Gambling Industry under Benford and Zipf’s Laws Approach on Financial Statements

Ionel Jianu[1], Alexandru Isaic-Maniu[2], Claudiu Brandas[3], Marius Pompiliu Cristescu[4],

Corneliu Bente[5], Claudiu Herteliu[1,6]

[1]Bucharest University of Economic Studies
[2]National Institute of Economic Research (INCE), Romanian Academy
[3]West University of Timisoara
[4]”Lucian Blaga” University Sibiu
[5]University of Oradea
[6]London South Bank University

Newcomb (1881) and Benford (1938)

- signaled that, counterintuitive, **first digit** of numbers in various datasets are not uniform (11.1% each) distributed
- the occurrence probabilities can be computed as

$$\log_{10}(1 + 1/d) \text{ where } d=1,2,\ldots,9$$

Hence, the expected probabilities for **first digit** are:

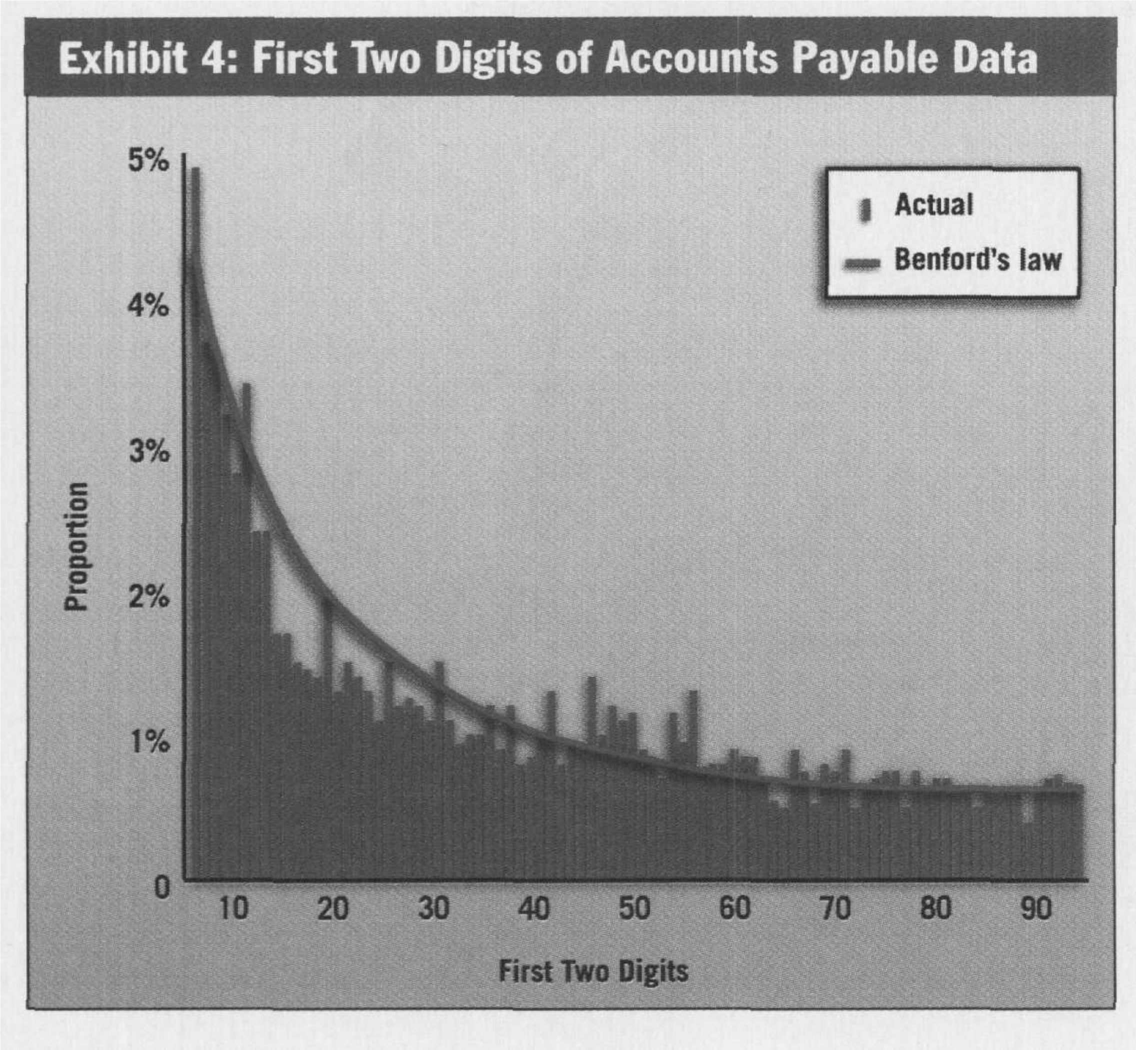
1	2	3	4	5	6	7	8	9
0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

