Economics Letters 196 (2020) 109573

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Benford's Law and COVID-19 reporting[☆]

Christoffer Koch[a], Ken Okamura[b,*]

[a] Research Department, Federal Reserve Bank of Dallas, 2200 North Pearl Street, Dallas, TX 75201, United States
[b] Saïd Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, United Kingdom

ARTICLE INFO

Article history:
Received 19 April 2020
Received in revised form 9 September 2020
Accepted 12 September 2020
Available online 14 September 2020

JEL classification:
E61
E65
H41
I18

Keywords:
Corona COVID-19
Statistical reporting
Government accountability
World Health Organization

ABSTRACT

Trust in the reported data of contagious diseases in real time is important for policy makers. Media and politicians have cast doubt on Chinese reported data on COVID-19 cases. We find Chinese confirmed infections match the distribution expected in Benford's Law and are similar to that seen in the U.S. and Italy. We identify a more likely candidate for problems in the policy making process: Poor multilateral data sharing on testing and sampling.

1. Chinese Reporting on the coronavirus

Contrary to popular speculation, we find no evidence that the Chinese massaged their COVID-19 statistics. We use a statistical fraud detection technique, Benford's (1938) Law, to assess the veracity of the statistics. This empirical finding is important because China was affected first. Policies to combat the global pandemic are informed by its response. Skepticism about the Chinese data may result – and may indeed already have resulted – in poor policy choices. Data sharing practices at the early stages of the pandemic were inadequate and led to costly policy errors.

The media frequently claim the Chinese government has understated the numbers of those affected.[1] Politicians echo these claims with President Trump declaring the reported death toll and infections seemed "a little bit on the light side". Much of the concern about Chinese data manipulation can be attributed to geopolitical tensions and foreign governments' need for a scapegoat.[2] The on-going doubts over the credibility of its published data are problematic as it impacts subsequent policy choices by countries that saw epidemics later. Papers that rely on Chinese data for calibration and analysis include: Models of economic activity and the trade-off with deaths such as Atkeson (2020), Jones et al. (2020) and Alvarez et al. (2020); Fang et al. (2020) predict the effect of movement restrictions on the spread of the disease.[3] Since countries patterned their social distancing and lockdown policies on the choices made by China,[4] policy makers need to know the data is reliable.

Lack of confidence in Chinese data may have contributed to a slower response in Europe to the emergent pandemic. Chinese provinces neighboring Hubei province, the Chinese epicenter, imposed movement controls, quarantines and checks on January 23rd at a time when the number of confirmed cases in Hubei was 444 and the number of deaths was 17.[5] In comparison Italy, Europe's initial pandemic hotspot, reached 445 cases on February

☆ The views expressed in this paper are those of the author and are not necessarily reflective of views at the Federal Reserve Bank of Dallas or the Federal Reserve System. Any errors or omissions are the sole responsibility of the authors.

* Corresponding author.
E-mail addresses: christoffer.koch@dal.frb.org (C. Koch), ken.okamura@sbs.ox.ac.uk (K. Okamura).

1 "What to make of China's Coronavirus figures" Foreign Policy, April 1st 2020, James Palmer. "Can China's COVID-19 statistics be trusted?" The Diplomat, March 26th 2020, Scott Romaniuk and Tobias Burgers.

2 "China's ambassador to US slams Trump for COVID-19 blame" ABC News, August 5th 2020, Mike Levine.

3 One interesting note is that an early and influential paper on the macroeconomic effects of pandemics Eichenbaum et al. (2020) does not rely on the Chinese data for calibration.

