ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022
ISBN 978-987-28590-6-0

On the Possibility of Using Benford's Law to Analyze Manipulation of Inflation Statistics

Miranda Zanetti, Maximiliano Gabriel
Delbianco, Fernando
Tohmé, Fernando Abel

On the Possibility of Using Benford's Law to Analyze Manipulation of Inflation Statistics

Maximiliano Miranda Zanetti*[1], Fernando Delbianco†[2], and Fernando Tohmé‡[3]

[1] *Departamento de Economía - Universidad Nacional del Sur*
[2] *Departamento de Economía - Universidad Nacional del Sur, , INMABB, CONICET*
[3] *Departamento de Economía - Universidad Nacional del Sur, , INMABB, CONICET*

Dated: September 1, 2018

Abstract

Benford's law is a well-known regularity in the frequency distribution of first (or most significant) digits found in many data sets. This regularity has been employed in the past to perform forensic analysis of data, including macroeconomic series.

In this paper, we propound the usage of the regularity to study possible manipulations of price and inflation index series. To this end, we evaluate fulfillment of the regularity in macroeconomic series of price and inflation for some representative countries. We resort to simulations to study the relationship between price and inflation indexes regarding compliance of Benford's law, possible problems of analyzing the regularity using price indexes (mainly, due to splicing of series, and base-year re-scaling), and how the regularity could be lost in the inflation series when some simple manipulations are performed over inflation data.

Keywords: Benford's Law; data tampering; inflation; price series. **JEL Codes**: C12 C46

*Electronic address: `mmiranda@uns.edu.ar`; Corresponding author
†Electronic address: `fernando.delbianco@uns.edu.ar`
‡Electronic address: `ftohme@criba.edu.ar`

1 Introduction

Benford's Law is a regularity well documented in general data sets. It has been observed that general sources of data show a logarithmic distribution of the first significant number [1]. This well-known regularity has also been found to be present in some economic data sources [8]. Since such aggregates are found to generally satisfy Benford's law, absence of this regularity has been used in forensic accounting and fraud detection [12][4][11]. However, to the best of our knowledge, Benford's law has never been applied to detect manipulation of consumer price or inflation index manipulation.

The aim of this work is to analize the possibility of using Benford's law to detect manipulation in such indexes. Firstly, we perform a statistical analysis to detect whether Benford's regularity is present in the aforementioned macroeconomic indexes for data of some selected countries.

Besides the US data (as representative of a developed economy), we consider Chile (a developing country with a solid, stable macroeconomic performance and controlled levels of inflation), and Venezuela and Zimbawe (countries experiencing high and very high inflationary processes).

We find that inflation indexes on these representative countries fulfill the regularity quite well. On the other hand, series of price indexes seem to be somewhat less prone to satisfy this "'law".

In view of these findings, we explore which connections there must be between price and inflation indexes, and whether some usually applied, not fraudulent operations on the indexes could alter the fulfillment of the regularity on price indexes. Such operations include rounding, splicing and base-year changes.

In order to answer these questions, we run simulations of both price and inflation series. Benford's law is found to be a robust property of the related series, in the sense that if a price series satisfies the regularity, then the derived inflation index series also fulfills the law. Conversely, an inflation index series which is found to fulfill Benford's law, integrates into a price series that also satisfies the regularity.

For the purpose of analyzing why the regularity does not show in price series as clearly as in inflation index series, we also resort to simulations. The simple operations of index splicing and base-year rescaling[1] are performed over sets of

[1] Two typical operations performed in price indexes, specially when considering merging two periods or indexes calculated with diverging methods.

price indexes simulated to fulfill Benford's law; then, the amended series are tested for the regularity.

The results suggest simple operations like index splicing could explain the occasional failure of price index series to fulfill Benford's regularity.

To further the usage of Benford's law to study possible inflation data manipulation, we also simulate the effect of simple manipulation schemes on inflation reports that turn a complying inflation index into a series that fails to satisfy Benford's law.

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 n of digits drawn from $\{0, \ldots, 9\}$ (which by a slight abuse of language, can be seen as a natural number n): the probability of n is $\log_{10}(1 + \frac{1}{n})$. While there are series that do not satisfy this property, an 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 [9]. 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 [10] [2]. But other than in those cases, the validity of the law is pervasive. Thus,

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

it has been used as a general principle to check for fraud and manipulation of data [12] [8].

Many interesting analyses of fraud have been run on scientific data ([3]), accounting data ([4]) and, closer to our own work, macroeconomic data ([11]). The procedure is basically the same, independently of the source of data. For each series, it starts by finding the actual distribution of digits, which yields, for each $d \in \{1, \ldots, 9\}$, 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

We use series of representative countries as reference for Benford's law compliance. As indicated in the Introduction, we consider the inflation series of the USA, Chile, Venezuela and Zimbawbe, all of them in annual terms for the period 1980-2016, which we get from the World Bank database[3].