The expected probabilities can be generalized for any digit. For **second digit** we will have:

0	1	2	3	4	5	6	7	8	9
0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085

Nigrini (1999)

- suggests that Benford law can be used “to uncover fraud and other irregularities”



Source: Nigrini (1999), p. 83

The current paper's aim:

- Context:
  - fiscal evasion proven to be high in Romania
  - gambling industry could be per se a non-conformist distribution

- *The main hypothesis which is tested basically claims that the gambling data in Romania should not obey Benford law.*

Data and methods (1)

- Fiscal records for all 574 221 registered enterprises in Romania in 2013 are used
- Public information is available for:
  - turnover (in local currency – lei)
  - gross outcome (lei) (profit or loss)
  - number of employees
  - the main activity domain is available (4 digit NACE classification)

Data and methods (2)

- Sleeping enterprises are signaled by three flags:
  - If turnover is zero
  - If number of employees is zero
  - If outcome is zero

- An enterprise is considered to be in **sleeping** mode if all three flags are simultaneously raised.

Data and methods (3)

- The gambling sector is identified through NACE classification
- Section R - Arts, Entertainment and Recreation, #92 division, #920 group (Gambling and betting activities)
- The analyses variable is **gross outcome** (both profit or loss are considered)

Data and methods: Zipf law (4)

- Is derived from power law type distributions family
- The general form of the Zipf's model is:

$$y(r)=\beta_0 r^{-\alpha}$$

- where y(r) is the size of y variable written as a function of its rank (r)
- After using logarithm in both parts the linear version is obtained:

$$\log(y(r)) = \beta_1 - \alpha \log_{10} r$$

Data and methods: Yule-Simon model (5)

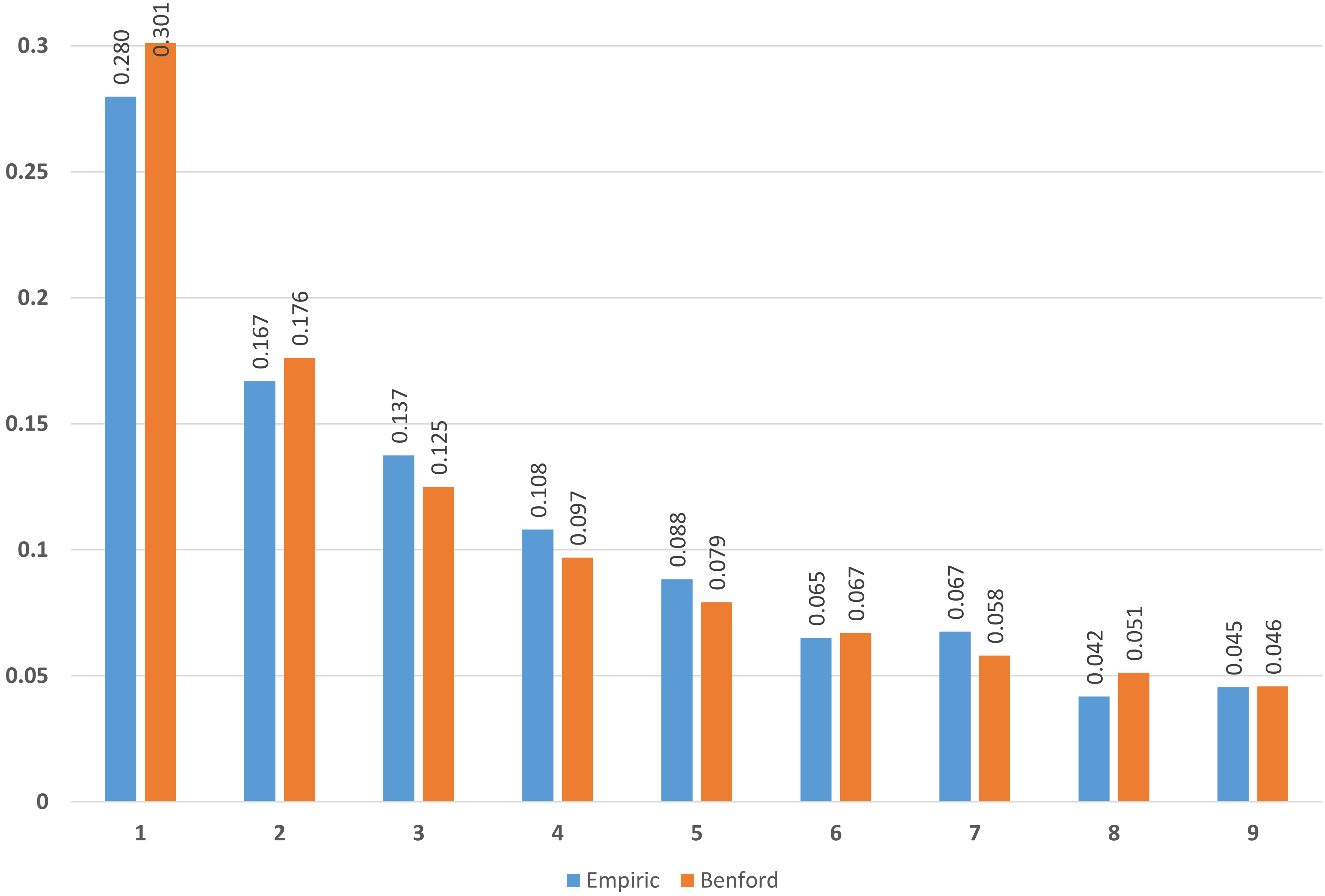
- Is derived from Zipf's law
- The general form of the Yule-Simon model is:

