1. **Title:** Prediction bias in times of Covid-19: Use of a linear infection-prediction model when the true model is exponential

2. **Authors and affiliations and contact emails:** Ritwik Banerjee (Indian Institute of Management Bangalore, India), Joydeep Bhattacharya (Iowa State University, USA), Priyama Majumdar (Indian Institute of Management Bangalore, India). Emails: ritwikbanerjee@iimb.ac.in, joydeep@iastate.edu and priyama.majumdar@iimb.ac.in

3. **Abstract:** The set of measures designed to halt the unrelenting transmission of COVID-19, prescribed by the World Health Organization and widely disseminated by local governments, include frequent washing of hands, use of hand sanitizers and face masks, social distancing and self quarantine, in case of potential exposure to COVID-19. Most countries in advanced stage of the disease and have now locked down parts of (or even entire) cities. Despite the widespread lockdown, there is a sense that people in general and policy makers in particular realized the scale of the issue a little too late, which resulted in delayed action. One possible reason for such delay is the mental model we turn to when estimating the likelihood of infection in the future. In the case of COVID-19, the true underlying data generating process governing the growth of coronavirus cases is exponential in nature but policy makers may end up delaying their action, if they apply a linear model to predict the scale of the problem in a week or two’s time. We conjecture that people in general use linear models to predict the future when the true underling DGP is in fact exponential in nature, resulting in systematic underprediction or behavioral bias. Our second related conjecture is that those with a larger bias are more likely to underestimate the danger and less likely to observe the prescribed best practices by W.H.O.

4. **Data description:** The incentivized experiment will be conducted on Amazon’s Mechanical Turk, an online platform. We hope to recruit about 600 participants form over 50 countries, based on statistical power considerations. In a given task, participants will be shown three, weekly data points of the number of Coronavirus-positive cases from an unnamed, but real country. They will, then, be asked to predict the number of Coronavirus-positive cases for Weeks 4 and 5, the actual numbers of which we know. We wish to know how close participants' forecasts are to the true numbers. Following this, a short survey will be administered to capture the participants’ behavior during the ongoing pandemic; additional demographic information will also be collected. The totality of information collected will allow us to test a) whether people systematically underpredict the 2-week ahead number of positive cases during an ongoing epidemic (prediction-bias) b) whether such biases are associated with a more or less lax attitude towards the afore discussed W.H.O measures. We introduce a further randomized treatment where some participants are shown a graphical (as opposed to numeric) representation of the exponential growth trajectory and thereafter, their predictions are elicited. The randomized experiment is designed to test whether representation of data in terms of raw numbers of graphs help achieve better prediction accuracy.

5. **JEL codes for the project:** I12, I18, D91

6. **Key-words:** COVID-19, prediction bias, public health, W.H.O guidelines