Accepted Manuscript

Tampering with inflation data: A Benford law-based analysis of national statistics in Argentina

Maximilano Miranda Zanetti, Fernando Delbianco, Fernando Tohmé

PII: S0378-4371(19)30411-X
DOI: https://doi.org/10.1016/j.physa.2019.04.042
Reference: PHYSA 20806

To appear in: *Physica A*

Received date : 27 September 2018
Revised date : 18 February 2019

Please cite this article as: M.M. Zanetti, F. Delbianco and F. Tohmé, Tampering with inflation data: A Benford law-based analysis of national statistics in Argentina, *Physica A* (2019), https://doi.org/10.1016/j.physa.2019.04.042

Tampering with Inflation Data: A Benford Law-based Analysis of National Statistics in Argentina

Maximilano Miranda Zanetti [†] - Fernando Delbianco[‡]* - Fernando Tohmé[‡]

[†] *Universidad Nacional del Sur, Departamento de Economía*
mmiranda@uns.edu.ar
[‡] *Universidad Nacional del Sur, Departamento de Economía and INMABB-CONICET*
fernando.delbianco@uns.edu.ar, ftohme@criba.edu.ar
San Andrés 800, 8000 Bahía Blanca, Argentina

Abstract

There is a widespread consensus that the national statistics on inflation were manipulated by the Argentinean government from 2006 to 2015. The best known tool to run a forensic analysis of this claim is to check for the validity of *Benford's law* in the data series. We find that indeed, the inflation for that period fails to satisfy this statistical regularity. We further compare this behavior to that of Argentina's inflation series for the same period but recorded independently of the government; to that of the national records of 1943-2006, as well as to historical series of other countries. We find again that Argentina in 2006-2015 is the only one in our sample that can be singled out as candidate for statistical manipulation.

Alternative hypotheses for why the inflation series failed to satisfy Benford's law can be formulated. One is that, it may be due to rounding price level figures to the significant digits. Or that it is due to changes in the base years which leads to splicing different series of general level

*Corresponding author.

of prices. We consider these alternative hypotheses and run simulations to assess them. We find that, independently of these possible changes in the underlying series of prices, the ensuing series of its variations, i.e. the series of inflation rates, always satisfies Benford's law. Therefore we can claim that, indeed, inflation data was tampered with in Argentina for an entire decade.

Keywords: Benford's Law; data tampering; inflation; price series.

1 Introduction

The inflation index produced by INDEC (National Institute of Statistics and Census of Argentina) from 2006 to 2015 are widely believed to have been manipulated by the Executive in order to downplay the actual magnitude of price increases [6], [2], [8]. Recent studies show that weak democracies (as Argentina at that period) tend to manipulate other macroeconomic variables mostly as a promotional tool of the incumbent government [7]. Inflation in a country with a history of price instability seems, in this sense, a perfect candidate for manipulation.

The goal of this paper is to assess the soundness of this claim, by resorting to a well-known strategy, namely checking the validity of Benford's law in both the price index and the inflation series. Furthermore, we compare the results obtained in that study with those for series reported by an independent source (`www.inflacionverdadera.com`[1]). We find that the INDEC 1943-2017 inflation series fails to satisfy Benford's law, while the shorter 1943-2006 INDEC series and that of `InflacionVerdadera.com` agree with it.

To ensure the robustness of our results we perform a similar analysis for the same series in other countries. Besides the US data (as representative of a developed economy), we consider Chile (a neighbor country of Argentina with similar degree of development), and Venezuela and Zimbabwe (countries experiencing, like Argentina, high and very high inflationary processes). The idea is to check whether the results of analyzing the INDEC data series does not reveal some hidden property of inflation indexes in either the general case, in middle-income countries or in high inflation economies. Our test allows us to reject the null hypothesis of the validity of Benford's law only in the case of the INDEC 1943-2017 series . We can thus discard the idea that there is something

[1]Supported by the joint MIT Sloan and Harvard Business School *Billion Prices* project.

in inflation series in general that precludes them to satisfy Benford's law. On the other hand, series of price indexes seem to be much less prone to satisfy this regularity.

Then, we consider two alternative hypotheses, other than the manipulation of the series, for the failure of inflation in satisfying Benford's law. One is that rounding to significant digits, a usual procedure to simplify the reporting of the series, might distort its properties. The other is that splicing price index series corresponding to different base years introduces a distortion in the derived inflation series which fail to satisfy Benford's law.

In order to discard these alternatives we resort to running simulations of both price and the derived inflation series, in the presence of either rounding and splicing. The results, again, lead us to reject the possibility that either of those hypotheses may explain the failure of Benford's law in the case of the INDEC series. The plan of the paper is as follows. Section 2 briefly discusses Benford's law and its uses in "statistical forensics". Section 3 presents the data and their properties. Section 4 runs the comparison with inflation series from alternative sources, different periods and corresponding to other countries. Section 5 discusses the simulations that allow us to show, in abstract terms, that inflation-like series satisfy Benford's law and that splicing and rounding do not affect these properties. Section 6 concludes.

