


Comment

Benford Law to Monitor COVID-19 Registration Data. Comment on Farhadi, N.; Lahooti, H. Forensic Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak. *COVID* 2022, 2, 472–484

Francisco Morillas-Jurado ¹, Maria Caballer-Tarazona ^{1,*} and Vicent Caballer-Tarazona ²¹ Department of Applied Economics, University of Valencia, 46022 Valencia, Spain; francisco.morillas@uv.es² Department of Corporate Finance, University of Valencia, 46022 Valencia, Spain; vicent.caballer@uv.es

* Correspondence: maria.caballer@uv.es



Citation: Morillas-Jurado, F.;

Caballer-Tarazona, M.;

Caballer-Tarazona, V. Benford Law to

Monitor COVID-19 Registration Data.

Comment on Farhadi, N.; Lahooti, H.

Forensic Analysis of COVID-19 Data

from 198 Countries Two Years after

the Pandemic Outbreak. *COVID* 2022,2, 472–484. *COVID* 2022, 2, 952–953.[https://doi.org/10.3390/](https://doi.org/10.3390/covid2070069)

covid2070069

Academic Editor: Guiseppe Novelli

Received: 10 June 2022

Accepted: 6 July 2022

Published: 12 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

In a recent study published in *COVID* by Farhadi & Lahooti, 2022 [1], the authors claim that the work carried out by Morillas-Jurado et al., 2022 [2] is an example of a poor application of Benford Law (BL) in the context of pandemic. Morillas-Jurado et al., 2022 examined the COVID-19 epidemic specifically during the first wave of the pandemic and document anomalies in the data that occurred in six Spanish regions.

The main argument to Farhadi & Lahooti's criticism is that the regional data sets were too small to assess the conformity of COVID-19 data to BL.

However, in Morillas-Jurado et al., 2022, in order to avoid the sample size problem, we ran a sensitivity analysis just to verify that our decision about the conformity of the BL applied also for our sample size. Specifically, in the Methodology section, within Section 2.3, "Sensitivity Analysis Steps", we explain in detail the steps of the sensitivity analysis devoted to validating our results for a specific number of observations, n .

In particular, as developed in step 2 of the sensitivity analysis, we artificially generated an arbitrary number of perturbations for each series obtained through a Monte-Carlo simulation; specifically, 1000 perturbations for each series.

Therefore, given the observed series $\{x_1^{obs}, x_2^{obs}, x_3^{obs}, \dots, x_n^{obs}\}$, and assuming that the observed values could be not correct, we obtained the i^{th} series modified as $x_k^i = x_k^{obs} (1 + u_k^i)$, $k = 1, \dots, n$ where $i = 1, \dots, 1000$, $\{u_k^i\}_{k=1}^n$ are n values obtained by simulation from the distribution $U [-0.1; +0.1]$. The U distribution introduces equal variability in all directions.

Then, we moved on the next step in which the BL test is applied to each series i_0 , $\{u_k^{i_0}\}_{k=1}^n$ generated synthetically, by calculating the statistics distance of χ^2 and the p -value test for that series. Then, we obtained 1000 synthetic series, with 1000 p -values $\{p_1^i, p_2^i, \dots, p_n^i\}_{i=1}^{1000}$ and 1000 χ^2 distances.

Summarizing, given an observed series, $\{x_k^{obs}\}_{k=1}^n$, we can generate 1000 synthetic simulations $\{\{x_{ik}\}_{k=1}^n\}_{i=1}^{1000}$ and 1000 p -values $\{p_1^i, p_2^i, \dots, p_n^i\}_{i=1}^{1000}$, then, we calculate both the average p -value \bar{p} and the average distance χ^2 , in addition we calculate quantiles of α -order for those p -values, q_α . Moreover, from q_α , it is possible to obtain the equivalent to a confidence interval that allow validation about the decision of the BL fulfilment with the observed data.

Farhadi & Lahooti, 2022 also point out, as a weakness of the aforementioned work, that we did not provide precise figures on the observed frequencies.

In fact, the figures of observed frequencies for each region were not provided in the paper to avoid redundancies. The observed frequencies at a national level and for each region were both shown graphically in Figures 2 and 3, respectively, where we compared

observed frequencies with the Benford's Law in order to graphically show anomalies in the data. In addition, as we mentioned in Section 3, "Data and Sources", we ran our analysis with "datadista Git-Hub repository" data [3], which is publicly available.

Finally, Farhadi & Lahooti, 2022 also commented that epidemic management limitations at the beginning of the outbreak that may have affected the distribution of leading digits. The authors stated that we ignored such limitations.

However, the literature supports our methodology design. The suitability of the BL for monitoring the registration of pandemic data has been tested for many countries. Several papers in the field show how BL is useful, especially when the number of cases is increasing; however, when the number of cases stabilizes, BL is no longer useful [4–8]. Therefore, the period selected for our analysis (March to June 2020) should fit the assumptions for applying this methodology.

Furthermore, the period selected for our analysis (the lockdown during the first wave) allows us to analyse a situation in which the confinement rules were homogeneous for the whole of Spain. Afterwards, decisions on socio-sanitary standards were taken by each of the regions, leading to non-comparable situations.

In addition, within the discussion section, a reflection on the complexity of managing an unexpected pandemic of the magnitude of COVID-19 has been added, especially for a country like Spain where health responsibilities are decentralized among regions and, therefore, adopting common tools to monitor data registration is required.

Author Contributions: Conceptualization, F.M.-J.; methodology, F.M.-J.; validation, F.M.-J., M.C.-T., and V.C.-T.; investigation, F.M.-J., M.C.-T. and V.C.-T.; resources, F.M.-J., M.C.-T. and V.C.-T.; data curation, F.M.-J.; writing—original draft preparation, M.C.-T. and V.C.-T.; writing—review and editing, M.C.-T. and V.C.-T.; visualization, F.M.-J. and V.C.-T.; supervision, F.M.-J., M.C.-T. and V.C.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting reported results can be found <https://doi.org/10.7910/DVN/GPFFAQ>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Farhadi, N.; Lahooti, H. Forensic Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak. *COVID* **2022**, *2*, 472–484. [[CrossRef](#)]
2. Morillas-Jurado, F.G.; Caballer-Tarazona, M.; Caballer-Tarazona, V. Applying Benford's Law to Monitor Death Registration Data: A Management Tool for the COVID-19 Pandemic. *Mathematics* **2022**, *10*, 46. [[CrossRef](#)]
3. Datadista. Coronavirus Disease 2019 (COVID-19) in Spain. Harvard Dataverse. 2020. Available online: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GPFFAQ> (accessed on 1 October 2020).
4. Campolieti, M. COVID-19 deaths in the USA: Benford's law and under-reporting. *J. Public Health* **2021**, *44*, e268–e271. [[CrossRef](#)] [[PubMed](#)]
5. Isea, R. How Valid are the Reported Cases of People Infected with Covid-19 in the World? *Int. J. Coronaviruses* **2020**, *1*, 53–56. [[CrossRef](#)]
6. Isea, R.; Lonngren, K. A quick Look at the Registered Cases of Covid-19 Throughout the World. *Int. J. Coronaviruses* **2020**, *1*, 16–21. [[CrossRef](#)]
7. Küstner, B.M. La información sanitaria se enreda en la informática. *Gac. Sanit.* **2011**, *25*, 343–344. [[CrossRef](#)] [[PubMed](#)]
8. Sambridge, M.; Jackson, A. National COVID numbers—Benford's law looks for errors. *Nature* **2020**, *581*, 384. [[CrossRef](#)] [[PubMed](#)]