Benford's Law-test on trial. Simulation-based application to the latest election results from France and Russia.

Verena Mack* and Susumu Shikano†

6. April 2013

Abstract

While the application of Benford's Law (BL) to detect election fraud has gained popularity in recent time, concerns about its validity have also been expressed. It is known that the application of the BL-test to first digits is problematic. Concentration of precincts votes in a certain range can boost the BL statistic. Some argue that the use of second digits instead of first digits would solve this problem. This is, however, no solution since concentrated precincts votes appear in certain circumstances which can affect the distribution of even second digits. In this paper, we apply 2BL and an alternative distribution systematically to different institutional settings. More specifically, we investigate the latest parliamentary and presidential elections of France (both 2012), with no suspicion of fraud, and Russia (2011 and 2012), with strong suspicion of fraud. Finally, we replicate another detection method for the purpose of cross validation and compute simple fraud scenarios to assess the performance and mechanisms of 2BL. We can identify a circumstance when 2BL gives misleading signals and have to conclude that 2BL is inappropriate for fraud detection.

*Graduate School of Decision Sciences, University of Konstanz

†Department of Politics and Public Administration, University of Konstanz

1 Introduction

Assessing the legitimacy of elections is what all studies have in common which investigate election frauds, however, approaches of such investigation differ strongly. In this field, digit based methods have gained popularity in recent time. This is mostly due to their simple application, straight forward interpretation and the minimally required information (vote counts of elections). Most famous is the second digit Benford's Law test (2BL) which is according to Pericchi and Torres (2011, 15) "becoming a standard tool on what has been termed by Mebane as "Election Forensics"". While 2BL has been increasingly utilized , concerns about its validity and usefulness have been increasingly raised as well (Mebane, 2010; Deckert, Myagkov and Ordeshook, 2011; Mebane, 2011; Shikano and Mack, 2011; Mebane, 2012).

Previous findings suggested that vote count distribution do not necessarily fit the assumptions to produce 2BL like digit distributions (compare Shikano and Mack, 2011). In line with these finding, we validate the performance of the 2BL test in different institutional settings. More specifically, we investigate the latest parliamentary and presidential elections of France (both 2012), with no suspicion of fraud, and Russia (2011 and 2012), with strong suspicion of fraud. This setting also follow up suggestions and argumentation of Deckert, Myagkov and Ordeshook (2011) and Mebanes replay (compare Mebane, 2011). According to Deckert, Myagkov and Ordeshook (2011), assessment of the value of BL requires its application to elections in which we have a priori knowledge whether significant fraud occurred or not. While they investigate actual fraudulent election, they only investigate simulated data as fraud free election. However, as (Mebane, 2011, 270) argues: "there is no reason to believe that the simulated data resemble the data from any actual election" . Further, we replicate another detection method concerning election fraud in Russia for cross validation of the BL method. We will show that neither deviation from the native nor deviation from the simulation based alternative distribution correspond in a satisfying way with obvious fraud indication. By focusing on specific deviating digits and the corresponding vote count distribution we will show violations of the BL assumption which have been neglected by most scholars. This aspect is missed by Deckert, Myagkov and Ordeshook (2011) and Mebane who investigate only the mean of digit distributions for reasons of simplicity. Instead, we keep focusing on digit distributions and get results which are in line with our previous findings in respect to the position and homogeneity of the vote count distribution.[1]

The paper is structured as following: The next section will shortly summarize the idea of digit based test and the main criticism which has been raised. The third section will introduce the data from France and Russia and its fit to the 2BL distribution. Subsequently we will conduct cross validations with another detection method and simple fraud scenarios to assess the performance and mechanism of 2BL. We will summarize and discuss the paper in the last section and suggest possible further research topics.

2 Controversy over Benford's Law based methods

Benford's Law has been found in the empirical observation that leading digits of numerical data are often not uniformly distributed (Newcomb, 1881; Benford, 1938). The formal proof by Hill (1995, 360) stats "if probability distributions are selected at random and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant-digit frequencies of the combined sample will converge to the logarithmic distribution". If digits of vote counts in normal or more specific fraud free election follow a Benford like distribution, deviation of such a distribution could indicate election fraud since the digit generating random process is interrupted by a certain artificial non-random process. If digit deviation could detect election fraud, we would have a powerful tool to assess the legitimacy of an election since we only need vote counts. This idea has gained a lot of popularity within and outside of science as many scholars could show that digits in vote count indeed seem to follow a Benford like distribution. More specifically, most scholars analyzed the distribution of not the first, but the second digit in vote counts. This is because of an inherent characteristic of many electoral systems that the more or less constant precinct size produces a certain frequent digit in the first digit of vote counts in the absence of election fraud (compare Brady, 2005; Mebane, 2006). Mebane (2011, 269) further argues that "[i]f this is true, however, then Benford's Law simply does not apply to vote counts at all" which means that

[1]Mebane only started to investigate the mean of digit distributions in his latest work (Mebane, 2007*b*, 2010, 2012), however this is when doubts about the 2BL test have been raised.

