Towards Detecting and Measuring Ballot Stuffing*

Dmitriy Vorobyev†

July 2013

Abstract

This paper proposes a method for detecting electoral fraud in the form of ballot stuffing. As ballot stuffing increases both turnout and the incumbent's vote share in precincts where it occurs, precincts with low reported turnout are more likely to be clean. Information on clean precincts is used to simulate counterfactual data for infected precincts, which are then compared to the observed data. The method is applied to the 2006 Finnish presidential elections. The test fails to reject the hypothesis of no ballot stuffing for the original presumably clean data, but detects artificially imputed fraud and provides a correct estimate of its magnitude. The same method implies that in the presidential elections in Russia held between 2000 and 2012 ballot stuffing was a significant issue, and the number of ballots stuffed in favor of the incumbents had been persistently growing over the period. Regional-level analysis suggests that this is a result of both increasing fraud magnitude and expanding of electoral falsification across the regions of Russia.

JEL Classification: D72, D73
Keywords: Elections, Fraud Detection

*This Research was supported by a grant RRC X-96 from CERGE-EI Foundation under a program of Global Development Network. All opinions expressed are those of the author and have not been endorsed by CERGE-EI or the GDN.

†Center for Economic Research and Graduate Education - Economic Institutes, a joint workplace of Charles University in Prague and The Economics Institute of The Academy of Sciences of the Czech Republic. Address: CERGE-EI, P.O. Box 882, Politickych veznu 7, Prague 1, 111 21, Czech Republic.

1 Introduction

Despite its importance, electoral fraud suffers from a relative lack of attention in the academic literature. Probably the main reason for this is the absence of a reliable measure of fraud. Indeed, not only measuring but even detecting fraud is problematic. The existing methods of fraud detection are more qualitative than quantitative, often based on the subjective assessment of electoral transparency and fairness by observers or other participants of the electoral process, and the results they produce may not always be treated as fully reliable. The few attempts to rigorously analyze electoral data for the presence of fraud have usually required a large amount of data, which handicaps efforts to measure fraud, proxy it, or even detect it with some confidence. It further precludes implementing reliable empirical research, which in turn discourages efforts towards a theoretical study of the nature and consequences of electoral fraud.

This paper proposes a statistical mechanism for testing the fairness of elections when the available data are limited. The methodology enables elections to be tested for the presence of electoral fraud in the form of ballot stuffing using official detailed electoral data. The mechanism is applied to test the fairness of the Russian presidential elections between 2000 and 2012, whose transparency and integrity are often put in doubt, and obtain an estimate of the magnitude of ballot stuffing in Russia.

The paper is organized as follows. The next section discusses existing approaches of detecting fraud. Section 3 presents a methodology that enables testing the fairness of elections based on official electoral results. In Section 4 the methodology is applied to several datasets. First, I create artificial electoral data, show that the test fails to reject the null hypothesis of no ballot stuffing, then impute fraud of about 2% and test the data once again, resulting in a strong rejection of the null hypothesis. Second, I perform the same exercise for data on the 2006 presidential elections in Finland. The test cannot reject the hypothesis of fair elections for the original data but rejects the hypothesis, once 2% fraud is imputed. Third, I apply the test to data on the Russian presidential elections 2000, 2004, 2008 and 2012 on both country and regional levels. The hypothesis of no ballot stuffing is strongly rejected in all the cases, and the test implies that the number of stolen votes in favor of incumbents had been persistently growing between 2000 and 2012.

2 Literature Review

Detecting fraud in data is not new. The general idea underpinning most fraud detecting statistical techniques is tracing unusual patterns in the observed data that might be explained by fraud. Such techniques have been successfully used for uncovering fraud in a variety of domains, from sports betting (e.g., Wolfers, 2006)) and education (e.g., Jacob and Levitt, 2003) to online auctions (e.g., Pandit et al., 2007) and banking (e.g., Quah and Sriganesh, 2008). A number of recent papers review fraud detecting techniques for specific fields such as telecommunications (Becker et al., 2010), health care (Li et al., 2008) and finance (Sudjianto et al., 2010). Though the specific design of fraud detecting techniques does depend on the nature of the data and type of expected fraud, all fraud detecting methods share enough features to be divided into two main groups: supervised and unsupervised. Methods of the supervised type assume that there are two data samples available for the analysis: the one which is affected by fraud, and the one which is not. In this case, labeling a new data set as clean or fraudulent is essentially a comparison with benchmark samples. When such samples are not available, the unsupervised methods are applied. They do not use benchmark samples and instead look for outliers in an observable sample. Due to the nature of data on elections and frequently changing electoral environments, electoral fraud detecting methods have to be of an unsupervised type.