4 See Qiu et al. (2020).

5 Hubei province has a population of 58.5 million, the city of Wuhan has a population of 11 million, making these entities roughly similar to Italy and the Lombardy region.

https://doi.org/10.1016/j.econlet.2020.109573

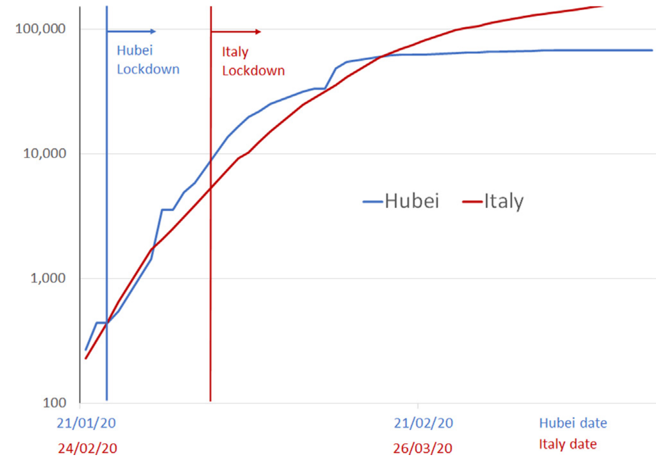

Fig. 1. Confirmed cases in Chinese provinces, U.S. States and Italian Regions.

26th and 17 deaths the following day. It took until March 9th for a national lockdown. Similarly, restrictions on international travel were too late and too mild. By February 26th, Hubei had seen cases rise to 65,187 and deaths to 2615.

Fig. 1 shows the similarity in trajectory between the Italian and Hubei number of confirmed cases. It took approximately a month for the number of cases to plateau in Hubei and this information was available to the Italians when their case numbers matched those of Hubei in late February. The 11 day delay explains the far higher number of cases in Italy.

Skepticism about politically motivated manipulation of Chinese state statistics is deeply rooted. Anecdotal and academic evidence point to lower level officials manipulating data to meet targets. In 2007, Chinese Premier Li Kejiang called all GDP measures "man-made and therefore not reliable" when discussing data on Liaoning province.[6] Lyu et al. (2018) find evidence that regional growth rates are manipulated to meet growth targets. In the case of the SARS outbreak in 2002–3, criticism of the Chinese response surfaced. The World Health Organization (WHO) suspected that China underreported the number of cases (see Parry, 2003).[7]

Data manipulation took place early in the epidemic.[8] The number of cases reported by the Wuhan authorities was "frozen" at 41 during the Hubei provincial Chinese People's Political Consultative Conference and the Wuhan People's Congress (*Lianghui*) between January 12th and 17th, 2020. A member of the WHO emergency committee, John Mackenzie, told the Financial Times on February 5th that China must have been withholding information on new cases.[9]

2. Benford's law

Benford's Law is used to detect fraud or flaws in data collection based on the distribution of the first digits of observed data. A Benford distribution of first digits arises naturally for exponential processes with multiple changes of magnitude, Michalski and Stoltz (2013). The spread of COVID-19 demonstrates exponential growth and changes of magnitude.

The frequency with which the first digit is "1" is 30.1%, the first digit is "2" is 17.6% etc., declining to the first digit being "9" only 4.6% of the time. Since it takes a 100% increase to go from "1" to "2" and a mere 11.1% increase to go from "9" to "1", this logarithmic distribution makes sense. See Table 1

$$P(d) = \log_{10} \frac{(1+d)}{d} \qquad \text{for} \quad d \in (1, \ldots, 9) \qquad \text{(Benford's Law)}$$

The use of Benford's Law to detect fraud has been widely demonstrated in economics and accounting (Varian, 1972). Benford's Law has been used to detect manipulation of economic statistics: Nye and Moul (2007), Gonzales-Garcia and Pastor (2009), Rauch et al. (2011), Holz (2014) and Nigrini (1996).

3. No evidence of data manipulation

We compile data from the Johns Hopkins University Corona Virus Research Center, for China and the Centers for Disease Control for the U.S. For Italy our data comes from the daily *Dipartimento della Protezione Civile* bulletins. The time period for each country matches the period when the pandemic goes through its exponential growth phase and then declines as measures to combat the infection such as quarantines and lockdowns are instituted as in Table 2.

We focus on the daily number of confirmed cases by political subunit — Chinese provinces, U.S. states, and Italian regions. The number of confirmed cases understates the true number of infections as China was unable to test those who did not present at hospitals. Limited testing capacity was also a problem in both Italy and the U.S. As long as the choice of sampling methodology does not change, the number of confirmed cases will also follow an exponential path and thus Benford's Law.