even if the second digit of vote counts has a similar probability distribution as 2BL it can only follow a Benford-like distribution. The test statistic he suggested and mostly applied by other scholars is based on the Pearson's χ^2 statistic:

$$\chi^2_{2BL} = \sum_{i=0}^{9} \frac{(d_i - dq_i)^2}{dq_i}, \tag{1}$$

where q_i denotes the expected relative frequency of i at the second digit, d_i is the empirical frequency of second digit i in a constituency and d is sum of precincts in the constituency.[2] This statistic is assumed to be distributed according to a χ^2 distribution with 9 degrees of freedom. Therefore, we can evaluate significance of deviation of empirical data from the Benford Law using a critical value of 16.9 at a significance level of 5%.

There are a series of application of 2BL(-like) to US-presidential election (Mebane, 2006, 2007*a*, 2008*a*), elections of Mexico, Indonesia, Russia, Iran, Germany, Ukraine, Puerto Rico and Venezuela (Mebane, 2007*a*; Mebane and Kalinin, 2009; Mebane, 2010; Shikano and Mack, 2011; Deckert, Myagkov and Ordeshook, 2011; Pericchi and Torres, 2011).[3] Further applications use variations in the investigated digit: Roukema (2009) investigated the first digit at elections in Iran, Beber and Scacco (2012) analyzed the last and next to last digit of election results in Sweden, Senegal and Nigeria. Weidmann and Callen (2012) also applied the last and next to last digit method to Afghanistan election. The mean of first digit is used to investigate the Argentinean election by Cantú and Saiegh (2011) . However, they combine this digit test with a calibration method based on simulation and they further implement a training set for fraud free and fraudulent distributions. Such a variety of different applications makes the deeper understanding of digit based tests beyond its simple technical use crucial.

Recently, scholars started a strong dispute whether Benford's Law is a valid and adequate tool to detect election fraud. Deckert, Myagkov and Ordeshook (2011, 260) conclude that it's not a "universally applicable magic box into which we plug election statistics and out of which comes and assessment of an election's legitimacy".

[2] This formula and more specific information can be found in Mebane (2008*b*, 179) and Shikano and Mack (2011).

[3] Some countries have been studied by different scholars and with slightly different approaches which we did not specifically repeat here.

In their study they first simulated fraud free and fraudulent election data and give the Type 1 and Type 2 errors for the mean digit deviation. Subsequently they apply 2BL to Ukraine and Russian elections for which they have prior knowledge of fraud. For both, simulated and empirical data the law seems to perform poorly for detecting election fraud. While Mebane (2011) agrees with Deckert, Myagkov and Ordeshook (2011) that the mean of second digit is significantly different from what we would expect under Benford's Law he also points out inconsistence in their argument and analysis as the inadequate simulation which does not necessary reflect any real election. Both scholars only consider deviation from the mean second digit which neglects important information, the deviation in digit pattern. In this regard, Beber and Scacco (2012) and earlier work from Mebane (2006) attributed inflating frequency of specific digits to human incapability to produce real random numbers.[4]

In contrast to such psychological interpretation we find a more simple and highly relevant feature in terms of BL. Inflated frequency of specific digits can give information about a specific range in which the vote count distribution has a high density. We have to remind us of the underlying assumption for vote counts behind Benford's Law. As Fewster (2009) pointed out, "[D]ata from *any* distribution will tend to be 'Benford', as long as the distribution spans several integers on the $\log_{10}$ scale - several orders of magnitude on the original scale - and as long as the distribution is reasonably smooth". This issue has only been addressed in Shikano and Mack (2011) where we can identify a boosted 2BL statistic within a specific range of vote counts and in combination with a certain homogeneity of the vote count distribution. As we are, a priori, quite certain that German parliamentary election data from 2009 are not fraudulent this finding can be attributed to the violation of the assumption to produce 2BL like digits. To make the 2BL test statistic independent from Benford's Law and capable to handle such vote count characteristic, Shikano and Mack (2011) suggest an alternative distribution based on simulated vote counts. The method extended an earlier approach by Mebane (2007*b*) by relying on the model for multi-party elections by Katz and King (1999) (For more detailed description see Shikano and Mack, 2011).

In the next section, we will follow up these findings by applying 2BL to different

[4]This implies a fraud mechanism where people specifically replace vote counts, voters or even eligible voters by numbers fitting better their preferred electoral outcome. This should reflect only one of many possible ways to fraud elections.

institutional settings, fraud free and fraudulent real election data. Subsequently, we conduct cross validation with further methods to find what is most decisive for digit deviation and the detection of election fraud.

3 Application to election from France and Russia

3.1 Data

We investigate election data from France and Russia for multiple reasons. First, we needed a set of election data which is a priori known for being fraud free and another for which we have further proof of fraud. Second, we still do not fully know whether the institutional setting may have some effect on digit distributions. Third, it is known that homogeneity in the vote count distribution and its limited range caused strong deviation from the expected digit distribution in German parliamentary election. Thus, election data with more variation in respect to unit size and heterogeneous vote count distribution should give more 2BL like digits distribution.