2 Benford's law and fraud investigation

Benford's "law" is a claim about the frequency distribution of first (or most significant) digits in the decimal expansion of the numbers in most numerical databases. More precisely, for any digit $d \in \{1, 2, \ldots, 9\}$ the probability of being the leading digit is[2]

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

which can be extended to the probability of any string of length n of digits drawn from $\{0, \ldots, 9\}$ as long as the first digit is $\neq 0$. By a slight abuse of language, such string can be seen as a natural number s_n. Then, the probability of s_n is $\log_{10}(1 + \frac{1}{s_n})$.[3] While there are series that do not satisfy this property, an

[2] This can be extended to any numerical base, just replacing 10 by the new base.

[3] Of particular interest for our simulations in Section 4 is the case of Benford's law in two digits, i.e. when $s_n \in \{10, \ldots, 99\}$.

interesting result is that scale invariance of a series (i.e. that are not affected by changes in the unit of measurement) implies that it verifies Benford's law [10]. This is particularly interesting in the case of inflation series that do not depend on the monetary unit in which prices are expressed. Similarly to spliced price series, which differ in some scale. Even series that are not scale invariant may satisfy Benford's law. On the other hand, series in which truncation and rounding have been applied tend to fail to satisfy it [11] [1]. But other than in those cases, the validity of the law is pervasive. Thus, it has been used as a general principle to check for fraud and manipulation of data [13] [9].

Many interesting analyses of fraud have been run on scientific data ([4]), accounting data ([5]) and, closer to our own work, macroeconomic data ([12]). The procedure is basically the same, independently of the source of data. Here we explain it for the case of a single digit, but as indicated in Footnote 2, can be easily extended to any number of digits. For each series, it starts by finding the actual distribution of digits, which yields, for each $d \in \{1, \dots, 9\}$, a frequency A_d. On the other hand, recall that $P(d)$ represents the probability of d according to Benford's law. Then, if the size of the series is m we compute:

$$\chi^2 = m \sum_{d=1}^{9} \frac{(A_d - P(d))^2}{P(d)}$$

to obtain the χ^2 statistics which allows us to reject the hypothesis that the series satisfies Benford's law and thus infer a possible manipulation of data. On the other hand, if this hypothesis cannot be rejected we can keep assuming that the series behaves according to Benford's law.

3 Data

The sources of data used in this work can be summarized in two parts. First, data on Argentinean inflation, for which we take the official data from INDEC[4] (monthly inflation in the 1943-2017 period) as well as from `inflacionverdadera.com`.[5]

On the other hand, we use series of other countries as control. As indicated in the introduction, we consider the inflation series of the USA, Chile, Venezuela

[4] Which can be found in `https://www.indec.gob.ar/`. Price series are reported with two significant digits while inflation series only with one.

[5] `https://www.inflacionverdadera.com/argentina/`. This site recorded independent measurements of inflation in Argentina to replace the suspected INDEC data. It ended its run on Argentinean data in February 2018. This series can now be found in `www.pricestats.com`.

and Zimbabwe, all of them in annual terms for the period 1980-2016, which we get from the World Bank database[6].

In figure 1 we depict the two time series of inflation from Argentina. We can see the discrepancy between the official inflation reported by INDEC, and the independent measures of [2]. The latter diverges from the former only in the data from 2007 to 2015 [7]. This difference disappears in 2016. This indicates that during the period of interest INDEC may have systematically underestimated the inflationary process.

In figure 2 we present our control cases. We can see that they reflect very different inflationary experiences, ranging from very stable low inflations to cases with very high and even hyperinflation.

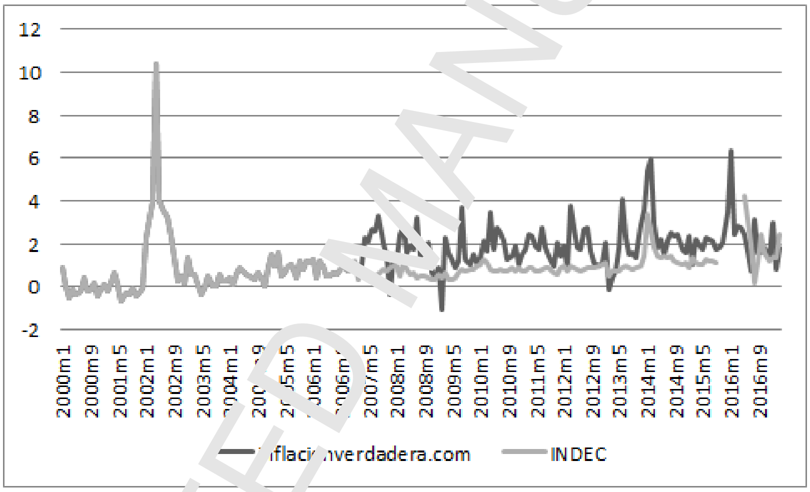


Figure 1: Inflation in Argentina

Tables 1 through 4 present the results of running the χ^2 test to the series from Argentina under the null hypothesis of the validity of Benford's law. The difference among these tables resides in the subsamples we choose in order to achieve more robustness. We run the analysis for both the monthly series of the general level of prices and inflation. Table 5 repeats this procedure for the countries selected as controls, although the series are instead annual. The results were obtained by using two R packages, Benford's.analysis