Attempts to detect fraud in the electoral process used to be rare and unsystematic. Lehoucq (2003) in his comprehensive review of studies on electoral fraud mentions a number of papers that look for traces of electoral fraud in elections in Argentina, Peru, Colombia, England, Ireland, Germany, Spain, Mexico and some Asian countries[1]. The majority of these studies detects fraud using descriptive evidence such as surveys, interviews and documents; none uses statistical methods. Even though such qualitative approaches can provide insight into the presence of electoral fraud in given elections, they require tremendous effort to collect relevant data and may yield results with limited application and replicability.

Due to limitations of qualitative approaches researchers have started to pay attention to the statistical analysis of electoral data with the aim of detecting electoral fraud. The largest and most rapidly growing approach to electoral fraud detection is digit analysis, which analyzes digit patterns in electoral data to identify anomalies that may appear due to fraud.

Beber and Scacco (2008) suggest a methodology based on the idea that people are bad random number generators: if elections are fair, the distribution of insignificant digits (e.g., digits at the third decimal place and further) in electoral outcomes (i.e. data on turnouts and vote shares) must be close to uniform, but if there are manual changes in outcomes there must be biases in generating digits. The idea is supported by a statistical comparison of outcomes from Swedish and Nigerian elections. However, such a method is limited to detecting manipulations with electoral returns; it is unlikely produce a result if electoral outcomes are shaped in a more sophisticated way than manually changing digits in election protocols; it does not provide any estimate of fraud magnitude.

In contrast to Beber and Scacco (2008), a number of recent papers have analyzed the first

[1]For the references see Lehoucq (2003).

significant digits in official electoral data to find deviations from Benford's Law (Benford, 1938). The law states that the first digits in real data are distributed in a specific non-uniform way. Deviations from the law are found by Roukema (2009) in data on the last Iranian presidential elections, Cantu and Saiegh (2010) in Argentinean elections, Pericchi and Torres (2004) in Venezuela, Mebane (2006, 2010) in the US and Mexico, Mebane and Kalinin (2009) in Russia, and by Breunig and Goerres (2011) in the Bundestag elections in Germany. Despite their merits, these methods of digit analysis are subject to criticism which puts in doubt their relevance for detecting fraud in electoral data (e.g., Brady, 2005). Recently Deckert et al. (2011) have shown that deviation from Benford's Law can arise in electoral data regardless of whether the elections are rigged or fair, and that the methods essentially do not differ from a random draw in their ability to mark elections as clean or fraudulent.

A number of authors have suggested alternative methods for discovering fraud in official electoral data. Myagkov and Ordeshook (2008) study the fairness of Russian federal elections between 1993-2007, by examining a variety of patterns in Russian electoral data such as turnout distributions across regions and precincts, and vote flows between different elections. They conclude that ballot stuffing as well as other fraud techniques of the 1990s were frequently used in a few Russian ethnic republics but then spread to other regions of the country. They apply similar techniques to several elections in Russia and Ukraine looking for the presence of fraud (see Myagkov et al, 2009, for a detailed discussion).

Electoral fraud in Russia is also discussed by Treisman (2009), who reviews the trends in voting in Russia since 1991. In a chapter devoted to electoral manipulations and fraud, by studying a variety of Russian electoral statistics in a way similar to Myagkov and Ordeshook (2008) and Mebane and Kalinin (2009), the author finds that in the early 1990s, the elections in Russia were nearly clean, whereas since 2000 electoral irregularities have become an integral part of electoral competition. Although Treisman's approaches are reasonable and capable of producing reliable conclusions, they are mainly based on a visual analysis and comparison of electoral data.

More recently Levin et al (2009) have elaborated on the approaches by Myagkov and Ordeshook (2008) and explore electoral data in Venezuela for the presence of fraud. They analyze the data on two consecutive state-level referenda in 2007 and 2009 and, assuming time-constant voters' preferences, discover unusual patterns in the voting behavior of selected regions that mainly benefit the incumbent. Specifically, most of the new votes in favor of Chavez in the 2009 referendum came from the regions with large abstention in 2007. To obtain this result Levin et al. explore three types of indicators. First, they perform digit analysis of the electoral outcomes. Second, they study the flow of votes between the two elections by estimating the proportion of vote share in the first elections that "flows" to each alternative in the second elections in order to see whether there is a noticeable increase in support of one of the alternatives in regions with a substantial increase in turnout. Finally, they look closely at the relationship between the time changes in turnout and share of votes cast for the incumbent. As in Myagkov and Ordeshook (2008), Levin et al's analysis is primarily based on a comprehensive investigation and description of data patterns rather than statistical testing.

