Truth or Dare? Detecting Systematic Manipulation of COVID-19 Statistics
Online Appendix

Fatih Serkant Adiguzel,* Asli Cansunar,† Gozde Corekcioglu ‡

August 29, 2020

Contents

*Email: serkant.adiguzel@duke.edu. Department of Political Science, Duke University

†Email: asli.cansunar@nuffield.ox.ac.uk. Department of Political Science and International Relations and Nuffield College, University of Oxford.

‡Email: gozde.corekcioglu@khas.edu.tr. Department of Economics, Kadir Has University.

1 Digit-Pair Tests

Digit-pair tests are based on a similar intuition to the last-digit test. Under a natural number generation process, the last two digits of any number are identical with a probability of 0.1; similarly, the distance between the last two digits is 1 with probability 0.2; finally, the distance between the last two digits is greater than 1 with a theoretical frequency of 0.7. Given that people avoid repetition, prefer serial sequences, and select pairs of distant numerals relatively infrequently when they are asked to produce numbers, any statistically significant deviation of the last two digits from these parameters should be taken as an indicator that there has likely been human manipulation of the data.

Figure A1 examines the patterns of the last two digits of the daily reported total of COVID-19 cases. Recall that the last two digits of a number are identical with a theoretical probability of 0.1, the distance between the last two digits is 1 with probability 0.2, and the distance between the last two digits is greater than 1 with a theoretical frequency of 0.7. We investigate whether the observed occurrences diverge from the theoretically expected frequencies. Ticks represent the 95% confidence intervals for each tested pattern, and the dots report the observed frequency of each pattern in the data from the nine countries under analysis.

Overall, we can plausibly state that the last-digit and last-two-digits analyses complement each other and provide consistent results. In particular, numbers from the US follow patterns perfectly consistent with human biases in number generation: lack of digit repetition, an excess of adjacent digits, and a dearth of distance numerals.