$$y(r)=\beta_0 r^{-\alpha}(N-r+1)^{\gamma}$$

- where N is the sample size.
- After using logarithm in both parts the linear version is obtained:

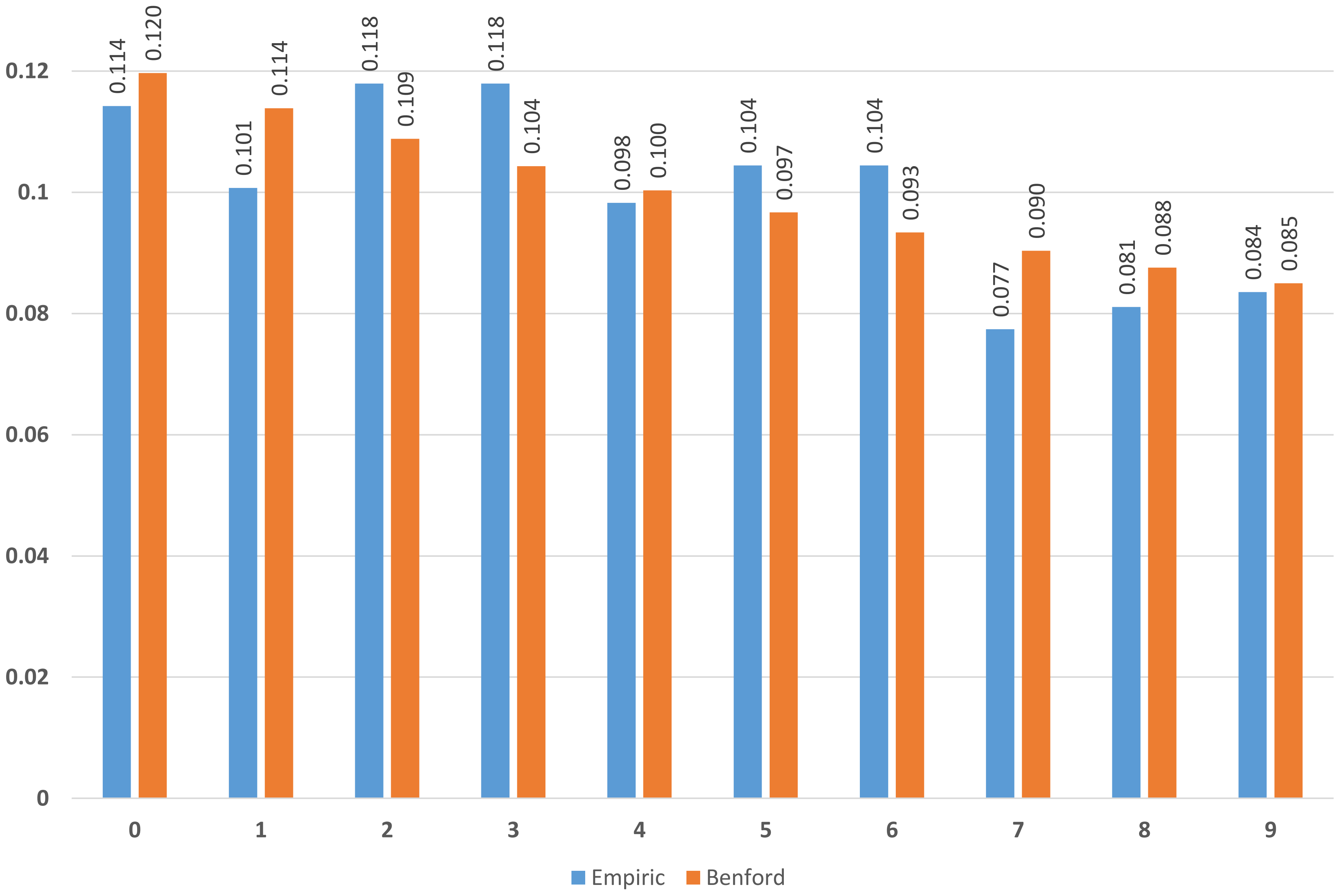
$$\log(y(r)) = \beta_1 - \alpha \log_{10} r + \gamma \log_{10}(N-r+1)$$

