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

Hiroyuki Matsuda[1]

1 Yokohama National University

Potential competing interests: No potential competing interests to declare.

In general, it is not surprising that correlations are higher with more explanatory variables; one should test whether a more complex model is a better fit than a simpler model, e.g., using AIC. The authors have not done so. In addition, there are fatal flaws in the statistical methods, such as not standardizing each variable (and thus relying almost exclusively on the population with large absolute values).

Fig. 1. The number of infected cases per population should be shown, not the total number of cases.

Fig.2: The meaning of the blue curves should be explained. I suspect that the regressions of the Income and Gini coefficients appear to show the maximum values, not the whole data, which is not well explained.

In general, such correlation analysis cannot prove causality; it can only show correlation.

Equations for Infection I(x): x should be P(x)

Equations for Duration D(x): 2,765 should be 2,765 D(x)

Denominator of P(x): It must be summed AFTER normarization.

Fig. 3: The model appears to be highly predictive, but in essence, it fits well because of the high correlation between the number of infections and the population, and the high absolute value of the population. It is not much different from a single correlation with population.

fig. 5: The horizontal axis is a continuous variable and this graph is odd. If the frequencies are tabulated for some interval of the abscissa, it should be a columnar graph, not a scatter plot.
It is usually expected to follow a normal distribution, which is not an interesting finding. In addition , the author did not test for significant deviations from the normal distribution.
Table III: The author should write the meaning of the numbers in brackets (does it mean not significant?)