[6] `http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators`

[7] We use monthly instead of annual inflation since otherwise we would have very few data-points representing the suspicious period

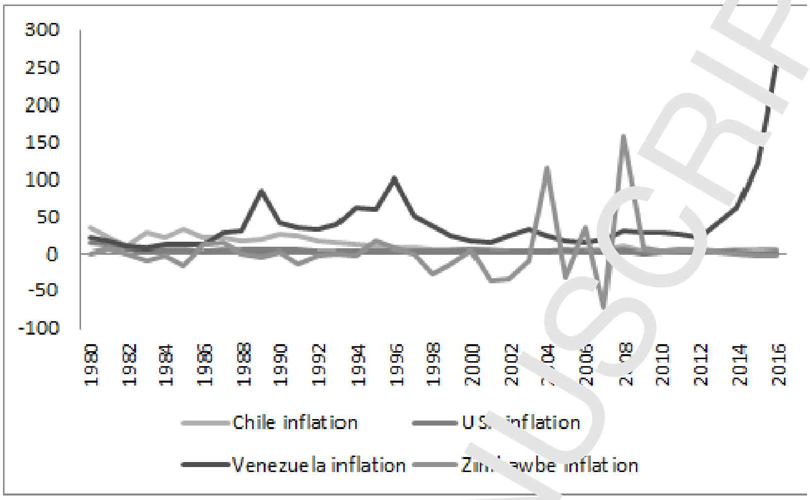


Figure 2: Inflation in control countries

(https://CRAN.R-project.org/package=Benford's.analysis) and Benford'sTests (https://CRAN.R-project.org/package=Benford'sTests).

We run the exercise for Argentina under different time windows in order to check the satisfaction of Benford's law in different periods, both including and excluding the period under suspicion. Besides checking the entire run of values with and without the period under suspicion, we include two highly relevant periods [3]. One goes from 1990 to 2017. 1990 was the last year in which Argentina experienced hyperinflation and can be seen the start of the last historical period in the inflationary history of the country, down from the high and hyper inflation experienced in the 1970s and 1980s. The idea is to discard the possibility that violations to Benford's law are due to the high levels of inflation (sometimes in the hundreds per month!), in which the first significant digit seemed irrelevant. In response to the last hyperinflation, convertibility of the currency to 1 Argentinean peso = 1 US dollar was enacted, which lasted until 2001. We take this convertibility period as another particular time window in our analysis, in which very low levels of inflation were also sometimes under suspicion. In all cases we run the χ^2 test to see if the null hypothesis of Benford's law in one digit can be rejected. We do that both using the price index and the derived monthly inflation rate.

The main result is that, unlike the case of the control countries, the null hypothesis can be rejected when testing for inflation on INDEC data at 5% in all cases except for 1943-2006, while the `inflacionverdadera.com` series does not reject the validity of Benford's law at any period under analysis. Running the same exercise in the case of the control countries we find that Benford's law cannot be rejected in any case, except Zimbabwe at 5% (but not at 1%), another country that experienced a hyperinflation.

On the other hand, the price index series for both INDEC and `inflacionverdadera.com` reject the null hypothesis. Close to rejection are also the price series in Chile and Zimbawe. This can be seen in Table 5.

Finally Table 6, summarizes all this information, where for each possible source and series we mark with X the rejection of the null hypothesis and with ✓ if Benford's law is statistically acceptable. On the other hand, we include another expression, namely X^* to denote that the rejection depends on the level of significance selected. The latter case allows us to point out when rejections (or acceptances) of the null hypothesis are weaker.

We can see that the INDEC inflation series differs from those of other countries as well as from the series recorded by independent sources. Clearly the tampering of data by political officers is the prime suspect, but the rejection in the case of price series, may indicate that there might exist another explanation. In particular that standard bookkeeping operations, like rounding and splicing may have induced the rejection of Benford's null hypothesis. The next section is devoted to analyze this possibility.