Gambling & Betting (NACE = 9200) – Benford law 1st digit fit



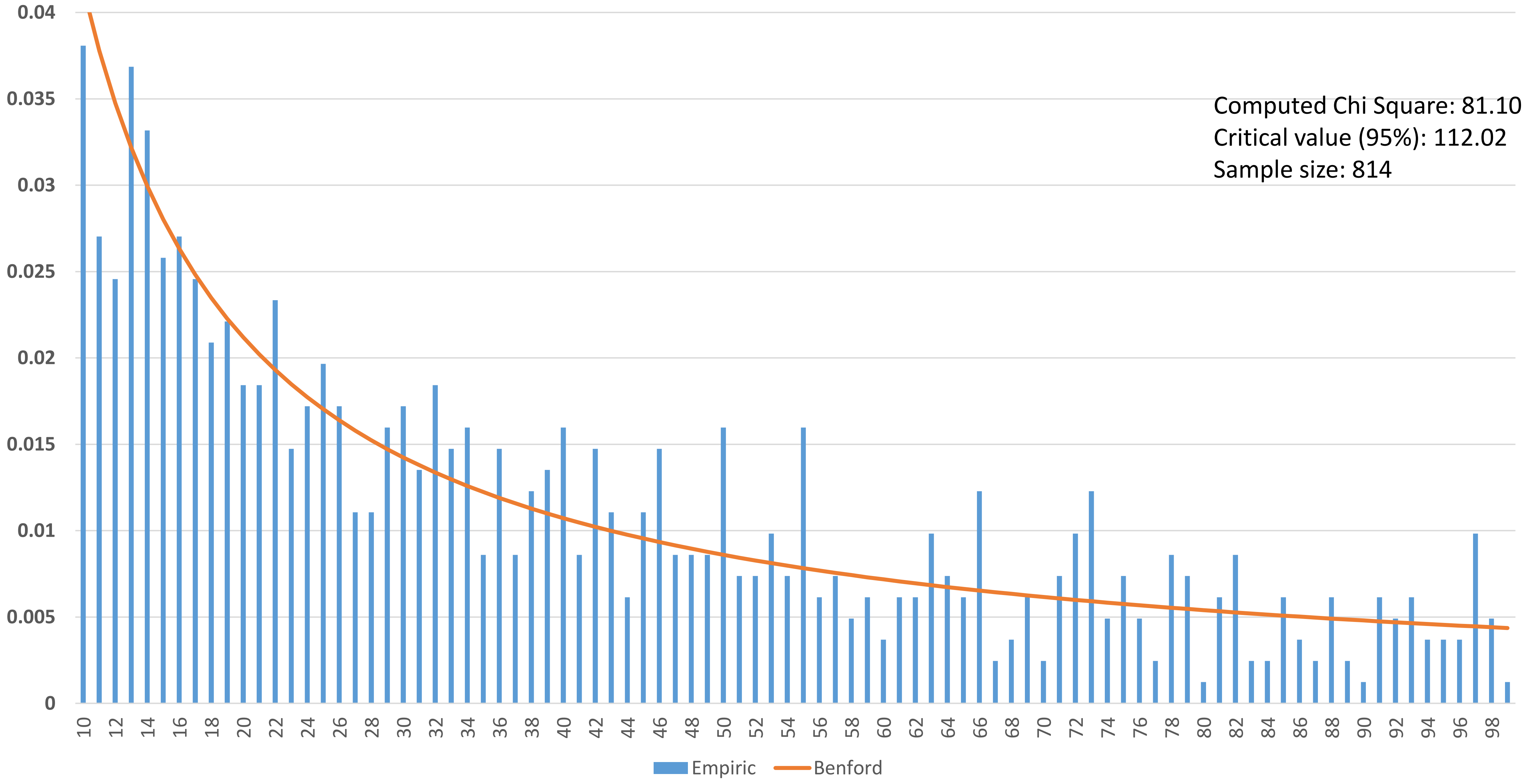
Computed Chi Square: 7.26
Critical value (95%): 15.51
Sample size: 815

