Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Data misreporting during the COVID19 crisis: The role of political institutions

Antonis Adam, Sofia Tsarsitalidou

1 University of Ioannina, Ioannina, Greece
2 Hellenic Open University, Greece
3 University of Patras, Patras, Greece
4 University of Piraeus, Piraeus, Greece

A R T I C L E   I N F O
Article history:
Received 24 November 2021
Received in revised form 1 February 2022
Accepted 1 February 2022
Available online 7 February 2022

JEL classification:
D72
P16
I18

Keywords:
Democracy
Data misreporting
COVID19

A B S T R A C T
We use Benford’s law of first digits to determine whether there is evidence of data misreporting in the total COVID19 reported cases across countries. We try to model the differences in the Mean Absolute Deviation of actual data from those predicted by Benford’s law to indicate the factors that lead to data misreporting using regression analysis. Using the Instrumental Variable model of Lewbel (2012) and Settler Mortality as an external instrument for democracy, we show that autocratic countries are more likely to misreport the COVID19 cases.

1. Introduction
During the COVID19 pandemic, governments worldwide faced serious incentives to misreport the actual number of patients. For example, travel restrictions were determined by the number of recorded cases in the targeted country. Similarly, all governments had incentives to appear efficient in dealing with the crisis, both to their “electorate” and external observers. Likewise, government response and effectiveness in dealing with the pandemic will be critical to all future elections.1

What are then the country characteristics which are related to data misreporting? First, it is natural to expect that countries with poor quality of political institutions are more likely to resort to methods of data manipulation: the absence of effective checks and balances, and powerful opposition, and the need of autocrats to project power both within the country as well as abroad, make autocracies more likely to try report manipulated data.

In the present paper, we exploit Benford’s law of first digits to answer the above question.2 To isolate the impact of strategic misreporting from other reasons, we use the Lewbel (2012) IV model to estimate the causal effect of the political regime on the reliability of the reported COVID19 cases. We find a negative association between the level of democracy and data misreporting, suggesting that data misreporting might be guided by the governments’ wishes.

2. Benford’s law and COVID19 data

Any number can be written in scientific notation as $S(x) \cdot 10^k$, where $S(x)$ is the significand and $k$ is an integer. Following Hill (1998), Benford’s law states that if random samples of numbers are taken from distributions selected at random, the significand’s first digit frequencies of the combined sample will converge to Benford’s distribution, i.e., the associated probability that digit $d$ occurs as the leading digit of $S(x)$ is given by

$$P(d) = \log_{10} \left( 1 + \frac{1}{d} \right)$$

The above states that in a collection of random numbers, 1 will appear as the first digit with a probability of 30.1%, 2 will appear with a probability of 17.61%, etc. So then, significant deviations of the actual from the theoretical distribution are attributed to data misreporting.

According to Nigrini (2015), Benford’s law can be applied to numbers (not, e.g., labels) data that represent the sizes of events...
Fig. 1. MADs across the world.

or facts, with no associated built-in minimum or maximum value distributed over several orders of magnitude. To evaluate the extent of data misreporting, we use the Total Number of COVID19 cases in each country, a variable that satisfies all the above conditions.3 Using the Total Number of Cases has an additional advantage. If we assume that data are not just misreported and governments manipulate them in a sophisticated manner, their focus would be on new cases, as this is, typically, the measure used to evaluate government efficiency. Thus, any data manipulation on new COVID19 Cases would make the total cases move erratically. The data we use covers 196 countries from the day of the first recorded case of COVID19 infection until July 6, 2021.4

To measure data misreporting, we compute the Mean Absolute Deviation (MAD), i.e., the mean absolute difference between the expected probability that a particular digit will appear with the realized one.5

Fig. 1 presents the estimated MADs, with darker areas indicating higher MAD.6 According to our computations, countries in the Middle East and Eastern Africa appear to have unreliable data. In contrast, Northern America and EU countries report data more in line with the theoretical predictions. Note that Australia and New Zealand have high MADs. This can be attributed to the fact that these two countries had a low number of active COVID19 cases due to the highly restrictive government measures.7

Failure to comply with Benford’s law provides evidence of data misreporting. Misreporting, however, can be the outcome of several reasons: poor data collection methods, COVID19 test unavailability, and even stringent lockdown policies may explain these deviations (Koch and Okamura, 2020). Our strategy, then, is to use an econometric framework to determine the factors that explain non-compliance with Benford’s law. If differences in MADs are determined by factors that correlate with the government’s data collection ability, or the policies to contain the spread of the disease, non-compliance is not an important issue. However, to the extent that MADs are associated with political variables, data manipulation by the government may explain cross-country differences.