In France, there have not been any accusation of electoral fraud or anomalies in the recent elections in 2012 nor for quite some time (compare Klimek et al., 2012). Further, French political system enables investigation of election data within different institutional settings, the presidential and parliamentary election, when everything else is equal. Last but not least, electoral units vary stronger in size than in Germany which should increase the fit of second significant digit in vote counts with the distribution expected by 2BL. Table 1 gives some information about the main quantities of the data.[5] However, we have to point out one potential drawback of the French election data. The data is aggregated at municipality levels; however, this is the smallest available aggregation level we could get in France. Therefore, vote counts of smaller municipalities will be provided at the precinct level while big municipalities which contain different precincts only give the aggregation of their data. As aggregated data in some communities have higher vote counts, this should contribute to the fit of the second significant digit distribution with 2BL.[6]

[5] What we present here as mean turnout is higher than the turnout given by official election statistics. This is because we compute turnout for each municipality which gives a not weighted average across municipalities. This equally applies to turnout of the Russian election in Table 1 .

[6] Data from France are available here: `http://www.interieur.gouv.fr/Elections`.

Table 1: Election from France 2012 and Russia 2011, 2012

Election	N obs.	N units	mean unit size	mean el. voter	mean voters	range voters	mean turnout
Fr Parl. 2012	36930	104	355	1247	714	4 - 48949	57%
Fr Pres. 2012	36785	103	357	1216	968	4 - 378905	84%
Ru Parl. 2011	95055	84	1132	1147	689	0 - 22671	66%
Ru Pres. 2012	95566	84	1138	1152	752	0 - 19711	70%

Russian elections do not appear to be controversial over the fact that there have been strong indications of fraud since 2003 (see Shakin, 2009; Mebane and Kalinin, 2009; Deckert, Myagkov and Ordeshook, 2011; Kalinin and Walter R. Mebane, 2011; Kobak, Shpilkin and Pshenichnikov, 2012; Klimek et al., 2012). Scholars showed that ballot box stuffing changes the shape of turnout distribution and gives high correlation between vote share and turnout (Shakin, 2009; Kobak, Shpilkin and Pshenichnikov, 2012; Klimek et al., 2012). In respect to the institutional setting Russia matches with data from France since we can assess presidential and parliamentary election. Further, electoral units are characterized by strong variation across units which should also imply less violation of the 2BL assumption. Table 1 shows that France and Russia mean eligible voters and mean voters are comparable while the mean size of units is way higher than the mean size in France.[7]

3.2 Summary of 2BL statistics in France and Russia

This part presents the naive and rescaled 2BL statistic from the four different elections in France and Russia. With the naive application we refer to the simple χ^2 test statistic which was presented in section 2. The rescaled 2BL statistic is based on the alternative distribution we obtain by simulating 1000 different sets of election results. For each set of simulated election results we compute the 2BL statistic and its 95% quantile ($\chi^2_{sim95\%}$) is the alternative critical value to 16.9 based on the chi^2 distribution. To make it more comparable with the naive 2BL statistic, we rescale the empirical 2BL using the following formula:

[7] Russian election data from 2011 and 2012 were made available by Klimek et al. (2012) at `http://www.complex-systems.meduniwien.ac.at/elections/election.html`.

$$\chi^2_{rescaled} = \frac{\chi^2_{2BL} \cdot 16.9}{\chi^2_{sim95\%}} \quad (2)$$

This rescaled 2BL statistic can be compared with the critical value of 16.9 just like the naive 2BL statistic while the former is independent from the χ^2 distribution (see Shikano and Mack, 2011).[8]

Table 2 presents the summary results for the naive and rescaled 2BL as percentage of statistics exceeding the critical value of 16.9. Emphasized values above 5% signal that digit deviation is not by chance and therefore indicate election fraud or anomalies (according to the original idea of the method). We give summary statistics for each election and candidate/party except for the parliamentary election in France 2012. The French parliamentary election does not allow us an analogous analysis to the other elections since each constituency is contested by multiple candidates from different or/and the same parties which varies across constituencies. Therefore, we compute the 2BL statistic for each candidate with a vote share more than 5% for each constituency and assess the percentage of statistics exceeding the critical value across all candidates. The percentage of deviation statistics is less than 5% which makes the comparison with the alternative distribution obsolete. While the three major candidates of the presidential election in France showed unproblematic percentage of significantly deviating 2BL statistic, the two least relevant candidates show the corresponding percentage higher than 5%. Comparing the statistics with the alternative distribution does not change this fact. On the one hand, this indicates that use of the alternative distribution does not solve the problem of misfit possibly caused by other factors than fraud. On the other hand, it is interesting to note that deviations from 2BL seem to increase with decreasing vote share (Candidates are sorted by their vote share in the table).