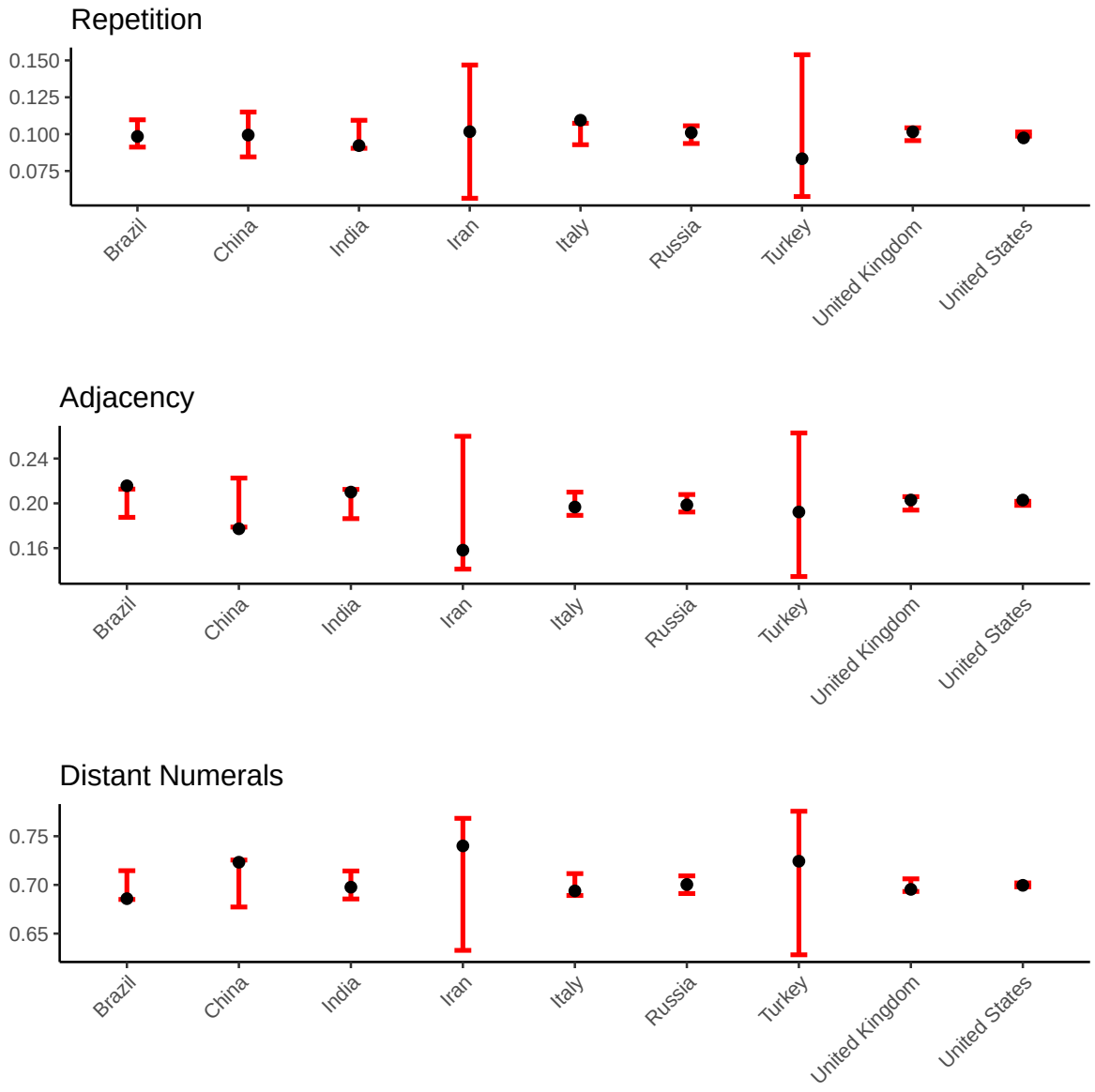

Figure A1: Frequency with which the last two digits are identical, adjacent, and have distance greater than one.

2 Cumulative Cases in China

In this section we provide additional complementary evidence on Chinese statistics. We refrain from imposing any rules on the data (such as dropping observations upon 7 consecutive days of no new cases) and take a closer look at the raw data as disseminated by the Chinese authorities. Figure A2 exhibits the last digit frequencies of daily announced cumulative confirmed cases before April 9. We detect several interesting patterns that might have remarkable implications

in the Chinese culture.

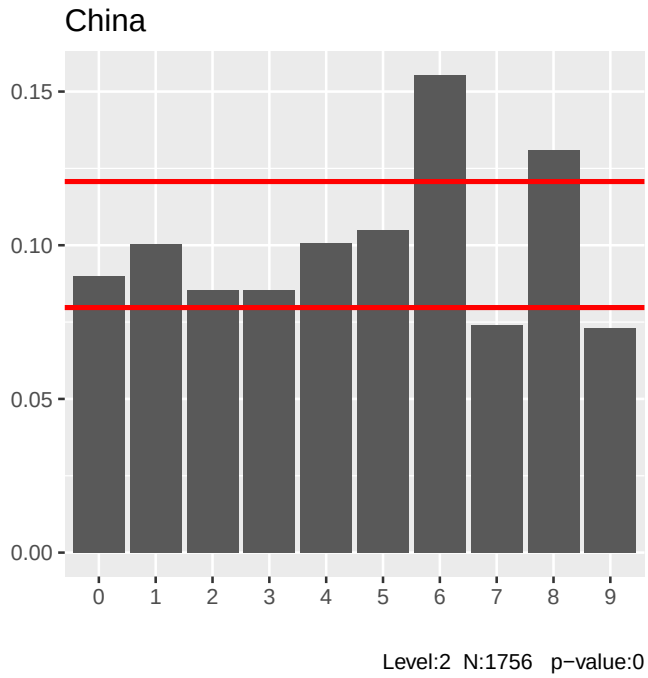

Figure A2: Frequency of last digits for cumulative COVID-19 cases in China. Administrative level at which data is used, number of observations, and p-values are reported. Administrative levels correspond to 1= national, 2 = state/province, 3 = sub-state/sub-province.