In short, even though electoral fraud appears to be very widespread, existing means of detecting fraud are primarily descriptive and qualitative and highly dependent on the nature of available data. Though statistical studies of electoral data with focus on fraud do exist, they are mainly focused on exploring unusual patterns in data and require a tremendous amount of information. Thus, there is still a need for a rigorous method to detect fraud and measure its extent, especially for cases when available data are restricted. This paper attempts to make a progress towards designing such a method.

3 Fraud Detecting Methodology

The suggested approach is based on the observation that ballot stuffing increases both turnout and votes cast for the corrupt candidate (hereafter I assume that fraud is implemented in favor of the incumbent). If elections are subject to ballot stuffing, which takes place not in all but in a selected number of precincts, this observation immediately implies that a precinct with lower reported turnout is more likely to be clean. The general idea behind the methodology described below is to use the information from such low turnout precincts to simulate counterfactual data for high turnout and likely fraudulent precincts, and test for systematic difference between counterfactual and observed data.

The suggested procedure allows for testing for the presence of ballot stuffing even when very limited data are available. Suppose that the only data available for the analysis are precinct level turnouts and candidates' vote shares. If elections are not fraudulent, and there is a certain degree of homogeneity between electoral precincts, the distribution of turnout across precincts should be close to bell-shaped. Clearly, if electoral districts are similar in terms of characteristics that might determine turnout, or alternatively if turnout is weakly affected by characteristics in which the districts differ, then distribution of turnout should be approximately normal (see, for example, Myagkov and Ordeshook (2008) or Levin et al. (2009) for more detailed discussion).

In turn, ballot stuffing, when takes place in a given precinct, results in an increase both in reported turnout and in the incumbent's vote share. Consequently, such fraudulent precinct moves in turnout distribution towards its right tail. As a result distribution of turnout in fraudulent elections will be skewed to the left and have thicker right tail. Furthermore, in precinct in the right distribution tail, i.e. those with high reported turnout, incumbent will have an advantage in comparison to the other precincts. The idea behind the suggested methodology is to check whether incumbent has such an advantage and whether it could be considered as natural.

Suppose the following statistic is computed from the available data:

$$s = \frac{V_I/V_R|t \geq t^*}{V_I/V_R|t < t^*}. \tag{1}$$

The nominator of the statistic is the ratio of the incumbent's and the runner-up's shares in the right tail of the turnout distribution (i.e., in those precincts where the turnout is above some threshold t^*). The denominator is the same ratio, but computed over the left distribution tail. I use this statistic to test the null hypothesis that elections are fair. Under the null hypothesis the statistic should be close to one if there is no objective systematic relationship between turnout and voting in favor of one or another candidate, meaning that if elections are fair, the ratio of the incumbent's and the runner-up's shares should not systematically differ in the precincts with high and low turnout. But if there is ballot stuffing in favor of the incumbent, the incumbent's share in the precincts with high turnout will be relatively higher, meaning that the statistic will be greater than one (less than one, if ballot stuffing is in favor of the challenger).

Indeed, there could be an objective correlation between the turnout and the vote shares for the candidates if the supporters of one candidate are more politically active than the supporters of the others. In this case this correlation will be present over the whole dataset, including the left tail. In other words, if such natural relationship between turnout and vote shares exists, it can be estimated using left tail clean data only, and fraud, if it exists, will make this correlation higher in the right tail. The procedure described below is designed to test not for the presence of correlation between turnout and incumbent's vote share in the right tail data, but rather for the presence of extra correlation in comparison to the left tail.

Using the statistic, I test the null hypothesis that there is no ballot stuffing. To conduct the test, one needs to know only a distribution of the statistic under H0. To obtain such distribution the following procedure is proposed.

First, recall that ballot box stuffing increases both turnout and the incumbent's share of votes. This means that a fraudulent precinct moves to the right tail of the turnout distribution. This in turn means that in the case of rigged elections, the left tail of the observed turnout distribution contains a fewer number of fraudulent precincts than the right one. Moreover, the higher the scale of fraud, the further to the right a fraudulent precinct moves, implying that a larger tail of the distribution stays clean. Second, recall that if there is no ballot stuffing and precincts are in some sense homogenous, turnout distribution across precincts should be approximately normal. Assuming some particular shape of the true turnout distribution (for example, normal), I choose its parameters such that the distribution fits the left-tail data, i.e. those with turnout below some threshold value t^* (I discuss the choice of the threshold below). Next, I estimate the relationships between turnout and vote shares in the clean left tail by simply regressing vote shares on turnout:

$$V_{Ii} = \alpha + \beta t_i + \epsilon_i. \tag{2}$$

Note that the purpose of these regressions is not to establish a causal effect of turnout on vote shares, but rather to find a correlation and then extrapolate it on the simulated right tail.