In contrast to the French elections, Russian election results indicate higher percentage of significant deviation of the naive 2BL statistic. Here, the most suspect Putin/United Russia do not have the highest percentage among the candidates/parties. This fact does not change much even if we use the alternative distribution (see res-

[8]If unites contained less than 10 observations (vote counts) we had to exclude them as this number is too small to compute the 2BL statistic. This is the case for three units in French parliamentary election and four in presidential election. In the case of Russia, we only had to exclude one unite from the presidential election 2012.

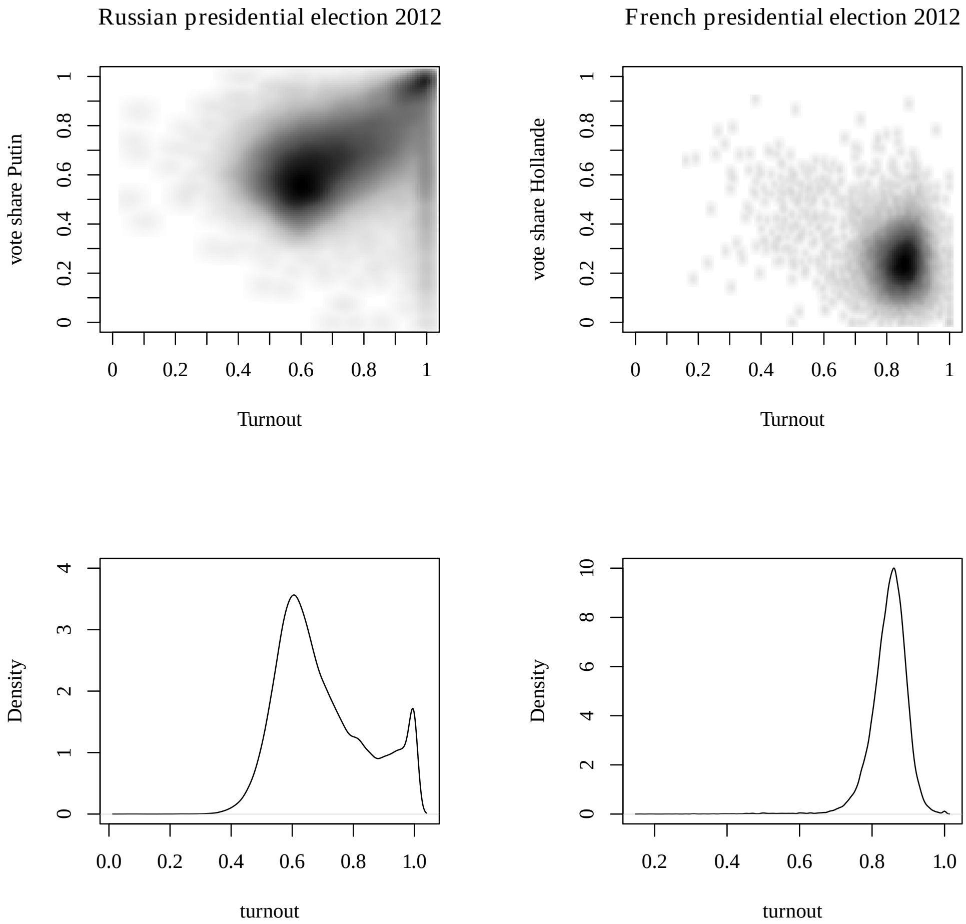
Table 2: Summary of 2BL statistics

		naive 2BL	rescaled 2BL	mean 2BL
France				
Parl. 2012	all candidates	3.9%	-	8.91
Pres. 2012	Hollande	1.0%	1.0%	8.39
	Sarkozy	1.9%	1.9%	8.20
	Melenchon	4.9%	3.9%	8.49
	Le Pen	**7.8%**	**6.8%**	8.96
	Bayrou	**9.7%**	**7.8%**	9.45
Russia				
Parl. 2011	United Russia	**17.9%**	**8.3%**	11.40
	Communist Party	**9.6%**	4.8%	9.98
	Ajust Russia	**14.3%**	**10.7%**	11.38
	LDP	**16.9%**	**13.6%**	13.62
Pres. 2012	Putin	**19.0%**	**6.0%**	12.69
	Zyuganov	**15.5%**	**8.4%**	10.89
	Prokhorov	**31.0%**	**21.7%**	14.47
	Zhirinovsky	**44.0%**	**31.1%**	20.53

caled 2BL). Interestingly we can find also in the presidential election that the candidates with lower vote share have higher percentage of deviation. The last column in Table 2 displays the mean 2BL statistic for each candidate/party. If 2BL statistic follows the χ^2 distribution we should observe values around 9. This is more or less true for French but not Russian elections. Even if we expected higher 2BL statistics for Russia, the statistic is generally high for all parties/candidates. Especially, very high 2BL statistics of the smallest parties are not in line with our prior knowledge where fraud is mainly suspected at Putin/United Russia.

