Review of: "A Multi-factor Model of COVID-19 Epidemic in California"

Amitava Datta[1]

1 University of Western Australia

This paper attempts to build a model for cumulative Covid infections from a purely data science perspective. This means, the author does not assume any infection model similar to the compartmental or agent-based models, rather use available data on population density, income and Gini. The model is simple and interesting, and the results are not unexpected. For example, to quote from the paper: "Correlation with population remains nearly the same as the state-wide result, but the other factors vary greatly. For the bottom half (fewer cases of infected people), population density (size) and income (Gini) are more strongly correlated. Specifically, lower population density and lower inequality (lower Gini) resulted in a lower number of infections. Cumulative cases are negatively correlated with size."

There are two important factors, if considered, will improve the quality of the paper in my opinion. First, the same model could be validated with data from other parts of the world, or at least from other states of the USA (the author has claimed that the model is applicable anywhere, to quote, "This approach is scalable – it can be applied on a country, county, or local level whenever data are available.", and hence more validation is required.). That will show how robust the model is. Secondly, there may be other important factors that are responsible, like mass transit systems (it is known that NY City underground system was a major cause of spread of Covid, in particular among the economically backward population). The testing of the model on other populations may result in modification of the model, thus making it more robust. So, I would consider this as a preliminary work that requires strengthening.