Argentina (1943-2017)	χ^2	**p-value**
price index `inflacionverdadera`	514.42	< 2.2e-16
price index indec	502.17	< 2.2e-16
inflation `inflacionverdadera`	89.686	0.686
inflation indec	168.36	7.78E-07

Table 1: Inflation and Prices in Argentina 1943-2017

Argentina (1990-2017)	χ^2	**p-value**
price index `inflacionverdadera`	1276.3	< 2.2e-16
price index INDEC	1237.8	< 2.2e-16
inflation `inflacionverdadera`	84.354	0.6195
inflation INDEC	216.62	1.22E-12

Table 2: Inflation and Prices in Argentina 1990-2017

Argentina (1943-2006)	χ^2	p-value
price index inflacionverdadera	136.31	0.0009388
price index INDEC	584.33	< 2.2e-16
inflation inflacionverdadera	103.96	0.1327
inflation INDEC	111.98	0.05025

Table 3: Inflation and Prices in Argentina 1943-2006

Argentina (1991-2001)	χ^2	p-value
price index inflacionverdadera	3068.4	< 2.2e-16
price index INDEC	2991.4	< 2.2e-16
inflation inflacionverdadera	66.536	[illegible]964
inflation INDEC	73.933	0.87[illegible]

Table 4: Inflation and Prices in Argentina 1991-2001

Chile (1980-2016)	χ^2	p-value
Inflation	89.097	[illegible]
Price index	[illegible]	[illegible]
USA (1980-2016)	χ^2	**p-value**
Inflation	[illegible]	[illegible].9727
Price index	[illegible]3.093	0.889
Venezuela (1980-2016)	χ^2	**p-value**
Inflation	78.893	0.7697
Price index	[illegible]	0.9723
Zimbabwe (1980-2016)	χ^2	**p-value**
Inflation	77.387	0.8054
Price index	112.67	0.04585

Table 5: Inflation and Prices in Control Countries 1980-2016

Period	INDEC price index	INDEC Inflation	inflacionverdadera price index	inflacionverdadera Inflation
1943/2017	X	X	X	✓
1990/2017	X	X	X	✓
1943/2006	X	✓	X	✓
1991/2001	X	✓	X	✓
Controls	Chile	USA	Venezuela	Zimbabwe
Inflation	✓	✓	✓	✓
Index	X*	✓	✓	X*

Note 1: X denotes the rejection of the null hypothesis.
Note 2: ✓ denotes the non-rejection of the null hypothesis.
Note 3: X* denotes that the rejection depends on the level of significance selected.

Table 6: Summary of results

4 Simulations

The results in the previous section indicate that there is nothing in the series of price indexes or inflation by themselves that may point to these kinds of data as culprits for the failure in satisfying Benford's law (as shown by its validity in the control cases). Nevertheless, we want to check with more generality this claim. Therefore, we run simulations of price index and inflation series in order to evaluate the satisfaction of Benford's Law.

Consider a random variable X distributed uniformly over an interval of real numbers $[a, b]$, denoted $X \sim U(a, b)$. Then, we generate a series $C\,10^X$, where C is a positive constant, which we interpret as a "'price" series satisfying Benford's Law since its logarithm is uniformly distributed.[8] Again, we apply the χ^2 test to detect deviations from Benford's law. Summary information about these simulations (see the corresponding pseudocodes in Appendix C) is presented in Tables 7 to 10, where $BL1$ and $BL2$ represent Benford's law in one digit and two digits, respectively.

A time series of inflation is derived from the price series as follows. Consider the price indexes at two consecutive periods t and $t+1$, denoted p_t and p_{t+1}. Then, the inflation rate at period $t+1$ is

$$\frac{p_{t+1} - p_t}{p_t} \times 100$$

Alternatively, we can use the same procedure to generate series we label as "inflation rate series" and by integration we obtain the "price index series". More precisely, given two consecutive periods t and $t+1$ and the inflation rate at $t+1$, I_{t+1}, the corresponding price p_{t+1} is obtained as

$$p_{t+1} = p_t(1 + I_{t+1})$$

up from an initial value p_0 at $t = 0$. This initial value is generated for each selecting at random a value of the corresponding $C\ 10^X$ process.

Table 7 presents the results of generating 5000 series of simulated data representing price indexes. The length of each series is 300, as to be similar to that of INDEC. We can see that in 4767 cases Benford's law is valid in one digit, in 4736 in two digits and in 4544 it is valid in both one and two digits. We then take these 4544 series and use them to derive corresponding inflation series (which, being based on the differences in the price series, have 299 values). We can see, in particular, that 4404 satisfy Benford's law in one or two digits.

In Table 8 we show the results we obtain by first simulating inflation series and then integrating them to yield price series. Thus, from the results reported in Tables 7 and 8, we can infer that, if the original series satisfies Benford's

[8]We choose $[a, b] = [0, 1]$ and $C = 10,000$ for price series, while $C = 1/100$ is used to generate inflation series. We run the simulations with `Mathematica` using its built-in random generator. Alternative generators (Mersenne-twister, etc.) did not yield noticeably different results.

law, the ensuing series, obtained either by differentiation or integration, also satisfies it. This provides further proof that the way in which inflation rates are computed does not affect its compliance with Benford's regularity.