4 Validation 2BL by using another detection method

The results presented above provide no clear picture whether the 2BL test can detect election fraud or not. To answer this question, we combine 2BL findings with other methods which could give strong indication that both Russian parliament



(a) (b)

Figure 1: Both upper graphics display a scatter plot of the turnout rate and vote share and the intensity of the black colour gives the density of data points. The lower graphics display the turnout distributions for France and Russia.

and presidential election are fraudulent. In particular, we will focus on the method employed by Klimek et al. (2012) who assessed the latest Russian election. Accordingly, fraud should be indicated in a peculiar shaped and bimodal turnout distribution and/or extremely high turnout in combination with extremely high vote share for the winning party (Putin/United Russia). In Figure 1(a) and Figure 1(b) we display both measures for France presidential election 2012 and Russian presidential election 2012.[9] As expected, figures for France (Figure 1(b)) indicate no election fraud, which should serve as a good benchmark for Figure 1(a). Here, we can observe two remarkable results: First the turnout rate has a bimodal distribution where a mode is at 100%. Second, the scatterplot of the turnout rate and vote share for Putin shows a concentration of units in the upper right corner. That is, there is a very high correlation between turnout and the vote share for Putin. Both facts, especially the second fact, seem to support the suspect of fraud at Putin. In the following analysis we focus more on the second fact by using the correlation between the turnout rate and vote share, which we will call "Klimek measure".

We compute the Klimek measure for each unit and candidate of our empirical and our simulated data. Since we computed the 2BL statistics at the same aggregate level, we can directly compare both measures in the different ways.[10]

First, we simply tested the possibility of any linear relationship between the 2BL statistic and Klimek measure. The result is negative. There is no linear relationship for any candidate.

If we further consider not only high vote share for Putin but also the Klimek measure we get the same insignificant results indicating that there is no relationship between obvious indications of fraud and the 2BL statistics. These findings are best summarized by Figure 2 for Russia presidential election and Figure 3 for French pre-

[9]Results from parliamentary election for both countries are very similar. They are available from the authors upon request.

[10]Beside the Klimek measure we check further: First, if the log number of problematic observations (more than 95% votes for Putin) in each unit correlates with the 2BL statistic, which is not the case. Second, divide the data into four quantiles according to their vote share which should give high concentration of likely fraudulent vote counts within the quantile with the largest vote share. This should boost the 2BL statistic and decrease across quantiles. No such pattern is reflected in the computed 2BL statistic. Third, if the suggested alternative distribution accounts for characteristics of the vote count distribution, we should be able to exclude such cases and identify relations between the rescaled 2BL statistic and the number of problematic vote counts, which we don't find.