3 In contrast, the number of new cases have a minimum value (i.e., zero) and has a rather low spread. Moreover, the number of new cases is not monotonic, with several ups and downs, making it an inappropriate variable for an application of Benford’s law. The size of total cases, in contrast, is distributed over several orders of magnitude (from zero to several thousands or millions).
4 Data are taken from https://ourworldindata.org/covid-cases.
5 Other measures of conformity with Benford’s law have been also proposed in the literature, e.g., a chi-square test for the difference in the two distributions. However, we opt for the MAD as these alternative measures are sensitive to the number of observations.
6 White areas indicate data unavailability.
7 In the Appendix, we also present the MADs, taking only the periods where there was an increase in the total number of cases. These MADs are highly correlated (0.83) with the MADs reported here.
weak autocratic ruler will be eventually disposed of, especially if he cannot protect the well-being of his citizens and, thus, gain their loyalty. As we want to derive the causal effect of the incentives and constraints placed by the political regime on the government to misreport data, we employ an Instrumental Variable Model as in Lewbel (2012). This method uses external instruments complemented with internal ones to increase efficiency. The internal instruments are constructed from the auxiliary equation’s residuals multiplied by the mean-centered exogenous variables. The external instrument follows Acemoglu et al. (2001), which shows that different colonization strategies created different sets of political and economic institutions. The colonization strategy was influenced by the conditions where colonizers were settled. Hence, we use the log of mortality rates by the first European settlers in colonies (Settler). Our model then takes the following form:

$$MAD_i = X_i’\beta + \delta Democracy_i + \epsilon_i$$

$$Democracy_i = X_i’\gamma + \zeta Settler_i + u_i$$

Where $X_i$ is a set of exogenous regressors, which include an index of quality of government, the share of imports plus exports to GDP (Openness), the Stringency index of government policies during the pandemic, the total number of Hospital beds (per 1000 people), the Cardiovascular death rate, the log of the total population and the log of GDP per capita. More details about the control variables and the data employed are given in Appendix.

### 4. Results

Table 1 presents our results. The results indicate a causal effect between democracy and data misreporting. However, this effect is conditional both on the ability of the government to collect data, as determined by the quality of government indicator, the level of development, and the rest of the variables that control for factors that affect the deviation from Benford’s law (e.g., the intensity of the lockdown policies). Moreover, the estimated effect of democracy is quantitatively significant, as democratic countries have on average a MAD approximately 2.7 points lower than autocracies.

### 5. Conclusions

COVID19 data from autocratic countries can generally be considered unreliable, and if the policy is structured using data from these countries, particular caution should be placed.

### Acknowledgments

We would like to thank Nicholas Giannakopoulos, Elias Mossialos, an anonymous referee and the journal editor for their constructive comments and suggestions.

### Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econlet.2022.110348.

### References

Acemoglu, D., Johnson, S., Robinson, J.A., 2001. The colonial origins of comparative development: An empirical investigation. Am. Econ. Rev. 91, 1360–1401.
Bjørnskov, C., Rode, M., 2020. Regime types and regime change: A new dataset on democracy, coups, and political institutions. Rev. Int. Organ. 15, 531–551.

Hill, T.P., 1998. The first digit phenomenon. Am. Sci. 86, 7.

Karabulut, G., Zimmermann, K.F., Bilgin, M.H., Doker, A.C., 2021. Democracy and COVID-19 outcomes. Econom. Lett. 203, 109840.

Koch, C., Okamura, K., 2020. Benford’s Law And COVID-19 reporting. Econom. Lett. 196, 109573.

Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. J. Bus. Econ. Stat. 30, 67–80.

Nigrini, M., 2015. Detecting fraud and errors using Benford’s law. In: Benford’s Law. Princeton University Press.