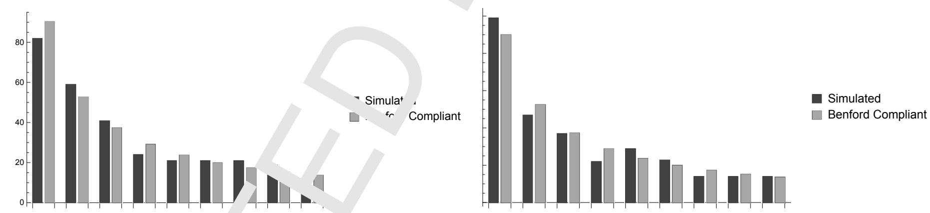
Generated price index series (length of each series: 300)			
Number of series	Satisf. BL1	Satisf. BL2	Satisf. BL1 and BL2
5000	4767	4736	[illegible]
Derived inflation rate series (length of each series: 299)			
Number of series	Satisf. BL1	Satisf. BL2	Satisf. BL1 or BL2
4544	3988	4113	[illegible]

Table 7: From price index series to inflation series.

Generated inflation rate series (length of each series: 299)			
Number of series	Satisf. BL1	Satisf. BL2	Satisf. BL1 and BL2
5000	4743	4742	**4533**
Integrated price index series (length of each series: 300)			
Number of series	Satisf. BL1	Satisf. BL2	Satisf. BL1 or BL2
4533	4531	[illegible]	4533

Table 8: From inflation series to price index series.

A generic pair of a price index and an inflation rate series is represented in Figures 3a and 3b, respectively, indicating that both series satisfy Benford's law.



(a) Price series (b) Inflation index

Figure 3: Summary histogram of the behavior of simulated price series and their derived inflation series. $BL1$ is satisfied by both of them. The domain is the ordered set of digits from 1 to 9, representing the first significant digit.

We analyze the robustness of Benford's's regularity on both price and the derived inflation series. The alternative hypothesis we test is that the failure of satisfying Benford's law in inflation series may be due not to tampering but to the application of standard statistical "bookkeeping" operations by national statistics offices. The comparison we presented previously with the evidence of other countries (in which these operations are routinely applied) is a strong indication of the untenability of this alternative hypothesis. Nevertheless, we use simulated series to check whether those operations (rounding and splicing) may induce violations of Benford's law.

We start by analyzing the consequences of *rounding* to the first significant digit. We present in Table 9 the results of rounding the second significant digit in an inflation series satisfying both $BL1$ and $BL2$. For this we use the 4533 inflation series reported on the top row of Table 8, which also shows that the corresponding price series by and large satisfy Benford's law. We round the values of the inflation series at .01%, i.e. at the second digit. We can see that this does not affect Benford's law as long as the first characteristic digits remain unaffected. Since the figures reported in by official institutions like INDEC include at least one digit, they have not been subject to rounding at the level that can affect the validity of Benford's law.

Inflation index series (length of each series: 299)			
Number of series	Satisf.BL1	Satisf. BL2	Satisf. BL1 or BL2
4533	4510	4437	4531

Table 9: Rounding the second digit of inflation series.

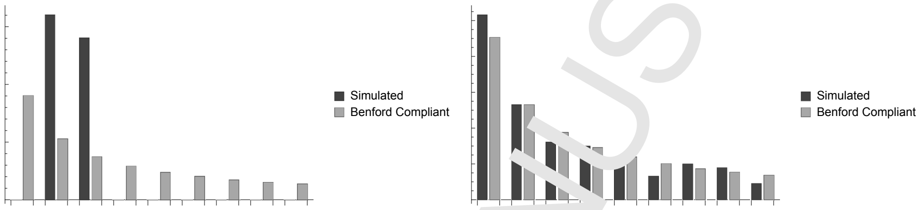
The other operation that can be applied on index series is splicing two or more series. This operation is required when the base year is changed, usually because the consumption basket is updated. For instance, if we have two series, $\{X_t\}_{t=0}^{T}$ and $\{Y_t\}_{t=T}^{T'}$, a single series is created by changing backwards the values of $\{X_t\}_{t=0}^{T}$ as to become compatible with those of $\{Y_t\}_{t=T}^{T'}$. Then, we generate values $\{\hat{Y}_t\}_{t=0}^{T-1}$ such that for each $t = 0, \ldots, T-1$, $\hat{Y}_t = \frac{X_t Y_T}{X_T}$.