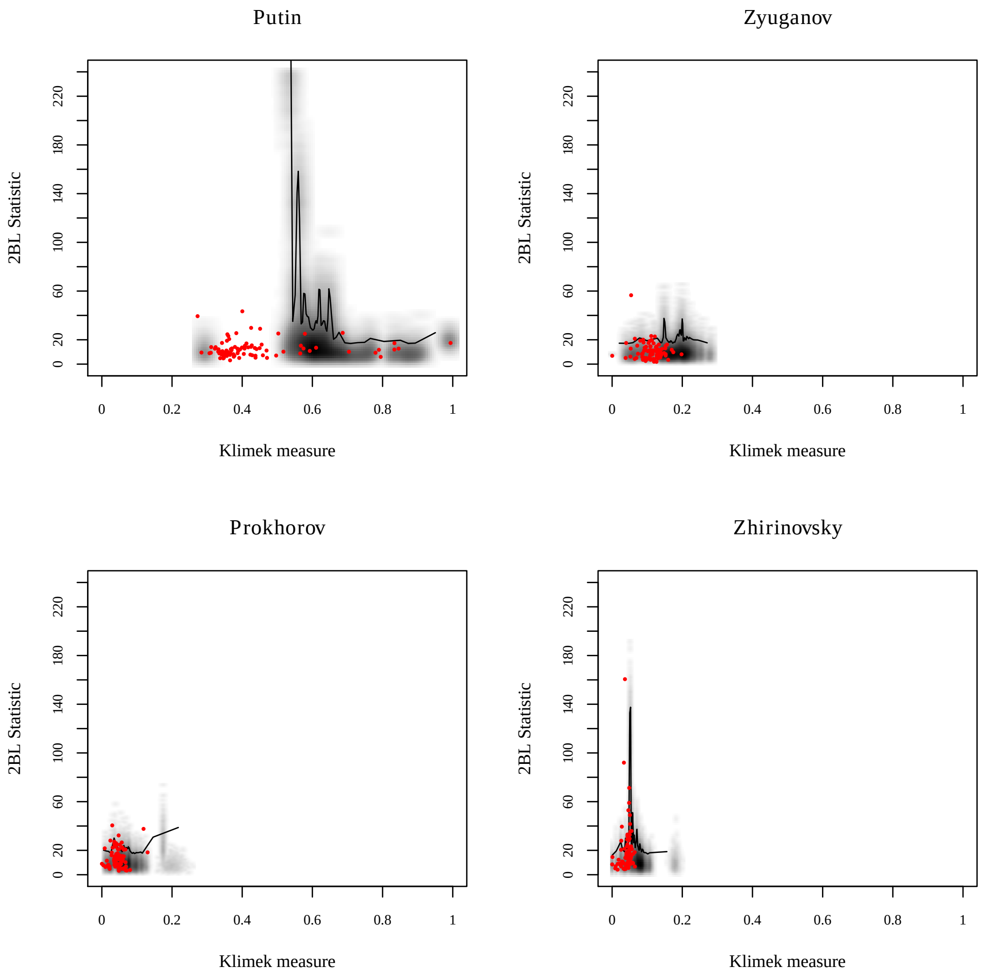


Figure 2: The red dots mark the bivariate relationship between the Klimek measure and the 2BL statistic for empirical observation. The grey shade displays the relationship between both based on the simulated data and the intensity of data points (darker colours show higher density). The black line displays the 95%-quantile of the 2BL statistic for every 2%-quantile of the Klimek measure of the simulated data.

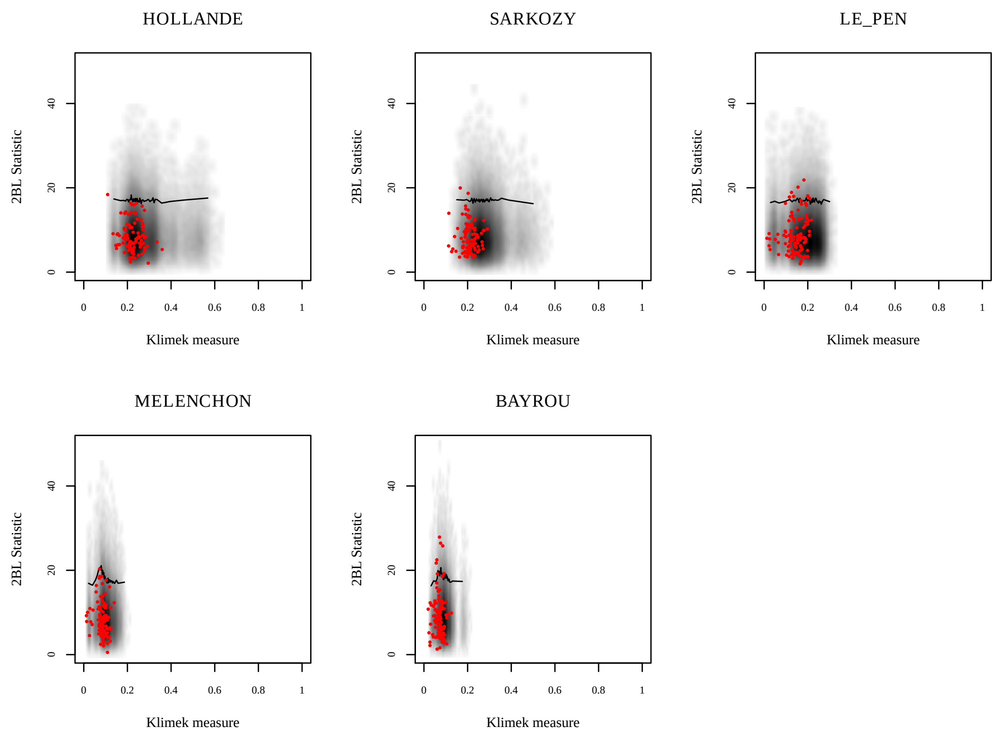


Figure 3: The red dots mark the bivariate relationship between the Klimek measure and the 2BL statistic for empirical observation. The grey shade displays the relationship between both based on the simulated data and the intensity of data points (darker colours show higher density). The black line displays the 95%-quantile of the 2BL statistic for every 2%-quantile of the Klimek measure of the simulated data.

sidential election. The red dots mark the bivariate relationship between the Klimek measure and the 2BL statistic for empirical observation. The grey shade displays the relationship between both based on the simulated data and the intensity of data points (darker colours show higher density). The black line displays the 95%-quantile of the 2BL statistic for every 2%-quantile of the Klimek measure of the simulated data. The Figures show no relationship between the Klimek measure and the 2BL statistic for any candidate, neither in France nor in Russia. Outstanding is the exploded 2BL statistic for the moderate value of around .55 for the Klimek measure of the simulated data. Those data points come from the unit 58 (the Moscow region). If we assess the second significant digits of the simulated vote counts we find strong deviations for digits 1, 2, 3, and slightly less strong deviations for digits 0, 4 and 5 as we would expect according to 2BL. If we focus on the vote count distribution we find high density for vote counts between 1000 and 1500. In this range only every additional 100 votes changes the second digit. If the distribution is concentrated in such a critical range of numbers the assumption of vote count distribution for the Benford's Law is violated. This supports our previous findings of Shikano and Mack (2011) which emphasizes the importance of the dispersion and central tendency of the vote count distribution for the 2BL statistics.