As sampling changes, through testing different groups or changing definitions, we would expect to find a drift away from Benford's Law. For example, if a country only tests those who present at hospital with symptoms, this sample will grow exponentially along with the overall number of infected. If the country then increases testing to those who are symptomatic, but not hospitalized, this sample will also grow exponentially, but will not be the same series as the hospital sample. Frequent sampling changes, in terms of the populations, definitions, or accuracy will lead to deviation from Benford's Law.

The sample periods match the exponential growth phase of the pandemic and the subsequent deceleration. We expect the period pre-lockdown to follow a Benford distribution, the period post-lockdown is a treatment period that should disrupt the Benford distribution. We follow Kissler et al. (2020) and assume 9 days between infection and hospitalization and assume infections are detected with hospitalization. We use the dates given in Fang et al. (2020) as proxies for Chinese provincial lockdowns. For Italy we use the national lockdown on March 9th for all regions. For U.S. states, we use the date of a "stay-at-home order".

Fig. 2 shows that for Chinese provinces, U.S. states and Italian regions the number of confirmed cases the distribution of the first digits shows a decline from 1 to 9 in line with the expected distribution of Benford's Law pre-lockdown.

Tests of significance for Benford's Law require that the "true" distribution should follow the Benford distribution. Our null hypothesis is that the observed distribution follows the theoretical (Benford) distribution. The most common test is the Chi-Square test of Goodness of Fit:

$$D^2 = n \sum_{d=1}^{9} \frac{(h_d - p_d)^2}{p_d} \qquad \text{(Chi-Square test)}$$

6 U.S. diplomatic cable March 15th, 2007. Leaked to Wikileaks.

7 Since the SARS outbreak, there have been improvements to Chinese domestic disease control. In the 2013 H7N9 outbreak the Chinese public health system worked well (Wang, 2013).

8 The obvious example being the treatment of Li Wenliang who died after contracting the coronavirus Green, 2020.

9 "WHO expert says China too slow to report coronavirus cases" Financial Times, February 5th 2020, Primrose Riordan and Sue-Lin Wong.

Table 1
Benford's law distribution of first digit.

First digit	1	2	3	4	5	6	7	8	9
Benford distribution probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Table 2
Data sample periods for confirmed cases.

Country	Start	End	Number of geographic units
China	Jan 21, 2020	Mar 16, 2020	31 Provinces
U.S.	Feb 29, 2020	Jun 30, 2020	50 States and D.C.
Italy	Feb 21, 2020	Apr 16, 2020	19 Regions and 2 Autonomous Provinces

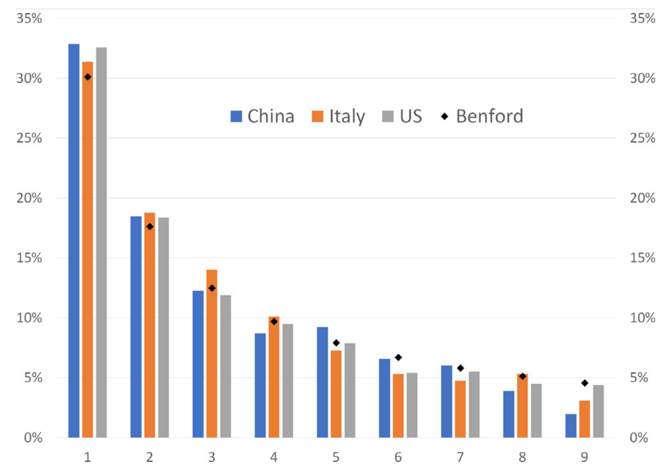

Fig. 2. First Digit Distribution Pre-Lockdown number of confirmed cases in Chinese Provinces, U.S. States and Italian Regions.