Price index series for splicing (length of each series: 1200)			
Total n	Satisf. BL1	Satisf. BL2	Satisf. BL1 and BL2
5000	4762	4728	**4527**
Spliced price series by joining three segments (length of each series: 300)			
Total n	Satisf. BL1	Satisf. BL2	Satisf. BL1 or BL2
4527	0	0	0
Derived inflation series (length of each series: 299)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1 or BL2
4527	4117	4327	4482

Table 10: Splicing price index series

We run a particular numerical experiment, reported in Table 10 to check the result of splicing series. We generate 5000 series of length 1200. From them 4762 satisfy Benford's law in one digit, 4728 in two digits and 4527 in both one and two digits. We take these latter series, and extract three segments from each: $(1, \ldots, 100)$, $(501, \ldots, 600)$ and $(1101, \ldots, 1200)$. We then splice them (so that both the first and the last segment have values compatible with the medium section) to generate a series of length 300. We check the validity of Benford's law in the resulting 4527 series. Interestingly *none* of them satisfies it at either one or two digits. But we then generate the derived inflation rate series, and find that 4482 of them satisfy Benford's law at one or two digits.

We can see in Figure 4a a generic case in which the spliced price series fails to satisfy Benford's's law, but the derived inflation series, shown in Figure 4b, still satisfies the regularity.



(a) Price series (b) Inflation index

Figure 4: Summary histogram of the behavior of splicing simulated price series and their derived inflation series. $BL1$ is not satisfied by the price series but it is verified by the inflation series. The domain is the ordered set of digits from 1 to 9, representing the first significant digit.

Thus, our simulations indicate that the failure to satisfy Benford's law in one case but not in others cannot be ascribed to the usual bookkeeping operations on time series.

5 Conclusions

In this paper we assessed the claim that Argentinean statistics on inflation were manipulated between 2006 and 2015. We find that the official series of the period do not satisfy Benford's law while alternative sources (including data from longer INDEC series) as well as series of other countries, all do satisfy it. The latter evidence also indicates that there does not seem to exist any reason to think that inflation data should, in general, fail to satisfy Benford's law.

Indeed, simulating series satisfying Benford's regularity we show, on derived series, that usual operations on data like rounding up figures splicing and merging, or changing base dates, does not preclude the validity of the law. This means that the problem with INDEC data cannot be due to those operations, and thus reinforces the idea that it is just due to manipulation. This confirms the widespread suspicions of political tampering with official statistics.

A Appendix: syntax of pseudocode

`arg`	argument of a function.
`f[x]`	f(x)
`{a,b,c}`	List of elements (a,b,c) [elements can be scalar or vectors].
`{}`	An empty list.
`l[[i]]`	Returns the $i-th$ element of vector or list l.
`M[[i,j]]`	Returns the element M_{ij} of matrix or array M.
`Map[f&,list]`	For $list = a_1, a_2 \ldots$, command returns $f(a_1), f(a_2) \ldots$
`Map[f(#)&,list]`	Same as above, but with function expressed with # as placeholder for the argument.
`f(#1,#2,...)&`	Body for a function of more than one argument.
`Boolean[test]`	Returns the Boolean values True or False.
`RandomReal[top,n]`	Returns n pseudo-random numbers generated from the interval $[0, top]$

B Appendix: General command definitions

- For sintax:

  `For[start,test,increment,body]`

  Performs `start`, then `body` and `increment` as long as `test` is true.

- Clears non-significant zeroes and takes z significant digits:

  `firstsignif[x,digits]= FromDigits[RealDigits[x, 10, digits][[1]]]`

- Creates the necessary categories to sort values into significant-digits clusters:

  `cats[len]= Table[i,(i,` $10^{(\texttt{len - 1})}$`,` $10^{\texttt{len}} - 1$]

- Main command for chi2 test and alpha level of significance:

  1. `chitest[list,digits,alpha]=newlist={}; index=0; lon=Length[list]; labels=cats[digits];` %Initial computations.
  2. `For[index=1; index <= lon; index++; newlist=Append[newlist,firstsignif[list[[index],digits]]]];`%Loop that strips numbers from non-significant digits.
  3. `frecobs=Map[Count[newlist,#]&,labels];` %Maps a counting routine to each digit or category.

4. `probs=Map[Log10[1 + 1/#]&,labels];` %Determines expected probability for each category according to Benford's Law.

5. `chistat=Sum[(frecobs[[index]]-lon*probs[[index]])`^{2}`/(lon*probs[[index]]),index, 1,` $10^{\texttt{digits}} - 10^{(\texttt{digits-1})}$`];` %Computes the chi-square statistic.

6. `critval= InverseCDF[Chisquare Distribution] [` $10^{\texttt{digits}} - 10^{(\texttt{digits-1})} - 1$`,1-alpha];` %Computes the critical chi-square value.

7. `Return[Boolean[chistat<=criticval]]` %Returns whether the series passes chi-square test for Benford's law.

C Appendix: pseudocodes

- Price series satisfying Benford's law translate into inflation series fulfilling the regularity

  1. `RandomSeed[seed1]` %sets the random number generator to a preset value for reasons of replicability.