Last but not least we test the 2BL statistic by using three hypothetical fraud mechanisms which are plausible in respect to the Russian election. The first assumes that falsifier know the number of voters and pretend that all of the votes have been cased for Putin/ Hollonde. The second mechanism assumes that all eligible voters cast votes for Putin and the third assumes that falsifier get the order that the vote share for Putin should be between 90% and 100%. Table 3 gives the corresponding 2BL statistic for election data in unit one for Russia and France. The first row of the Table gives the 2BL statistic for Putin and Hollonde in unit one of the Russian and French presidential elections 2012. The second row lists the computed 2BL statistic for all voters of unit one in Russia and France according to the first mechanism. This is adequately applied to all eligible voters as suggested in mechanism two. To generate votes for Putin/ Hollande according to the third mechanism, we use random draws from a uniform distribution between 90% and 100%. Afterwards, the random draws are multiplied by voters of unit one for both elections and the 2BL statistics is computed. The 2BL statistic for scenario 1 and 3 with Russian data is

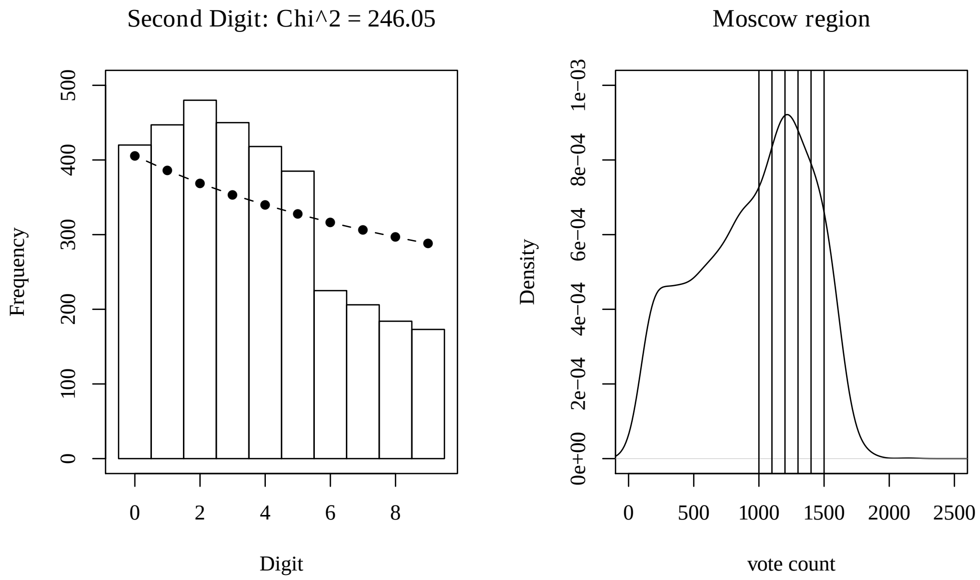


Figure 4: The Figure to the left shows a histogram which gives second digits of vote counts and the black line the expected frequency of digits according to 2BL. The Figure to the right shows the density distribution of votes for Putin in unit 58. The vertical line marks the numbers 1000 - 1500.

boosted but the same mechanism does not trigger significant boost in the French data. Therefore 2BL seems not capable to detect the presented fraud mechanism.

Table 3: Possible fraud mechanisms

	2BL Russia	2BL France
Putin/Hollande	10.62	10.28
scenario 1	**23.5**	13.44
scenario 2	6.79	8.42
scenario 3	**24.86**	15.01

5 Discussion

To give an answer to the controversy over the validity of 2BL test, this paper investigated presidential and parliamentary elections in countries with and without certain suspects of fraud. In particular, we applied 2BL test to elections in France and Russia and validated it with another detection method suggested by Klimek et al. (2012). According to our findings, 2BL test results are neither consistent with our a priori suspects of fraud nor with the results of the other method. Here, some might contest the validity of a priori suspects and/or that of the detection method of Klimek et al. (2012). Concerning the validity of a priori suspects, we can never certainly know where fraud occured e.g. in election results of Le Pen at the French presidential election. However, we have at least some reasons to trust the validity of the results based on the Klimek measure more than that of 2BL statistic. The Klimek measure has a very simple, but comprehensible model of the fraud generation process: the faker of election results replace all eligible ballots by ballots for his candidate, independently from how many ballots were actually casted. From this model, we can expect co-existence of an abnormally high turnout rate and a certain candidate's extremely high vote share if election is fraudulent. In contrast, 2BL statistic is missing such a model of the fraud generation process. Together with the fact that the Klimek measure gives more consistent test results with a priori suspects we would conclude that 2BL is inappropriate for fraud detection.