Given that numbers that are considered to be 'lucky' in Chinese culture often affect economic, and business transactions, an overrepresentation of 6 and 8 in the last digits of the confirmed COVID-19 cases suggests that the data were manipulated in a meaningful direction (Shum, Sun and Ye, 2014; Yang, 2011). Consistent with this is that the two underused digits, 7 and 9, also have interesting connotations. In Chinese, the number nine sounds like "long-lasting", which is an undesirable characteristic for a pandemic, and number seven, which sounds like "to deceive", is considered an unlucky number.

3 Additional Figures

In this section we include screenshots of daily public announcements of COVID-19 statistics by the primary official sources in China, Russia and Turkey. While the Chinese government reports both daily cumulative and new COVID-19 cases, Russian and Turkish statistics emphasize daily

new COVID-19 cases, as illustrated in Figures A3-A6 below.

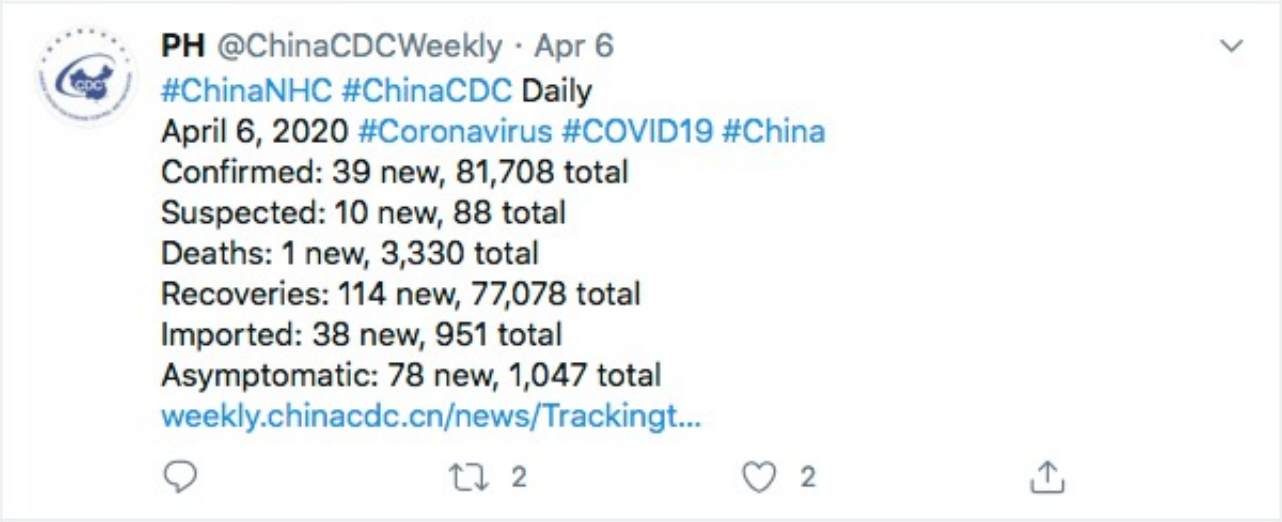

Figure A3: Screenshot taken from the official Twitter account of the Chinese Center for Disease Control and Prevention. The data are provided by the Chinese National Health Commission. Source: `http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm`

Figure A4: The image is taken from the Instagram account of `stopcoronavirus.rf`, the official Russian Health Ministry website dedicated to the coronavirus pandemic.

Figure A5: Screenshot of a tweet by Turkey's Health minister, who releases daily official coronavirus statistics from his Twitter account.

Figure A6: Anadolu Agency, a state-run news agency, tweets the daily official coronavirus statistics in both Turkish and English.

References

Shum, Matthew, Wei Sun and Guangliang Ye. 2014. "Superstition and "lucky" apartments: Evidence from transaction-level data." *Journal of Comparative Economics* 42(1):109–117.

Yang, Zili. 2011. ""Lucky" numbers, unlucky consumers." *The Journal of Socio-Economics* 40(5):692–699.