



Munich Personal RePEc Archive

Theory and practice of falsified elections

Kapustenko, Oleg

Statistical Institute for Democracy

23 December 2011

Online at <https://mpra.ub.uni-muenchen.de/35543/>
MPRA Paper No. 35543, posted 23 Dec 2011 15:46 UTC

Theory and practice of falsified elections¹

Oleg Kapustenko²

Statistical Institute for Democracy

Abstract: An analysis of falsified election results is presented. A model of the falsification process is proposed and simulations are performed. The model fits well the data of the parliamentary elections in Russia on December 4, 2011. It is shown that the "noise" of false votes is well separated from the fair "signal", which can be extracted with high statistical accuracy (less than 1%) allowing quantitative reconstruction of the falsification patterns.

Introduction

Statistical analyses of the Russian elections show anomalous effects, as evidenced by numerous calculations [1-2]. The purpose of this article is the development of a model of election falsification and its simulation, as well as development of a methodology of data analysis which is able to isolate the "noise" bogus votes and obtain the "signal" of the actual election results.

Before start let's clarify the terminology an some specific of Russian vo. **The electorate** is all peoples, which are valid or entitled or eligible to vote. **The voters** or **voted** peoples are part of the electorate, which exercised their rights to vote. Their fraction relatively to the electorate is **turnout** or **attendance**. The **precinct** or the **polling station** (also called UIK according to the Russian acronym УИК) is atomic local unit, which is organized specially for elections. On average, a thousand of citizens can vote on a typical polling station and their total number is nearly 100,000. The precincts are not related to the administrative-territorial division of the country; they are organized and managed by the Central Election Commission (CEC) of the Russian Federation (CEC)³.

Theory

Practice shows that for elections without fraud, the distribution of voter turnout, as well as votes for candidates, in electoral precincts is close to a normal distribution; this is the "pure fair signal" (see Fig. 1). This is a critical assumption and its accurate test goes beyond the scope of this article. However, qualitatively and intuitively it can be assumed that elections should obey the law of large numbers. The vote of each citizen is an independent, random value with negligible effect on the final result. According to the central limit theorem, the mean of a large number of such random values should be normally distributed; all roads lead to a Gaussian. In the case of falsification of almost all types this distribution of votes is inevitably altered. In most cases the votes of "dead souls"

¹ Ver. 0.4.

² The name and affiliation are fictional, contact e-mail: oleg.kapustenko@gmail.com

³ <http://cikrf.ru/eng/>

are added for the Big Candidate (BC). Regardless of the mechanism of falsification this leads to the following two effects:

1. Fictitious increase in votes for the BC;
2. Fictitious increase in voter turnout.

It is crucial that these phenomena are **100% correlated with each other**, so that the effect is seen immediately in the statistical distributions of votes per precinct.

Simulation. D.Agafontsev made the first simulations to assess the rigging of the parliamentary elections in Russia, December 4, 2011 [3].

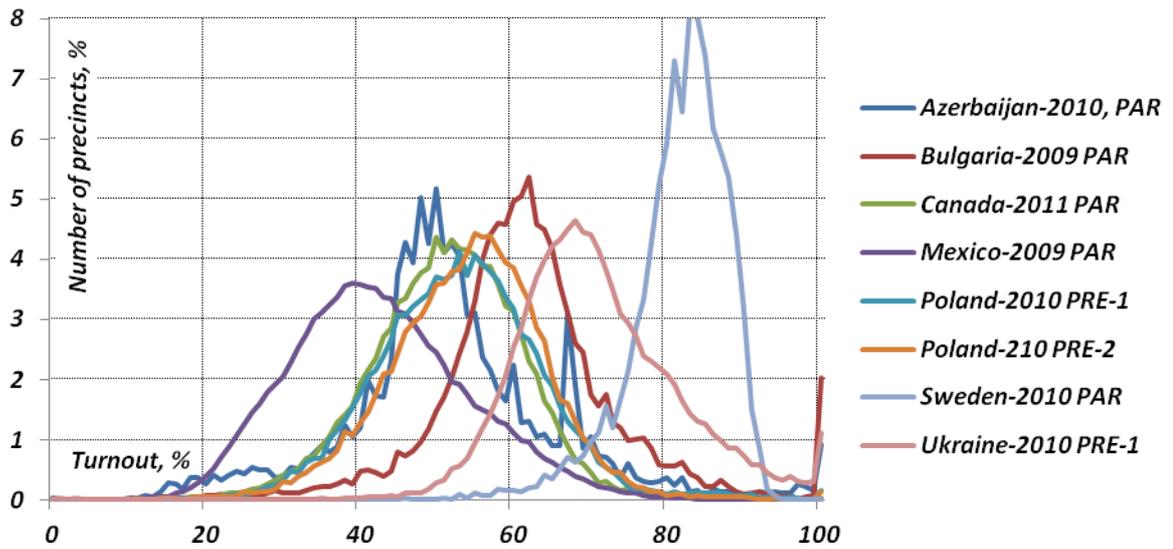


Figure 1. Fractional voter turnout per precinct in parliamentary (PAR) and presidential (PRE and stage) elections in different countries. Data from Sergey Shpilkin [9].