At the end, our conclusion that 2BL test is inappropriate for detection of election frauds is the same as those of further studies (e.g. Deckert, Myagkov and Ordeshook, 2011). However, this paper contributed to the existing controversy at least in the following two points: First, we systematically compared the test results based on 2BL and another method by applying it to two different political systems with and without certain suspect of fraud. Deckert, Myagkov and Ordeshook (2011), for example, analyzed the actual election results of Russia with certain suspects, while they only investigate a non-empirical data set as election results without such suspects. Second, we also identified the circumstance when 2BL gives misleading signals: violation of distributional assumption of the underlying vote counts. This aspect has been raised by Shikano and Mack (2011) who, however, investigated the election results of one specific political system. In contrast, we showed that violation of the distributional assumption can also boost the 2BL statistic under diverse political institutions.

References

Beber, Bernd and Alexandra Scacco. 2012. "What the numbers say: A digit-based test for election fraud." *Political Analysis* 20(2):211–234.

Benford, F. 1938. "The Law of Anomalous Numbers." *Proceedings of the American Philosophical Society* 78(4):551–572.

Brady, H. E. 2005. "Comments on Benford's Law and the Venezuelan Election.". Unpublished manuscript, Stanford University, January 19, 2005.

Cantú, Francisco and Sebastián M Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning." *Political Analysis* 19(4):409–433.

Deckert, Joseph, Mikhail Myagkov and Peter C Ordeshook. 2011. "Benford's Law and the detection of election fraud." *Political Analysis* 19:245–268.

Fewster, R. M. 2009. "A Simple Explanation of Benford's Law." *The American Statistician* 63(1):26–32.

Hill, T. P. 1995. "A Statistical Derivation of the Significant-Digit Law." *Statistical Science* 10(4):354–363.

Kalinin, Kirill and Jr. Walter R. Mebane. 2011. Understanding Electoral Frauds through Evolution of Russian Federalism: from Bargaining Loyalty to Signaling Loyalty. Prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Katz, J.N. and G. King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review* 93(1):15–32.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel and Stefan Thurner. 2012. "Statistical detection of systematic election irregularities." *Proceedings of the National Academy of Sciences* 109(41):16469–16473.

Kobak, Dmitry, Sergey Shpilkin and Maxim S Pshenichnikov. 2012. "Statistical anomalies in 2011-2012 Russian elections revealed by 2D correlation analysis." *arXiv preprint arXiv:1205.0741* .

Mebane, W. R. 2006. Election Forensics: The Second-Digit Benfords Law Test and Recent American Presidential Elections. Election Fraud Conference, Salt Lake City, Utah.

Mebane, W. R. 2007*a*. Election Forensics: Statistical Interventions in Election Controversies. Prepared for presentation at the Annual Meeting of the American Political Science Association.

Mebane, W. R. 2007*b*. Election Forensics: Statistics, Recounts and Fraud. pp. 12–16. Annual Meeting of the Midwest Political Science Association, Chicago, IL, April.

Mebane, W. R. 2008*a*. Election Forensics: Outlier and Digit Tests in America and Russia. In *American Electoral Process conference, Center for the Study of Democratic Politics, Princeton University*. American Electoral Process Conference, Center for the Study of Democratic Politics, Princeton University.

Mebane, W. R. 2008*b*. Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections. In *Election Fraud. Detecting and Deterring Electoral Manipulation*, ed. R. M. Alvarez, T. E. Hall and S. D. Hyde. Brookings Institution Press pp. 162–181.

Mebane, W. R. and K. Kalinin. 2009. Comparative Election Fraud Detection. Prepared for presentation at the 2009 Annual Meeting of the American Political Science Association, Toronto, Canada, Sept 3–6.

Mebane, Walter R. 2010. Election Fraud or Strategic Voting? Can Second-digit Tests Tell the Difference? Summer Meeting of the Political Methodology Society, University of Iowa.

Mebane, Walter R. 2011. "Comment on Benford's Law and the Detection of Election Fraud." *Political Analysis* 19:269–272.

Mebane, Walter R. 2012. Second-digit Test for Voters' Election Strategies and Election Fraud. Prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Newcomb, S. 1881. "Note on the Frequency of Use of the Different Digits in Natural Numbers." *American Journal of Mathematics* pp. 39–40.

Pericchi, Luis and David Torres. 2011. "Quick Anomaly Detection by the NewcombBenford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 26(4):502–516.

Roukema, B. F. 2009. "Benford's Law Anomalies in the 2009 Iranian Presidential Election.". http://en.scientificcommons.org/46955838.

Shakin, Mikhail Myagkov ; Peter C. Ordeshook ; Dimitri. 2009. *The forensics of election fraud: Russia and Ukraine*. Cambridge University Press.

Shikano, Susumu and Verena Mack. 2011. "When Does the Second-Digit Benfords Law-Test Signal an Election Fraud? Facts or Misleading Test Results." *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)* 231(5-6):719–732.

Weidmann, Nils B. and Michael Callen. 2012. "Violence and Election Fraud: Evidence form Afghanistan." *British Journal of Political Science* pp. 1–23.