  2. `simulp = Table[`$10000 * 10^{\texttt{RandomReal[1,300]}}$`,5000]` %generates 5000 random series from U[0;1] of length 300, which (should) satisfy Benford's law.

  3. `simulpord=Map[Sort,simulp]` %Orders every sequence in simulp so that it resembles a price index series.

  4. `diagnost1=Map[chitest[#,1,.05]&,simulpord]` %performs chi-test of BL, 1 digit, alpha=.05 to each of the 5000 price series.

  5. `diagnost2=Map[chitest[#,2,.05]&,simulpord]` %performs chi-test of BL, 2 digits, alpha=.05 to each of the 5000 price series.

  6. `cases1=Position[diagnost1,True]` %Those series who fulfill BL1.

  7. `cases2=Position[diagnost2,True]` %Those series who fulfill BL2.

  8. `simulpok=simulpord[[Intersection[cases1,cases2]]]` %Selects price series that robustly fulfill BL1 and BL2.

  9. `infl[list]:= lon=Length[list];`
     `listdif=Table[(list[[index+1]]-list[[index]])/list[[index]],index,1,lon-1];`
     `Return[listdif]` %Command to compute inflation series from price series.

10. `simulpinfl=Map[infl,simulpok]` %Computes the corresponding inflation indexes.
11. `diagnost1infl=Map[chitest[#,1,.05]&,simulpinfl]` %performs chi-test of BL, 1 digit, alpha=.05 to each of the inflation series.
12. `diagnost2infl=Map[chitest[#,2,.05]&,simulpinfl]` %performs chi-test of BL, 2 digits, alpha=.05 to each of the inflation series.
13. `cases1infl=Position[diagnost1infl,True]` %Those series who fulfill BL1.
14. `cases2infl=Position[diagnost2infl,True]` %Those series who fulfill BL2.
15. `simulpinflok=simulpinfl[[Intersection[cases1infl,cases2infl]]]` %Selects inflation series that fulfill BL1 or BL2.

- Inflation series satisfying Benford's law translate into price index series fulfilling the regularity

  1. `RandomSeed[seed2]` %sets the random number generator to a preset value for reasons of replicability.
  2. `simulinf = Table [`$1/100*10^{\texttt{RandomReal[1,299]}}$`,5000]` %generates 5000 random series from U[0;1] of length 299, which (should) satisfy Benford's law.
  3. `diagnost1i=Map[chitest[#,1,.05]&,simulinf]` %performs chi-test of BL, 1 digit, alpha=.05 to each of the 5000 inflation series.
  4. `diagnost2i=Map[chitest[#,2,.05]&,simulinf]` %performs chi-test of BL, 2 digits, alpha=.05 to each of the 5000 inflation series.
  5. `cases1i=Position[diagnost1i,True]` %Those series who fulfill BL1.
  6. `cases2i=Position[diagnost2i,True]` %Those series who fulfill BL2.
  7. `simulinfok=simulinf[[Intersection[cases1,cases2]]]` %Selects inflation series that robustly fulfill BL1 and BL2.
  8. `anchorexp=Table[`$1000*10^{\texttt{RandomReal[1,300]}}$`,{Length[simulinfok]}]` %Generates random anchor numbers for price indexes (p0) [anchor distribution itself satisfies BL, so no noise is added].
  9. `buildfrominf[anchor,series]:= index=1;result={anchor}; p=anchor; lon=Length[series]; Do[p=p*(1+series[[index]]); result=Append[result,p]; index++, lon]; Return[result]`

10. `simulinfp=Table[buildpfrominf[anclasexp[[i]],simulinfok[[i]]],{i,1,Length[anchorexp]}]` %builds price index series from inflation and anchor series.
11. `diagnost1ip=Map[chitest[#,1,.05]&,simulpinfp]` %performs chi-test of BL, 1 digit, alpha=.05 to each of the derivated price series.
12. `diagnost2ip=Map[chitest[#,2,.05]&,simulpinfp]` %performs chi-test of BL, 2 digits, alpha=.05 to each of the derivated price series.
13. `cases1ip=Position[diagnost1ip,True]` %Those series who fulfill BL1.
14. `cases2ip=Position[diagnost2ip,True]` %Those series who fulfill BL2.
15. `simulpinfpok=simulpinfp[[Union[cases1ip,cases2ip]]]` %Selects price series that fulfill BL1 or BL2.