In this paper a simple model of election falsification is proposed. Let’s assume that elections take place at N polling stations with false votes added at a fraction fN of the polling stations, $f \in [0,1]$. A normal distribution of voters per polling station is expected:

$$V(W) = \exp [-0.5(W-W_{MEAN})^2/\sigma_w^2] \tag{1}$$

where W_{MEAN} and σ_w are the average voter turnout and its standard deviation respectively. In a particular polling station, the turnout is W' , from distribution (1). Assuming no theft of votes the only gap to assign additional false votes for the BC is in the range $[W', 1]$ with some false vote probability function:

$$W_{FALSE} = F(W) \quad W \in [W', 1]. \tag{2}$$

A set of simulations was carried out in order to demonstrate the effects of this model. The parameters used are (percentages are relative to the electorate in the precinct): normal turnout $W_{MEAN}=60\%$ and $\sigma=10\%$; fraud probability $f=0.5$; real support for the BC is 40% with a standard deviation 4% scaled from that of the turnout for simplicity. A total sample of $80,000$ precincts was simulated with a Monte Carlo technique. As a starting point, the turnout W' was randomly chosen according to (1) for each precinct. False votes were then added in the range $[W',1]$, with probability f , using different functions of (2). As can be seen from Fig. 2 (left), any injection of false votes spreads the votes from the normal peak to its right tail (a linearly descending “noisy generator” function (2) was used).

The probability function (2) is determined by the circumstances, i.e. the possible vote falsification mechanisms. It can be modest (flat or decreasing function), medium (growing), or even shamelessly large and aggressive (left exponential tail of 100%) or more realistically a combination of these (see Fig. 2, right). A priori, one can assume that the modest regime with a decreasing function (2) takes place in semi-democratic countries: the more falsification, the harder it is to implement under the pressure of public observers in particular and society in general. However, the specific form of the function can only be identified by analyzing real data, which is done in the next section.

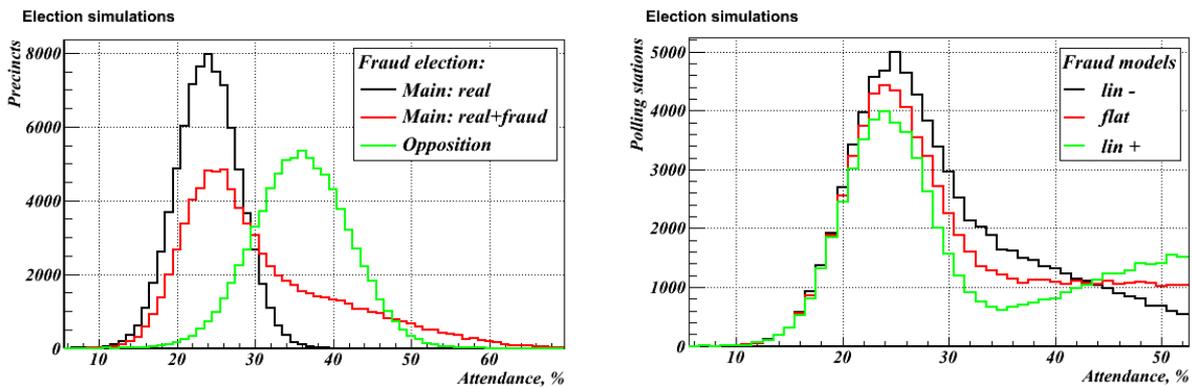


Figure 2. Left: simulated frequency distributions of votes per precinct (precinct) in elections without false votes (black and green histograms for the BC and all others, respectively) and with linearly decreasing injection of false votes (red curve). Right: the influence of weak (black), medium (red) and strong (green) falsification for the BC.

We also note an important fact: the higher the turnout, the less opportunity for falsification due to narrowing of the allowed $[W',1]$ gap! Thus, the tactic of boycotting the elections plays to the advantage of the BC.

The opposition candidate (OC). The violation of a normal distribution for the OC could take place by the stealing of votes, e.g. fake voting protocols in precincts. In this case a left tail should be observed for the OC, which is a mirror opposite to the right tail of the BC distribution in the case of a falsified election.