Gambling & Betting (NACE = 9200) – Benford law 2nd digit fit



Computed Chi Square: 7.03
Critical value (95%): 16.92
Sample size: 815

Gambling & Betting – Benford law: first two digits combined



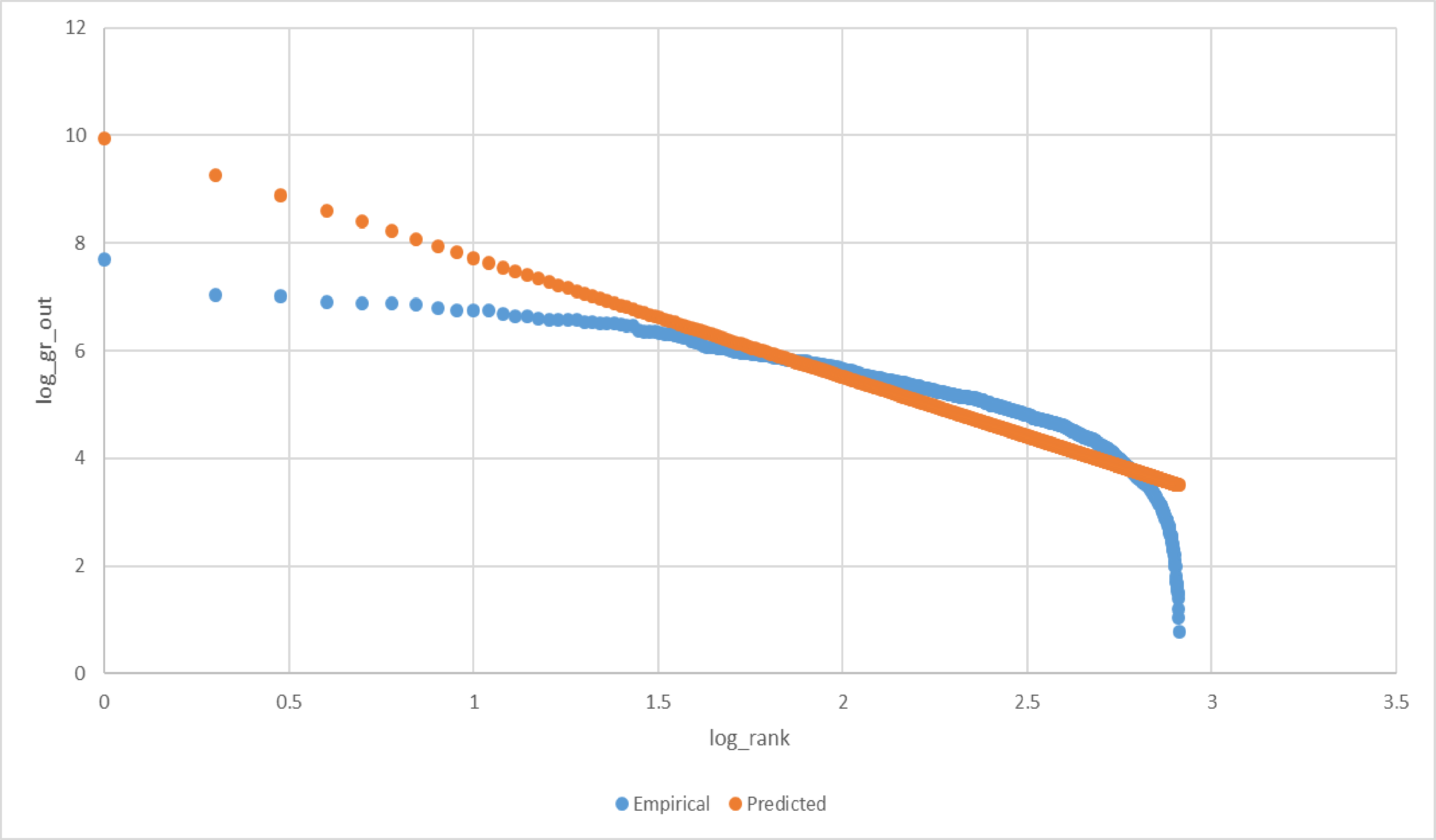
Brief information about Chi Square testing (χ^2) for checking (non)conformity within Benford's law – gambling industry

