provided in the text and in Table 2. All reported values have been carefully revised and corrected when necessary.

Reviewer Comment C 5 — In the Conclusions, there are some comments referred to the covariates included in the model. I would consider moving these results from the Supplementary Material to the main text.

Reply: Following the reviewer's suggestion, the estimates corresponding to the impact of the considered covariates have been included in the main text (Table 2) and removed from the Supplementary Material. TODO: S'ha de fer!!