In figure 1 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.

Table 1 presents the results of running the χ^2 test to both inflation and price (annual) series for each reference country, under the null hypothesis of the validity of Benford's law. The results were obtained by using the Benford's.analysis (`https://CRAN.R-project.org/package=Benford's.analysis`) and Benford'sTests (`https://CRAN.R-project.org/package=Benford'sTests`) packages in R.

The main result is that Benford's law cannot be rejected in any case for the inflation index, while the price index series is more problematic (for Chile and Zimbabwe), except for a significance level of 1%).

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

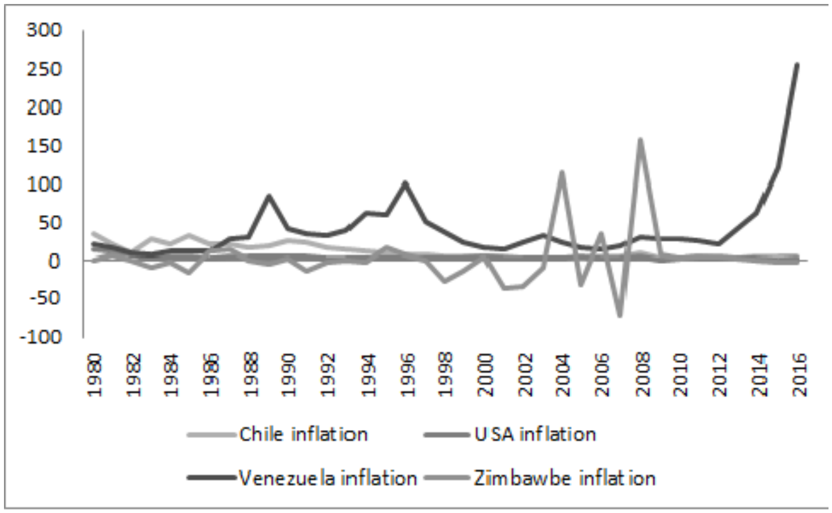


Figure 1: Inflation in control countries

Finally Table 2 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.

Weaker evidence for Benford's regularity in price series might be explained by standard bookkeeping operations, like series splicing, which may affect the distribution of first significant digits much in the same way that other types of restrictions or manipulations have been shown to affect the regularity.The next section is devoted to analyze this possibility.

Chile (1980-2016)	χ^2	**p-value**
Inflation	89.097	0.4772
Price index	110.45	0.06138
USA (1980-2016)	χ^2	**p-value**
Inflation	65.203	0.9727
Price index	73.093	0.889
Venezuela (1980-2016)	χ^2	**p-value**
Inflation	78.893	0.7697
Price index	65.264	0.9723
Zimbabwe (1980-2016)	χ^2	**p-value**
Inflation	77.387	0.8054
Price index	112.67	0.04585

Table 1: Inflation and Prices in Control Countries 1980-2016

Controls	Chile	USA	Venezuela	Zimbabwe
Inflation	✓	✓	✓	✓
Price 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 2: Summary of results

4 Simulations

The results in the previous section indicate that there is room for holding that both price index series and inflation index series satisfy Benford's law. 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, 10^X trivially satisfies Benford's Law, since its logarithm is uniformly distributed. We can thus use this procedure to generate "‘price" series satisfying Benford's Law.[4]. Again, we apply the χ^2 test to detect deviations from Benford's law.

First we analyze the robustness of Benford's's regularity on both price and the derived inflation series. Our results (see Figures **??** and **??**) indicate that even if the price index series does not satisfy Benford's law its derived inflation series may satisfy it. Table 3 shows the robust inheritance of Benford's law compliance.

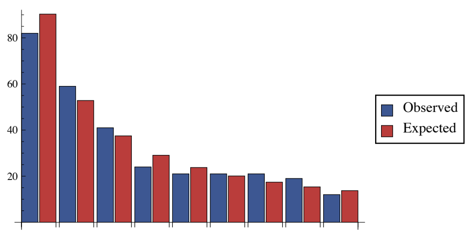
We also test whether an inflation index fulfilling Benford's law translate into an [integrated][5] price index that also satisfies the regularity. The results are also quite robust [see table 4.]

With respect to effects of splicing and merging series, we can see in Figure 3a that applying these operations on simulated price series, even if the resulting price serie fails to satisfy Benford's's law, the ensuing inflation series shown in Figure 3b indeed satisfies the law.