Indicator	First digit	Second digit	First two combined digits
Visual localization	Figure 1.	Figure 2.	Figure 3.
Computed χ^2	7.26	7.03	81.1
Critical value (χ^2) for a significance level of α=0.05	15.51	16.92	112.02
Degrees of freedom	8	9	89
Sample size	815	815	814

Zipf's model outcomes

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.875576					
R Square	0.766632					
Adjusted R Square	0.766345					
Standard Error	0.521119					
Observations	815					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	725.2882	725.288	2670.775	4.1E-259	
Residual	813	220.782	0.27156			
Total	814	946.0702				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	9.935438	0.107617	92.32229	0	9.724198	10.14668
log_rank	-2.21085	0.04278	-51.6795	4.1E-259	-2.29482	-2.12688

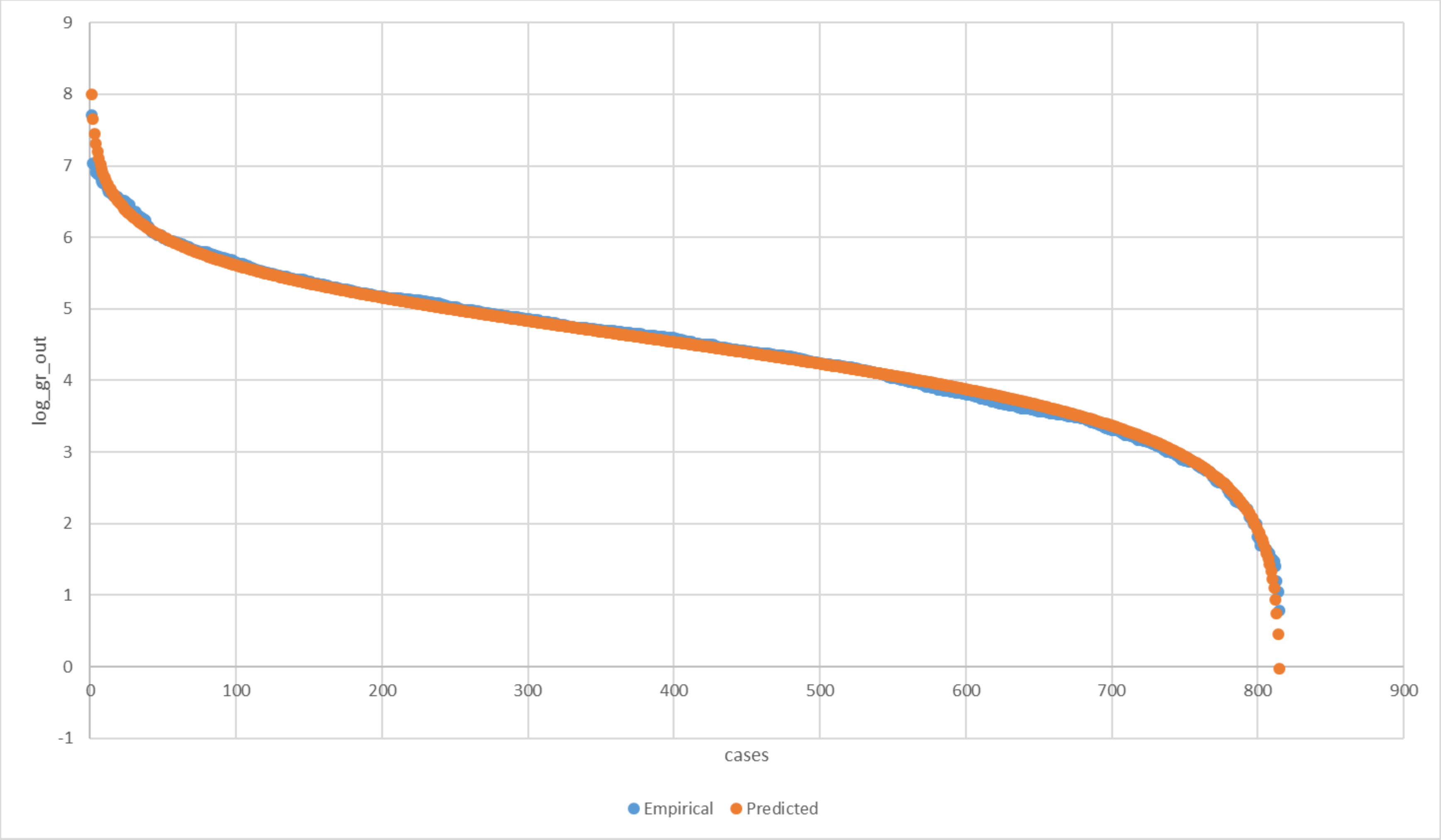
The Zipf model – theoretical versus empirical values



