Review of: "Quantile regression for identifying latent structures in COVID-19 pandemic – Examples from Nepal"

Giorgio Sonnino[1]

1 ULB Université Libre de Bruxelles

Potential competing interests: No potential competing interests to declare.

Summary of the Work

The aim of this manuscript is to determine the path of the daily infections traced during the initial phase (characterized by an exponential trend), the lockdown period, and after the vaccination campaign. To this end, on the base of data for Nepal ranging from the period January 2020 – February 2023, the author proposed two quantile regression models based. This investigation is motivated by the fact that there are Countries, such as Nepal and India, with underreporting of daily infections and deaths.

Main Results Obtained

- The author made several comparisons concerning the latent behavior of the daily incidence of COVID-19 in different countries with different data qualities.

- According to the author, quantiles and quantile regression are more robust with respect to underreporting.

General Considerations

I read the work with interest. However, there are two aspects that should be further explored:

- Quantile regression models can be useful for analyzing COVID-19 infection data, but they also have some limitations. the author has not discussed these limitations exhaustively;

- The list of references is not exhaustive and should be completed.

- The two models developed by the author have limited predictive power.

The following tips are meant to fill some gaps.

Suggestions

1) One of the main limitations associated with using quantile regression models for COVID-19 analysis is the following. Quantile regression models assume linearity between predictors and the quantiles of the response variable. However, this assumption may not hold true in complex and nonlinear systems like infectious diseases. COVID-19 infections can be influenced by various factors, and the relationship between predictors and infection rates may not be linear. The author is asked to discuss this point.

2) Another intrinsic limitation of a quantile regression model stems from the fact that it has primarily been designed for establishing associations between predictors and response variables, rather than inferring causal relationships. While they can identify factors that are associated with COVID-19 infections at different quantiles, they may not provide a complete understanding of the underlying causal mechanisms. We may object that the author's models are not suitable to analyze causal inference effects. To investigate these effects, other methods are more appropriate (such as *randomized controlled trials* or *structural equation modeling* are more appropriate for establishing causality). The author is asked to dispel this possible objection.

3) Another point that deserves to be discussed is the fact that, like any regression model, the selection of appropriate predictors is crucial for accurate analysis. In the context of COVID-19 infections, identifying relevant predictors can be challenging due to the complex nature of the disease. Incomplete or inadequate inclusion of important predictors in the model can lead to biased and unreliable results. What is the author's opinion on this?

4) As mentioned above, another objection that may be raised in this work is the fact that the author's models have quite limited prediction power. More precisely, as known, quantile regression models focus on estimating the conditional quantiles of the response variable rather than the entire distribution. While this can provide valuable insights into specific quantiles, such as the median or extreme quantiles, it may not capture the full variability and complexity of COVID-19 infections. Please, discuss this important point.

5) Other modeling approaches, different from the quantile values and quantile regression, might be more suitable for capturing the dynamics and making accurate predictions. These models use different methodologies such as time series analysis, machine learning algorithms, or study of the dynamics of the spread of SARS-CoV2 infection governed by stochastic partial differential equations. Below, the author can find some suggestions in this regard:

[1] Saswat S. *et al*., *Time Series Analysis of COVID-19 Data to Study the Effect of Lockdown and Unlock in India*, J Inst Eng India Ser B, **102**(6): 1275 (2021).

[2] Fuad A. C. *et al*., *Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning*, Journal of Virological Methods, **301**, 114433 (2022).

[3] Sonnino G., Peeters P, and Nardone P., Modelling the Spread of SARS-CoV2 and its variants. Comparison with Real Data. Relations that have to be Satisfied to Achieve the Total Regression of the SARS-CoV2 Infection*, European Society of Medicine* (ESMED), (2022). Available online:

https://esmed.org/MRA/mra/article/view/2969

[4] Sonnino G., Mora F., and Nardone P, A Stochastic Kinetic Type Reactions Model for COVID-19, MDPI-Mathematics: Math Mod and Ana in Bio and Med, **9**(11), 1221 (2021). Available online:

https://doi.org/10.3390/math9111221

Conclusions

In my opinion, the work is interesting and deserves attention. However, concerning the regression models proposed by the author, it is important to consider the above-mentioned limitations and complement the analysis with other approaches and methodologies to gain a comprehensive understanding of COVID-19 infections (or, at least, provide a brief discussion of them). I encourage the author to heed the above suggestions.