Where n denotes the number of observations, h is the observed frequencies of the digits and p is the Benford's Law distribution. We also use a Kuiper test (a modified Kolmogorov–Smirnov test).

$$T_K = \left(D_n^+ + D_n^-\right)\left[\sqrt{n} + 0.155 + \frac{0.24}{\sqrt{n}}\right] \quad \text{(Kuiper test)}$$

Where $D_n^+ = \sup(H_d - P_d)$ and $D_n^- = \sup(P_d - H_d)$ and H_d and P_d represent the cumulative frequencies of the first digit d in the observed data and the Benford distribution. We also calculate the m (max) statistic where $m = \max_{d=1,\dots,9} |h_d - p_d|$ and the d (distance) statistic $d = \sqrt{\sum_{d=1}^{9}(h_d - p_d)^2}$.

We find in Table 3, as expected, pre-lockdown matches Benford far more than the overall period for Italy and China. The U.S. distribution is close to Benford for the entire period and does not appear to change significantly before and after the lockdowns are announced.

Table 3 shows that the Chi-Square test does not support a Benford distribution. However, as noted in other papers on Benford's Law, the Chi-Square test is extremely sensitive for large sample sizes and tends to reject statistical significance even for small differences. The Kuiper test does not reject the null hypothesis that the distribution is Benford for China for the entire time period and pre-lockdown and for Italy pre-lockdown. For China and Italy pre-lockdown, the d and m tests also do not reject the null. The U.S. results show that for the full period, the Kuiper, d and m tests do not reject the Benford distribution, but show that pre-lockdown the null is rejected at the 10% level for d and m tests and at the 5% level for the Kuiper. Post-lockdown in the U.S. the m test does not reject the null that the distribution is Benford.

The U.S. follows Benford over the full time period, but does not display a Benford distribution pre-lockdown. We attribute this to relative lax adherence and the piecemeal policy measures (Dave et al., 2020).

4. The Italian delay

Given the doubts about the reliability of Chinese data, why did the Italians delay? The likely reason is in the sampling of infections. The Italian government believed they were detecting a far higher proportion of infected than the Chinese had managed at the same point in the pandemic. On February 27th, "Public health officials have said that Italy contributed to fears of an epidemic in Europe with its zealousness in testing"[10] and Dr. Walter Ricciardi, an Italian government adviser and World Health Organization (WHO) official, was quoted as saying there was "too much testing".[11] In retrospect, the Italians had not "tested too much" and their rate of virus detection was no better than that of the Chinese.

Information on the extent of testing early in the pandemic is lacking in both China and Europe. Hubei authorities were able to undertake 4000 tests per day by February 4th. Even Germany, lauded for its testing regimen, had not published data on the number of tests in late February, leading to Dr. Ricciardi claiming as late as March 7th "from an epidemiologic point of view, it is not plausible that Italy ... accounts for more cases than Germany and France".[12]

5. Focus on sampling

It is possible to create data series that fit Benford's Law (Diekmann, 2007). To manipulate the Chinese data requires coordination of daily announcements across all provinces while accurately forecasting future infection rates. This is improbable. With the benefit of hindsight, we now know that the Italians had a similar sampling method to China. Both countries' distributions of first digits for confirmed cases pre-lockdown follow Benford's Law. A key insight from our analysis is a focus on the underlying data sampling processes. The ongoing geopolitical dynamics and escalation of policy rhetoric between the U.S., European countries, and China have obscured one of the causes of poor policy responses. We conclude a refinement of the International Health Regulations Article VI §2 to share timely, accurate and sufficiently detailed information on the extent and reliability of *testing* is necessary.

10 "Italy blasts virus panic as it eyes new testing criteria" AP News, February 27th 2020, Frances D'Emilio and Nicole Winfield.

11 "Coronavirus accounting is looking vulnerable" Bloomberg Opinion, March 2nd 2020, Lionel Laurent.

12 "Leap in coronavirus cases tests limits of Italy's health system" Al-Jazeera, March 7th 2020, Michele Bertelli.

Table 3
Table of first digit distribution and tests of significance.