False votes in elections can be considered as background (noise), which should be isolated and cut off from the signal of fair votes.

Thus, summarizing the above, the simulation shows that injection of votes for the BC generates right tails in the distribution of votes, or voter turnout, per precinct, and stealing of votes for the OC gives the opposite effect. The sizes and shapes of the tails depend on the specific circumstances in the elections.

Analysis of real data

The results of the parliamentary elections in Russia, December 4, 2011, were analyzed. The data includes protocols from 95,073 precincts with 109,430,459 voters. One needs to emphasize that votes were normalized to the electorate, i.e. the list of eligible voters plus those eligible to vote outside their precincts, the so-called absentee ballots. Normalization to the number of voters introduces a systematic error in the case of false votes, which systematically shifts the total number of votes.

The distributions of votes per precinct for the state party "Edinaya Rossiya" (United Russia), here called the EdRo, and for all other parties (including invalid ones) are shown in Fig. 3. A significant right tail is observed for the EdRo as well as for the turnout, while the peak for the other parties is almost normally distributed.

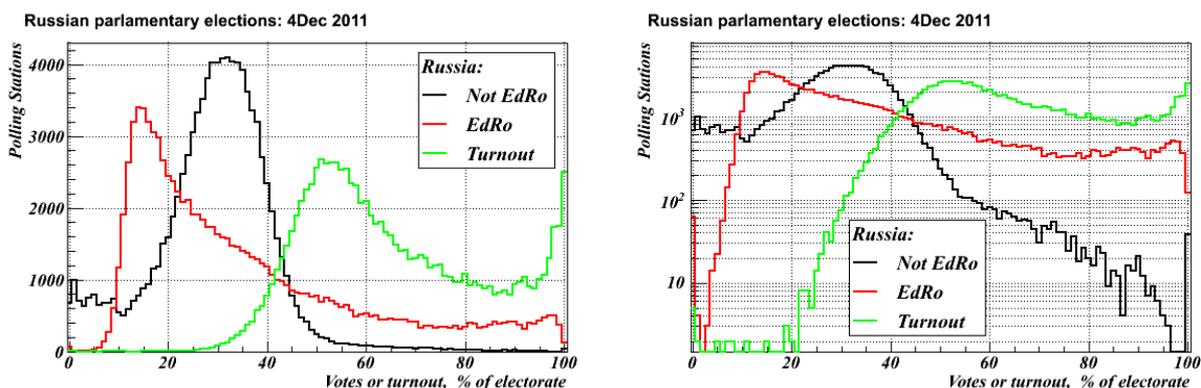


Figure 3. Votes for the EdRo (red) and the others (black) and the turnout (green) as fractions of the electorate in linear (left) and logarithmic (right) scales.

Fit of false tail

At this stage it is necessary to find a theoretical model that describes satisfactorily the right tail of the EdRo curve (red line in Fig. 3.). The exponential, Landau, and RooNovosibirsk[4] distributions give qualitatively close, but quantitatively poor fits. The best approximation is obtained using the Extreme Value Type II distribution, proposed in 1927 by French mathematician Maurice Fréchet [10]:

$$f(x) = \frac{a}{b} \left(\frac{b}{x}\right)^{a+1} \exp\left(-\left(\frac{b}{x}\right)^a\right) \quad a>0, b>0 \quad (3)$$

The result is shown in Fig. 4. As can be seen, the model fits the data quite well in the main middle range, while it's clear that some special mechanisms are contributing at higher turnout (>70%). The results of this fit allow us to separate the false votes modeled by the Frechet curve (purple line in Fig. 4) from the fair votes localized as a Gaussian peak (red line in Fig. 4).