- Splitting of price index series complicates the fulfillment of Benford's Law (but only on prices)

  1. `RandomSeed[seed3]`
  2. `splitlong=Table[10000*10^(RandomReal[1, 1200]),5000]` %simulation of 5000 series of length 1200 to be splitted and merged.
  3. `splitlongord=Sort[splitlong]` %Ordered, so as to give them aspect of price series.
  4. `diagnost1s=Map[chitest[#,1,.05]&,splitlongord]` %first digit.
  5. `diagnost2s=Map[chitest[#,1,.05]&,splitlongord]` %first two digits
  6. `cases1s=Position[diagnost1s,True];`
     `cases2s=Position[diagnost2s,True]`
  7. `splitok=splitlongord[[Intersection[cases1s,cases2s]]]` %As usual, for robustness we build a set of series that satisfy BL in both one and two significant digits.
  8. `splitseries=splitok[[Join[Table[i,i,1,101],`
     `Table[i,i,501,600],`
  9. `Table[i,i,1100,1200]]]]` %Builds merged series from cases 1-100,501-600 and 1101-1200 for each simulated price series (final length:300+2). Extra points are used to merge series.

10. `merge[list]:=m1=list[[101]];m2=list[[102]];`
    `m3=list[[201]];m4=list[[202]];a1=m2/m1;a2=m3/m4;`
    `Return[Join[a1*list[[1;;100]],list[[102;;201]],a2*list[[203;;302]]]]`
    % Merges list, makes all 3 segments compatible (as in a base-year change).
11. `splitmerged=Map[merge,splitseries]` %price series is merged.
12. `diagnost1sm=Map[chitest[#,1,.05]&,splitmerged]` %first digit, merged series.
13. `diagnost2sm=Map[chitest[#,2,.05]&,splitmerged]` %first two digits, merged series.
14. `cases1sm=Position[diagnost1sm,True];cases2sm=Position[diagnost2sm,True]`
15. `splitmergedok=splitmerged[[Union[cases1sm,cases2sm]]]`
    %Selects price series that fulfill BL1 or BL2 [none].
16. `splitmergedinf=Map[infl,splitmerged]` % computes inflation to merged series.
17. `diagnost1si=Map[chitest[#,1,.05]&,splitmergedinf]` %first digit.
18. `diagnost2si=Map[chitest[#,2,.05]&,splitmergedinf]` %first two digits.
19. `cases1si=Position[diagnost1si,True];cases2si=Position[diagnost2si,True]`
20. `splitinf=splitmergedinf[[Union[cases1sm,cases2sm]]]` %Selects inflation series that fulfill BL1 or BL2.

- Rounding does not affect fulfillment of Benford's Law

  1. `simulinfokr=Map[Round[#,0.0001]&,simulinfok]`%Rounds previous simulated inflation series to two decimals.
  2. `diagnost1r=Map[chitest[#,1,.05]&,simulinfokr]` %first digit, rounded series.
  3. `diagnost2r=Map[chitest[#,2,.05]&,simulinfokr]` %first two digits, rounded series.
  4. `cases1r=Position[diagnost1r,True];cases2r=Position[diagnost2r,True]`
  5. `simulinfokrok=simulinfokr[[Union[cases1r,cases2r]]]` %Selects rounded inflation series that fulfill BL1 or BL2.

References

[1] Beer, T. (2009). Terminal Digit Preference: Beware of Benford's Law. *Journal of Clinical Pathology* **62**: 192.

[2] Cavallo, A. - Cruces, G. - Pérez-Truglia, R. (2016). Learning from Potentially-Biased Statistics: Household Inflation Perceptions and Expectations in Argentina. *NBER Working Paper* **22103** (http://www.nber.org/papers/w22103).

[3] Dabús, C., Delbianco, F. and Fioriti, A. (2016). High Inflation, Price Stability and Hysteresis Effect: Evidence from Argentina. *Revista de Análisis Económico – Economic Analysis Review* **31**: 59–71.

[4] Dieckmann, A. (2007). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics* **34**: 321–329.

[5] Durtschi, C. - Hillison, W. - Pacini, C. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data.*Journal of Forensic Accounting* **7**: 17–34.

[6] Editorial Staff (2017). Argentina's New, Honest Inflation Statistics: the End of Bogus Accounting. *The Economist* **May 25** (https://www.economist.com/news/americas/21722694-end-bogus-accounting-argentinas-new-honest-inflation-statistics).

[7] Martinez, L. (2018). How Much Should We Trust the Dictator's GDP Estimates? *Harris School of Public Policy - University of Chicago working paper.*

[8] Ng, S. (2017). Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data. *NBER Working Paper* **23673** (http://www.nber.org/papers/w23673).

[9] Nigrini, M. (1996). A Taxpayer Compliance Application of Benford's Law. *Journal of the American Tax Association* **18**: 72–91.

[10] Pinkham, R. (1961). On the Distribution of First Significant Digits. *Annals of Mathematical Statistics* **32**: 1223–1230.

[11] Raimi, R. (1976). The First Digit Problem. *American Mathematical Monthly* **83**: 521–538.

[12] Rauch, B. - Göttsche, M. - Brähler, G. - Engel, S. (2011). Fact and Fiction in EU-Governmental Economic Data. *German Economic Review* **12**: 243–255.

[13] Varian, H. (1972). Benford's Law (Letters to the Editor). *The American Statistician* **26**: 65.