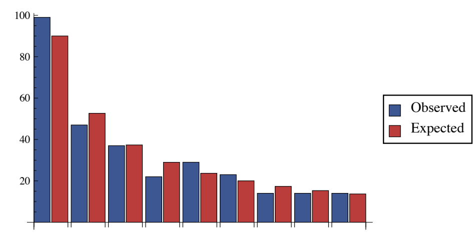
Figure 3 shows how fulfillment of Benford's law could be lost for a price series subject to standard bookkeeping manipulations, while the regularity is maintained in the derived inflation series.

[4]To run the simulations with `Mathematica` using its built-in random generator. Alternative generators (Mersenne-twister, etc.) did not yield noticeably different results.

[5]To build the price series, a series of anchors satisfying Benford's law were employed, though using instead a uniform distribution did not change results.

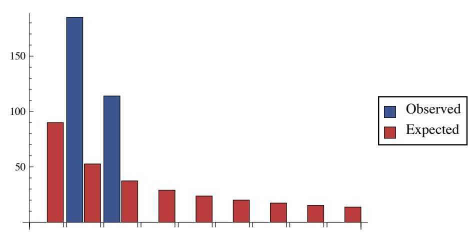


(a) Price series

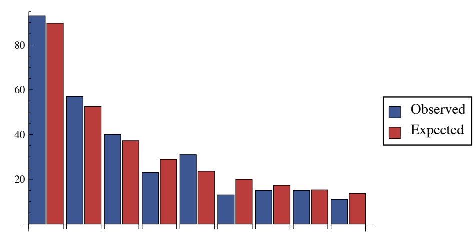


(b) Inflation index

Figure 2: Representative histogram of (a) a simulated price series and (b) its derivated inflation index, showing fulfillment of Benford's's law for both series.



(a) Price series



(b) Inflation index

Figure 3: Histograms of (a) a splitted and merged simulated price series and (b) the derivated inflation index, showing non-compliance of Benford's's law for the merged price series but fulfillment of the law for the inflation index.

According to these results, the conclusion is that if Benford's law is a general feature of types of economic data like price indexes [and the reference countries' data seem to support such claim], then: a) inflation indexes could also be used to test Benford's law, since the regularity is quite roboustly transferred between series; b) some simple operations on the price index can make the series lose fulfillment of Benford's law. This suggests using mainly the inflation index to analyze possible fraudulent manipulations.

Is it possible to detect deceitful manipulation of inflation data with Benford's law testing? To answer this question, we perform additional simulations to study how simple fraudulent operations on the inflation series can affect fulfillment of Benford's law. Three manipulations are analyzed:

1. Chop of very large inflation numbers [considering the ceiling as two standard deviations above the mean, any number above such ceiling is trimmed to that limit].

2. Omission of large inflation numbers [inflation data that lie one standard

deviation above the mean are discarded[6]].

3. Reduction of very large inflation numbers (as defined above) by the factor $\frac{\mu}{\mu+\sigma}$.

In every case, the manipulation destroyed the original compliance of Benford's law.

5 Conclusions

In this paper we assessed the claim that price and inflation index series are part of the great set of data that seem to fulfill Benford's law. As discussed above, inflation series show a better fit in terms of χ^2 tests.

Since price indexes are subject to simple, commonplace manipulations that may force the series to reject the null hypothesis of fulfilling Benford's regularity, a good piece of advice derived from our study is to make use of inflation indexes, rather than price series, to detect fraudulent manipulation of inflation data. There does not seem to exist any reason to think that inflation data should, in general, fail to satisfy Benford's law.

Simulations indicate that even quite simple tampering of data could in principle been detected via Benford's law forensics. Further work [6] is devoted to apply checking of Benford's law presence in inflation data to detect fraudulent manipulations in inflation series where there could be widespread suspicions of political tampering with official statistics.

[6] Inflation data cannot in general be omitted; however, we consider this kind of omission since it is the most difficult to be detected.

A Appendix: syntax of pseudocode

`arg`	argument of a function.
`f[x]`	f(x)
`a,b,c`	List of elements (a,b,c).
`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(arg),list]`	Same as above, but with function expressed with a body of argument.
`f(arg1,arg2,...)`	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

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

  `firstsignif[x,z]= FromDigits[RealDigits[x, 10, dig][[1]]]`

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

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

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

  `chitest [list,digits,alpha] = newlist=... ; index=0 ; lon =Length[list] ; labels=cats[digits]`

- Determines expected probability for each category according to Benford's Law:

- computes the chi-square statistic

  `critval= InverseCDF[Chisquare Distribution] [`$10^{\texttt{digits}} - 10^{\texttt{(digits-1)}} - 1$`,1-alpha]`