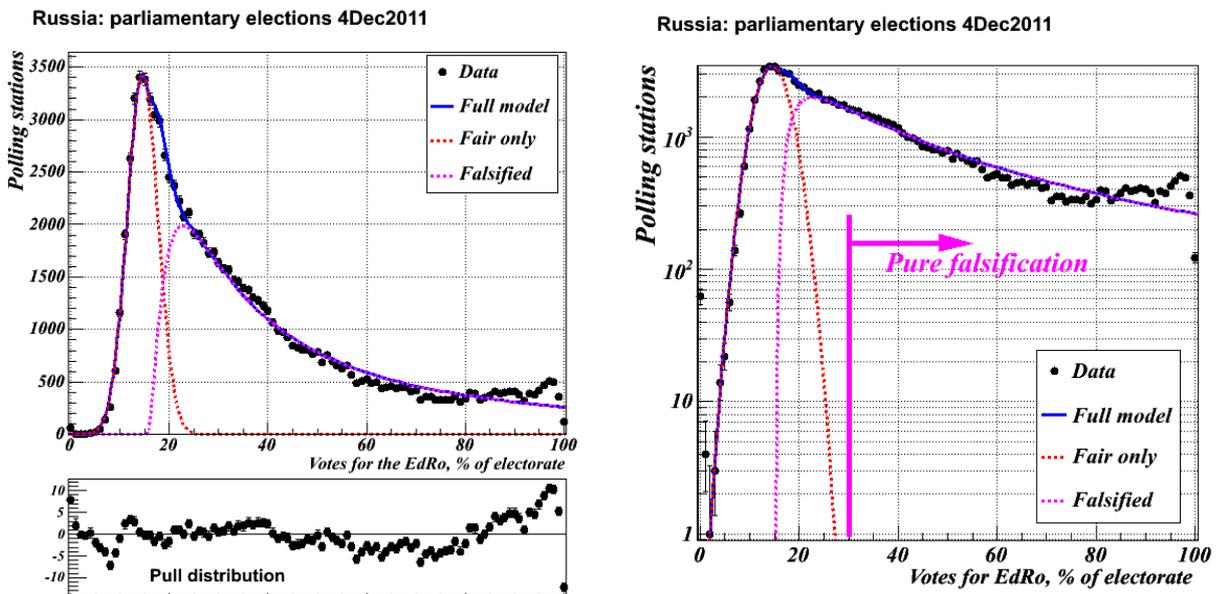


Figure 4. The fit of the election results for the EdRo in linear (left) and logarithmic (right) scales with the fair Gaussian signal (red dotted line) and falsified noise modeled by a Frechet curve (purple line). The full model (blue line) fits reasonably the data (black dots).

A recalculation of election votes based on the result of the fit is presented in the Table 1.

The fit on the logarithmic scale (Fig. 4 right) shows that the Gaussian curve of fair votes ends at a turnout of ~30%. This means that most precincts above this threshold are cheating ones, excluding a small number of some specific, unusual cases.

The fit result gives ~39% for the true percentage of votes for the EdRo, which is consistent to within a few percent of other estimates based on different approaches [5].

The opposition. The fit of the votes for the non-EdRo parties is shown in Fig.5. It was found that the “broadened” Gaussian signal on exponential pedestal right slope fits well the data except ranges at both edges of the scale. The non-EdRo peak is actually a sum of different Gaussians with different sigmas (votes for different parties), what’s why the parameter of broadness B was introduced to imitate the imperfectness of integral peak (Fig.5 zoom at left picture) and was fit to $B=1.2\%$. The fit gives $31.44 \pm 0.03\%$ of the electorate voting for the non-EdRo parties, using a real turnout of 50.2% , corresponding to 62.4% of the votes cast, which is consistent with the result $60,4\%$ based on the fit of the distribution of votes for the EdRo (see Table 1). The consistency of the results obtained in two different ways validates the approach used.

Table 1. Comparison of the election results with and without the "noise" in the model described.

Parameter	Value	Parameter	Value
Citizens, millions		The ratio of false votes, %	
All eligible voters (electorate)	109,4	Relative to all electorate	9,8%
False (bogus) votes	10,7	Relative to number who voted including false votes	16,3%
Number who voted		Precincts (UIKs)	
- Including false votes	65,7	Cheating UIKs, % of all UIKs	72,0%
- Excluding false votes	55,0	The average ratio of false votes for the EdRo, % of the electorate	22,5%
Number who voted for the EdRo			
- Including false votes	32,3	Votes for the EdRo, % of voted	
- Excluding false votes	21,6	Including false votes	49,3%
All other votes	33,2	Excluding false votes	39,4%
Turnout of the entire electorate, %		Votes for all others, % of voted	
- Including false votes	60,0%	Including false votes	50,6%
- Excluding false votes	50,2%	Excluding false votes	60,4%

Turnout dependence. The scatter plots of turnout vs. party votes (see Fig.6) are another good way to see the effects discussed. Two “comet tails” are clearly visible on the plots, but the tail of the EdRo comet is directed toward 100% of votes, while the other tail drops to zero votes. This signature is totally consistent with the assumption of injection and theft of votes.