Yule-Simon model outcomes

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.99777					
R Square	0.995545					
Adjusted R Square	0.995534					
Standard Error	0.072048					
Observations	815					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	941.8552	470.9276	90722.06	0	
Residual	812	4.214997	0.005191			
Total	814	946.0702				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3.325066	0.035619	93.3497	0	3.255149	3.394983
log_rank	-1.15133	0.007867	-146.35	0	-1.16677	-1.13589
log(N-r+1)	1.60688	0.007867	204.2564	0	1.591438	1.622322

Yule-Simon model– actual versus theoretical data points (universal rank-size law: Ausloos &Cerqueti, 2016)



Conclusions

- financial reports (more precisely the gross result obtained) of the companies operating in Romania in the gambling industry are folded on a distribution derived from the power function
- contrary to expectations, however, the distribution was not a Zipf type but one specific to the Yule-Simon law
- very likely the fact that in the gambling industry we are dealing with a player having a dominant position (Romanian National Lottery Company) has led to this state of affairs (law of distribution)

Selected references

- Ausloos, M., & Cerqueti, R. (2016). A universal rank-size law. PloS one, 11(11), e0166011.
- Ausloos, M., Herteliu, C., & Ileanu, B. (2015). Breakdown of Benford's law for birth data. Physica A: Statistical Mechanics and its Applications, 419, 736-745.
- Benford, F. (1938). The law of anomalous numbers. Proceedings of the American philosophical society, 551-572.
- Mir, T. A., Ausloos, M., & Cerqueti, R. (2014). Benford's law predicted digit distribution of aggregated income taxes: the surprising conformity of Italian cities and regions. The European Physical Journal B, 87(11), 261.
- Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics, 4(1), 39-40.
- Nigrini, M. J. (1999). I've got your number. Journal of accountancy, 187(5), 79-83.
- Nigrini, M. J. (2015). Persistent patterns in stock returns, stock volumes, and accounting data in the US capital markets. Journal of Accounting, Auditing & Finance, 30(4), 541-557.
- Reed, W. J. (2001). The Pareto, Zipf and other power laws. Economics letters, 74(1), 15-19.
- Shi, J., Ausloos, M., & Zhu, T. (2018). Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries. Physica A: Statistical Mechanics and its Applications, 492, 878-888.
- Stanley, M. H., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firms. Economics letters, 49(4), 453-457.

Thank you!