- evaluates whether Benford's law is satisfied

  `Boolean[chistat<=criticval]`

- For sintax: start, test, increment, body:

  1. **for** index=1

2. index <= lon
3. index++
4. newlist= Append[newlist, firstsignif[list[[index],digits]]]
5. frecobs = Map[Count[newlist,arg]
6. **return** labels

- Maps a counting routine to each digit or category:

  `probs= Map[Log10[1 + 1/arg], return labels`

C Appendix: pseudocodes

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

1. RandomSeed[seed1]
2. simulp = Table $[10000 * 10^{\mathrm{RandomReal}[1,300]}$,5000]
3. diagnost1=Map[chitest[arg,1,.05], **return** simulpord]
4. diagnost2=Map[chitest[arg,2,.05], **return** simulpord]
5. cases1=Position[diagnost1,True];
6. cases2=Position[diagnost2,True];
7. simulpok=simulpord[[Intersection[cases1,cases2]]]
8. infl[list]:= lon=Length[list];
9. listdif=Table[(list[[index+1]]-list[[index]])/list[[index]],
10. index,1,lon-1]
11. simulpinfl=Map[infl,simulpok]
12. diagnost1infl=Map[chitest[arg,1,.05], **return** simulpinfl]
13. diagnost2infl=Map[chitest[arg,2,.05], **return** simulpinfl]
14. cases1infl=Position[diagnost1infl,True];
15. cases2infl=Position[diagnost2infl,True];
16. simulpinflok=simulpinfl[[Union[cases1infl,cases2infl]]]

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

1. RandomSeed[seed2]
2. simulinf= Table[$10000 * 10^{\texttt{RandomReal[1,300]}}$,5000]

3. diagnost1i= Map[chitest[arg,1,.05],**return** simulinf]
4. diagnost2i= Map[chitest[arg,2,.05],**return** simulinf]
5. cases1i=Position[diagnost1i,True];
6. cases2i=Position[diagnost2i,True]
7. simulinfok=simulinf[[Intersection[cases1i,cases2i]]]
8. anchorexp=1000 10 (RandomReal[1,Length at simulinflista])
9. buildfrominf[anchor,series]:= index=1;result=; p=anchor; lon=Length[series]; Do[p=p*(1+series[[index]]); result=Append[result,p]; index++, lon]; result
10. simulinfp=Table[buildpfrominf[anclasexp[[i]],simulinfok[[i]]],i, 1,Length at anchorexp]
11. diagnost1ip=Map[chitest[arg,1,.05],**return** simulinfp]
12. diagnost2ip=Map[chitest[arg,2,.05],**return** simulinfp]
13. cases1ip=Position[diagnost1ip,True];cases2ip=Position[diagnost2ip,True]

- Splicing of price index series complicates the fulfillment of Benford's Law

1. RandomSeed[seed3]
2. splitlongord=Sort[splitlong]
3. diagnost1s=Map[chitest[arg,1,.05],**return** splitlongord]
4. diagnost2s=Map[chitest[arg,1,.05],**return** splitlongord]
5. cases1s=Position[diagnost1s,True];cases2s=Position[diagnost2s,True]
6. splitok=splitlongord[[Intersection[cases1s,cases2s]]]
7. splitseries=splitok[[Join[Table[i,i,1,100], Table[i,i,501,600],
8. Table[i,i,1101,1200]]]]
9. diagnost1sm=Map[chitest[arg,1,.05], **return**splitseries]
10. cases1sm=Position[diagnost1sm,True];
cases2sm=Position[diagnost2sm,True]

D Results of the simulations

References

[1] Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society* **78**: 551.

simulp: price index series (length of each series: 300)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1and2
5000	4767	4736	4544
Robust series: simulpok Computed inflation series			
simulpinf (length of each series: 299)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1and2
4544	3988	4113	4404

Table 3: From price index series to inflation series

simulinf: inflation index series (length of each series: 300)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1and2
5000	4764	4716	4524
Robust series: simulinfok Computed price series			
simulinfp (length of each series: 300)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1and2
4524	4522	4522	4522

Table 4: From inflation series to price index series

splitlongord: Price index series for splicing (length of each series: 1200)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1and2
5000	4762	4728	4527
Robust series: splitok Computed price series (by splitting and merging three segments)			
splitseries (length of each series: 300)			
Total n	Satisf.BL1	Satisf. BL2	Satisf. BL1and2
4527	0	0	0

Table 5: Price index series splicing

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

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

[4] 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.

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

[6] Miranda Zanetti, M. - Delbianco, F. - Tohmé, F. (2018). Tampering with Inflation Data: A Benford Law-based Analysis of National Statistics in Argentina. Working Paper.

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

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

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

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

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

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