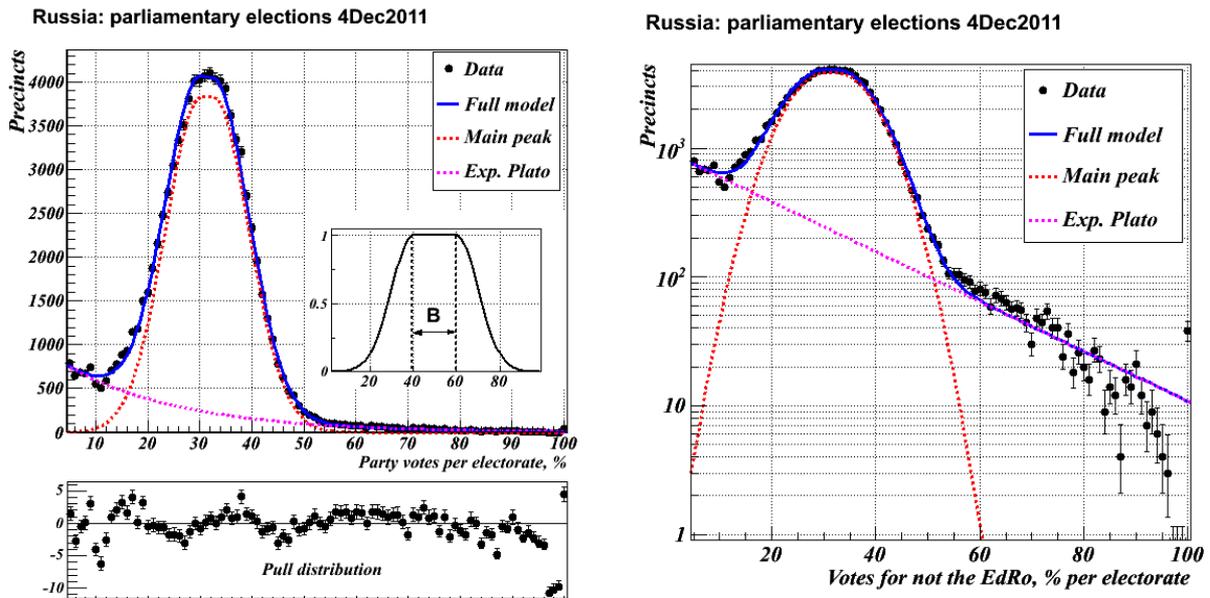


Figure 5. A fit of all other votes excluding those for the EdRo. The “broadened” Gaussian model used is shown in the zoom of the right picture, the parameter of broadness was fit to $B=1.2\%$.

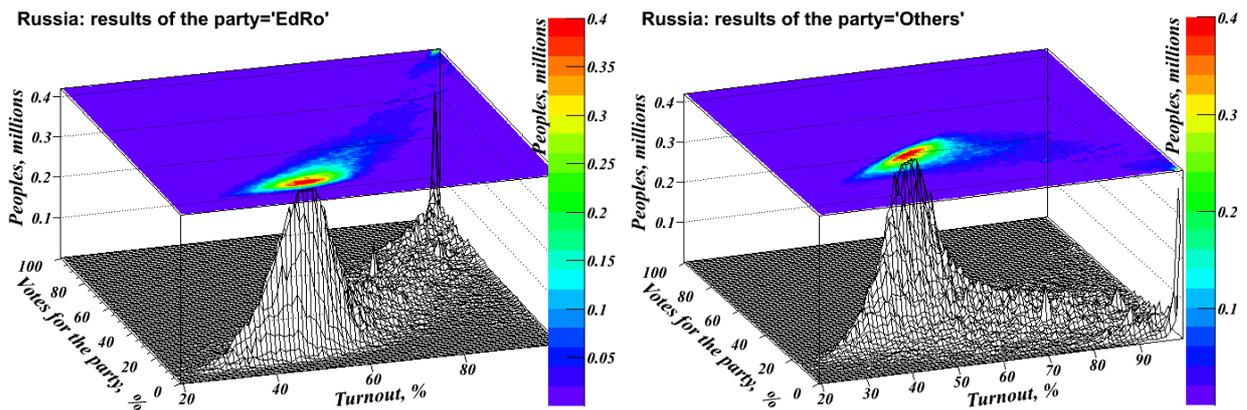


Figure 6. The turnout vs. party votes scatter plots for the EdRo (left) and all others (right). Shares of votes on the Y axis are relative to the electorate, which is also counted on the Z axis.

The evolution of voting distributions as the turnout increases is shown in Fig. 7. It is obvious that the shape of the voting distribution depends dramatically on the turnout. The more polling stations with a large turnout are taken into account, the stronger grows the right tail of the EdRo and the left tail of the all other votes. This fits perfectly with the assumption of falsified elections.