Country	Time	Leading digit									N	χ^2-Stat	d_n^*	m_n^*	Kuiper
		1	2	3	4	5	6	7	8	9					V^*
China	Full Sample	249	128	90	57	60	48	37	23	13	705	25.33***	1.72***	1.38***	0.91
China	Pre-Lockdown	194	106	72	51	52	38	36	22	10	581	16.04***	1.16	0.79	0.33
China	Post-Lockdown	63	28	19	7	10	11	3	1	3	145	23.78***	1.89***	1.61***	1.87*
Italy	Full Sample	326	142	98	94	91	61	65	53	50	980	18.13***	1.69***	0.99**	1.84***
Italy	Pre-Lockdown	113	68	50	36	26	19	17	19	11	359	5.00***	0.65	0.29	0.64
Italy	Post-Lockdown	213	74	48	58	65	42	48	34	39	621	39.61***	2.31***	1.42***	2.81***
U.S.	Full Sample	1682	962	700	578	438	351	295	263	210	5479	15.19***	1.07	0.64	0.91
U.S.	Pre-Lockdown	608	343	222	177	147	101	103	84	82	1867	11.40***	1.31*	1.06*	1.34**
U.S.	Post-Lockdown	1074	619	478	401	291	250	192	179	128	3612	20.03***	1.25*	0.85	1.53**

***Denotes statistical significance at the 1% level.
**Denotes statistical significance at the 5% level.
*Denotes statistical significance at the 10% level.

References

Alvarez, F.E., Argente, D., Lippi, F., 2020. A Simple Planning Problem for COVID-19 Lockdown. Working Paper Series 26981, National Bureau of Economic Research.

Atkeson, A., 2020. What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. Working Paper Series 26867, National Bureau of Economic Research.

Benford, F., 1938. The law of anomalous numbers. Proc. Am. Phil. Soc. 78 (4), 551–572.

Dave, D., Friedson, A.I., Matsuzawa, K., Sabia, J.J., 2020. When do shelter-in-place orders fight COVID-19 best? Policy heterogeneity across states and adoption time. Econ. Inq.

Diekmann, A., 2007. Not the first digit! using benford's law to detect fraudulent scientific data. J. Appl. Stat. 34 (3), 321–329.

Eichenbaum, M.S., Rebelo, S., Trabandt, M., 2020. The Macroeconomics of Epidemics. Working Paper Series 26882, National Bureau of Economic Research.

Fang, H., Wang, L., Yang, Y., 2020. Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-nCoV) in China. Working Paper Series 26906, National Bureau of Economic Research.

Gonzales-Garcia, J., Pastor, G., 2009. Benford's Law and Macroeconomic Data Quality. International Monetary Fund Working Paper 2009-10.

Holz, C.A., 2014. The quality of China's GDP statistics. China Econ. Rev. 30, 309–338.

Jones, C.J., Philippon, T., Venkateswaran, V., 2020. Optimal Mitigation Policies in a Pandemic: Social Distancing and Working from Home. Working Paper Series 26984, National Bureau of Economic Research.

Kissler, S.M., Tedijanto, C., Lipsitch, M., Grad, Y., 2020. Social distancing strategies for curbing the COVID-19 epidemic. medRxiv.

Lyu, C., Wang, K., Zhang, F., Zhang, X., 2018. GDP Management to meet or beat growth targets. J. Account. Econ. 66 (1), 318–338.

Michalski, T., Stoltz, G., 2013. Do countries falsify economic data strategically? Some evidence that they might. Rev. Econ. Stat. 95 (2), 591–616.

Nigrini, M.J., 1996. A taxpayer compliance application of Benford's law. J. Am. Tax. Assoc. 18 (1), 72.

Nye, J., Moul, C., 2007. The political economy of numbers: On the application of Benford's law to international macroeconomic statistics. BE J. Macroecon. 7 (1).

Parry, J., 2003. WHO Is worried that China is under-reporting SARS. BMJ 326 (7399), 1110.

Qiu, Y., Chen, X., Shi, W., 2020. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. J. Popul. Econ. 33 (4), 1–46.

Rauch, B., Göttsche, M., Brähler, G., Engel, S., 2011. Fact and fiction in EU-governmental economic data. Ger. Econ. Rev. 12 (3), 243–255.

Varian, H.R., 1972. Benford's law (letters to the editor). Am. Stat. 26 (3), 65.

Wang, Y., 2013. The H7N9 influenza virus in China–changes since SARS. New Engl. J. Med. 368 (25), 2348–2349.