Fraud mechanisms. In Russia each citizen is assigned by residence location to vote at particular polling station. However, (s)he may vote on another precinct, previously asked for and received the so-called **absentee ballot** (2.5% of electorate at this election). In addition, (s)he can **vote in advance (ahead)** or outside the polling station on election day(1.5%). The last option is called **the voting at**

home, as it is usually provided for the sick and disabled persons to whom the special mobile ballot boxes are delivered from the assigned polling station (0.2%). These three options are potential channels of the direct injection of votes, so-called **ballot stuffing**, and subjects to study.

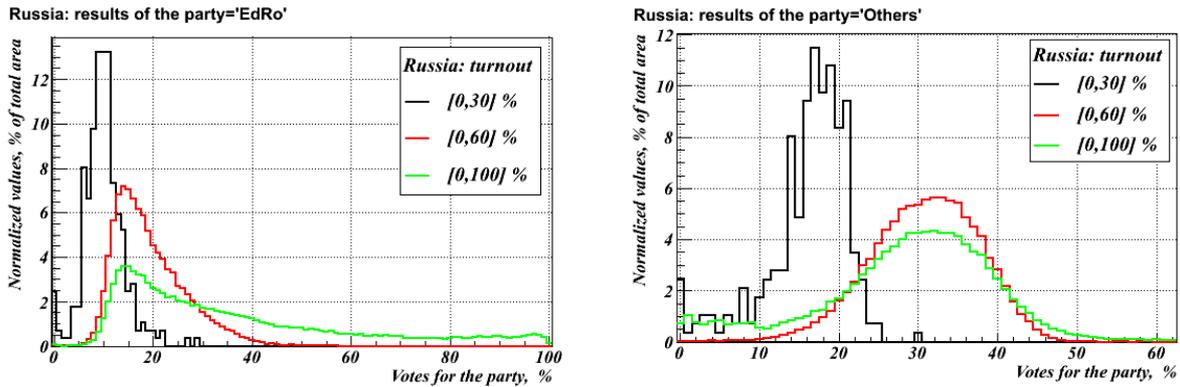


Figure 7. The voting distributions for the EdRo (left) and all others (right) for low (<30%, black), middle (30-60%, red) and high (>60%, green) turnouts. Vote fractions on X axis are relative to the electorate.

The distributions of votes for these categories are shown in Fig. 8. All these distributions have shapes close to the main trends shown in Fig. 3. The slight difference is that the mean value for all three kinds of votes for the EdRo is 36.1% of the electorate, which is 6.5% bigger than the final total result of 27% for the EdRo, which corresponds to the 0.3 million of votes. The other parties have 29.2% of these votes, which is 1.1% less than the final total value of 30.3% (the 0.05 million votes). Nevertheless, these differences are small in comparison with the 10.7 million false votes determined from the fit. Therefore, the ballot stuffing and/or the false reporting of the ballot results at precincts are the only ways to generate such a significant amount of false votes.

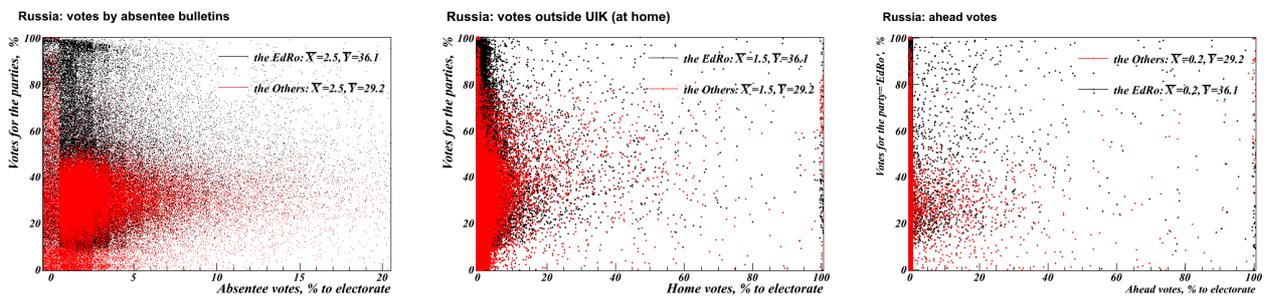


Figure 8. Distributions of votes from absentee ballots (left); home voters (middle), and advance votes for the EdRo (blue) and all others (red). Vote fractions on the Y axis are relative to the electorate.

The KOIBs. The systems of automated processing of ballots (KOIB, according to the Russian acronym КОИБ) are installed at some of the polling stations⁴ (see Fig.9, left). Only one bulletin at a time can be thrown into the KOIB, making it difficult to ballot stuffing. And the system scans the ballot inserted counting the final results automatically without human intervention. It means that both main falsification channels of the election are hampered at the polling stations equipped by the KOIB. We built the distributions of votes separately on polling stations equipped and not equipped with the KOIB. The total statistics covers 58 regions: 3189 KOIB-equipped polling stations with 5.75 million citizens entitled to vote⁵ versus 60,862 polling stations with 71.85 million of voters without the KOIBs. Although these groups differ in magnitude, the sample for automated polling stations covering 5.26% of the electorate is not negligible. The normalized shapes of distributions obtained are shown in Fig.10 (right).

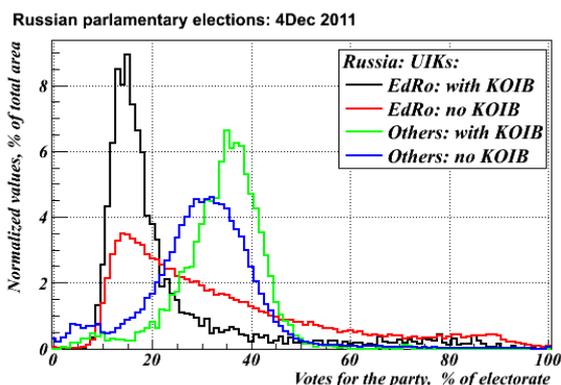


Figure 9. The KOIB, which is electronic ballot box⁶ (left) and the votes at polling equipped and not equipped with the KOIBs.

The shapes are dramatically different for the EdRo votes with a huge right tail for the polling stations without KOIBs perfectly consistent with the assumption of ballot stuffing. The shapes for the other parties also differ significantly: main right shift and right tail has appeared for the polling station without KOIBs, which could be explained by the theft of the votes.

Regional effects. Although further detailed analysis of the effects discussed here is potentially possible, probing one step down to regional level, it is far beyond the scope of this work. Here we will give one example which, nevertheless, indicates huge regional differences in the results of the elections. The results of voting in Moscow are shown in Fig. 10 (left). We assume that the second

⁴ E.g. <http://mysouth.su/2011/07/counting-the-votes-at-the-elections-in-primorye-will-be-automatic/>

⁵ Information about KOIB is from Sergey Shpilkin <http://podmoskovnik.livejournal.com/125484.html>.

⁶ Photo is from <http://rugrad.eu/communication/blogs/VadimKosuhin/cohiba-sleight-of-hand-and-no-cheating>.

peak of the EdRo distribution (observed also by others, e.g. [6]) cannot be explained by anything other than execution of the plan, publicly claimed by the EdRo before the election, to obtain more than half the votes. A fit shows (see Fig. 10, right) that authorities at almost one third of precincts (~1100 from a total of 3373) followed this order. Other detailed effects can also be investigated, e.g. Peaks around “nice numbers” of voting results (60%, 75%, 80%, etc.) have also been observed (see Fig. 11, right). Finally, significant differences are revealed between different groups of citizens with the most significant gap between urban and rural votes. For details we refer the reader to numerous publications [1].

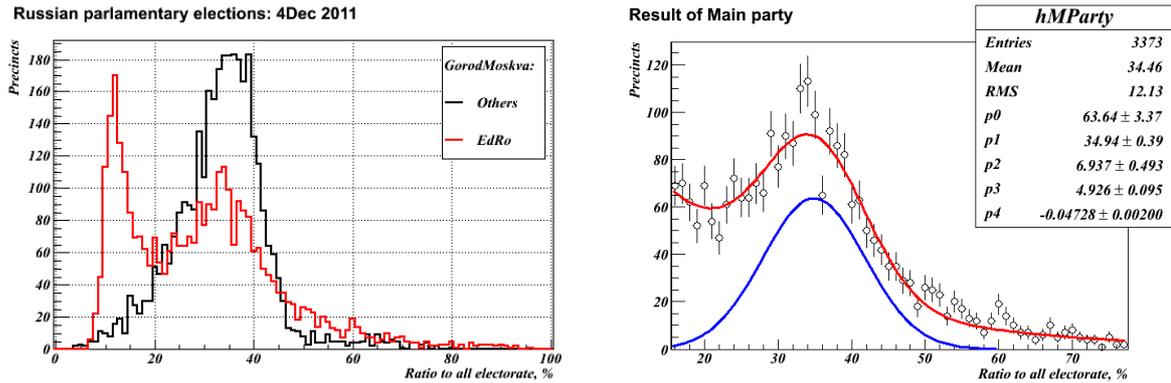


Figure 10. The Moscow votes (left) and fit of the second peak in the EdRo distribution (right).

The distribution of the tail. It is an interesting fact that the tail of false votes is described by the Frechet, extreme value, distribution (3) with the following parameters:

$$a = 0.93 \pm 0.02, b = 20.4 \pm 0.3 \tag{4}$$

As the name implies, the distribution is intended to describe the behavior of the maximum value of a random variable, for example, the maximum amount of rainfall in one day for a year [7]. If a set of random numbers is generated, the maximal number in the sample follows a Frechet distribution. The shape of the distribution depends crucially on the size N of the sample. For $N = 1$, the Frechet distribution is identical to the distribution of the random variable itself. With increasing sample size the maximal value is more and more likely to be at the upper boundary of the allowed range. In our case the maximum value is 100% voter turnout. The injection of false votes can be treated as virtual sampling of false votes introduced by precinct authorities. The size of this sample is limited by the amount of external pressure applied to the precincts by public control over their work (honest members of precincts; electoral observers, etc.). Thus, it appears that the Frechet distribution has found another use in real processes, in the analysis of falsified elections.

Finally, the scientific questions discussed have already moved to political posters (see Fig. 10).



Figure 10. Protest against unfair elections, 10 December 2011, Moscow [8-9]. Phrases in the posters are “For the normal distribution” (left) and “In Gauss we trust!”

Conclusion

The method employed in this article demonstrates the power of statistical analysis applied to election results from individual precincts. It is shown that the injection and theft of votes distorts their distributions. Despite fraud violating the election results, a statistical analysis is able to determine and cut off the tails of fraudulent votes and, ultimately, to separate the wheat from the chaff or flies from the cutlets (the Russian proverb with the same meaning).

The analysis shows that ~10 million of false votes have been added for the EdRo, which accounted for ~10% increase to the final results. Fraud took place at ~70% of the polling stations with average value of added false votes of 20% of the electorate.

The method is confirmed both theoretically (simulation) and experimentally. The model fits well the data, and the noise is separated from the signal with a high level of statistical accuracy (error less than 1%). Interestingly the fake votes obey the extreme value type II distribution of Frechet, showing that this distribution is suitable for the analysis of “noisy” elections.

Thus, access to voting data from individual precincts makes faking results in such a way as to be invisible in statistical analyses almost impossible; full control of all precincts would be required to achieve this, and this is not the case in normal elections. Alas, a more realistic way to hide election fraud is to deny access to detailed voting data and to declare only the final results.

The data used and scripts developed to obtained the result presented here can download from the Google docs [10].

Acknowledgments

Thanks to Mike, who suggested the use of extreme value distributions and to Julia for the correction of English⁷. Thanks to all those who have discussed the subject in social networks [1,2] stimulating the idea for this work. Also many thanks to the developer of ROOT (root.cern.ch) in general, and the package RooFit (<http://roofit.sourceforge.net>) in particular, without whom this analysis would have been impossible.

References

1. <http://podmoskovnik.livejournal.com/65386.html> # cutid1;
<http://www.echo.msk.ru/blog/echomsk/631692-echo>;
<http://trv.nauchnik.ru/40N.pdf>. <http://trv-science.ru/2011/12/06/vybory-2011>;
http://samarcandanalytics.com/?page_id=39, in English; <http://www.lif.univ-mrs.fr/~ashen/elections.pdf>; <http://trv-science.ru/94N.pdf>;
http://www.gazeta.ru/science/2011/12/10_a_3922390.shtml. In Russian if other is not pointed.
2. <http://oude-rus.livejournal.com/>, <http://podmoskovnik.livejournal.com/>;
<http://kobak.livejournal.com/>, see various internal references in the blogs, in Russian.
3. <https://plus.google.com/100208873201558961542/posts/Xp8okPaMTdv>, in Russian.
4. <http://root.cern.ch/root/html512/src/RooNovosibirsk.cxx.html>
5. <http://podmoskovnik.livejournal.com/129843.html>, in Russian.
6. <http://oude-rus.livejournal.com/540063.html>, in Russian.
7. http://en.wikipedia.org/wiki/Fr%C3%A9chet_distribution.
8. http://pics.livejournal.com/shultz_flory/pic/000ec4e6, in Russian.
9. <http://podmoskovnik.livejournal.com/133805.html>, in Russian.
10. http://docentes.deio.fc.ul.pt/fragaalves/FRAGA_ALVES_LEXICON.pdf.
11. <https://docs.google.com/open?id=0Bz-pwqutBJonZmJmYWQzYjEtMTM5Yy00OGE0LTk0MWEtZGUyNTBmMzdINmM0>

⁷ As the text has been updated after the correction, any mistake found is, most likely, my fault; I apologize for this.