Then I repeat the following simulation multiple times. At each simulation step I generate a new turnout distribution across precincts $\bar{t_i}$ as a random draw from the fitted normal distribution and predict vote shares. To make predicted vote shares consistent with clean left tail data I, first, maintain the same relationships between the vote shares and turnout as in the observed clean left tail, and, second, introduce additional noise into predicted vote shares such that their variances evaluated over the left tail are the same as the variances of observed left tail vote shares. Specifically, vote shares for the incumbent and challenger (runner-up) are predicted as

$$\bar{V_{Ii}} = \tilde{\alpha} + \tilde{\beta}\bar{t_i} + u_i. \tag{3}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are random draws from normal distributions with means α and β, and standard deviations equal to corresponding standard errors from regression 2. The latter means that when predicting the vote share, I do not just use coefficients obtained from regression 2, but allow them to vary across simulations according to the precision of the estimation. If a predicted vote share exceeds 1, it is equalized to 1. Errors u_i are drawn from a zero mean normal distribution. Variance of this distribution is chosen such that

$$Var(\bar{V_{Ii}})|t < t^* = Var(V_{Ii})|t < t^*. \tag{4}$$

By allowing coefficients to vary and making variance of predicted vote shares to be the same as variance of actual vote shares, I guarantee that the simulated right tail data are consistent with observed relatively clean left tail data.

Once the vote shares are predicted, statistic (1) can be computed. Repeating this simulation multiple times and computing the statistic on each step, one can obtain a distribution of the statistic under the null hypothesis that elections are fair, and tabulate critical values. Computing the statistic for observed data one can now test the hypothesis. Note that if the value of the statistic appears to be above the right tail critical value, it implies that ballot stuffing took place in favor of the incumbent. In contrast, a statistic below the left tail critical value signals ballot stuffing in favor of the challenger. Given critical and actual values of the statistic it is now easy to obtain an estimate of ballot stuffing magnitude by calculating the vote share that incumbent has to obtain in the right tail precincts in order to equalize the observed value of the statistic to its critical value of desired confidence level. The difference between actual vote share and counterfactual vote share calculated in this way would give an estimate of ballot stuffing magnitude.

When there are more than two candidates in elections the procedure is slightly different. Because turnout can be related differently to vote shares of each candidate and there are several candidates, challenger's vote share cannot be predicted by simply subtracting incumbent's vote share from one. Instead, his vote share should be obtained in a similar way that incumbent's one (formulas (2) and (3)).

The ballot stuffing detection procedure is based on a number of implicit assumptions. First, I assume that ballot stuffing occurs in a small number of precincts. Suppose instead that ballot stuffing of relatively the same magnitude would occur in all precincts. This means that turnout and share of votes cast for the incumbent increase in all precincts. Thus, there will be no systematic difference between left and right tail data due to fraud, which is needed for identification. Second, the fraud should be of reasonable magnitude in a sense that it should result in a noticeable increase in turnout to move the precinct to the right tail of the turnout distribution. Together these two assumptions say that fraud, in order to be detected should move the precinct where it occurred to the right tail of the distribution.

The methodology described above explicitly distinguishes between the left tail and right tails of turnout distribution by using a turnout threshold t^*. Ideally, t^* should be chosen such that all precincts with turnout below t^* are clean and the lowest reported turnout among the fraudulent precinct is slightly above t^*. In practice such choice is challenging. More likely, there will still be some fraudulent precincts even in the left turnout distribution tail but less than in the right one, meaning that fraud detection is still possible though the fraud magnitude will be underestimated in this case. On the one hand, the low value of the threshold allows for the capture of small-scale fraud and fraud in low turnout precincts since they are more likely to appear above the threshold. On the other hand, low t^* will not allow detecting even large-scale fraud if it appears in a very small number of precincts, as the contribution of the fraudulent precincts in the statistic will be relatively small due to a large number of clean precincts. Also, low t^* will result in a small number of data points in the left tail which are used for estimation of the natural relationship between turnout and vote shares which will decrease the power of the test. On the other hand, a high threshold value would make it easier to reject the null hypothesis of no ballot stuffing if there is large-scale fraud, but could fail to reject the hypothesis when fraud is balanced. Thus, the choice of threshold generally depends on the data as well as some priors about the nature and the extent of fraud.

One way to endogenize the choice of the turnout threshold is to analyze the values of coefficient on turnout from regression (2) for different thresholds. Clearly, going over different threshold values from low to high, at some point the left tail data would start containing fraudulent precincts. As a result coefficient β from regression (2) will start growing if fraud is in favor of incumbent and decreasing if fraud is in favor of challenger. Thus, value of turnout threshold at which coefficient of turnout starts growing (decreasing) would be a natural choice for t^*. The value of t^* specified in this way can itself signal about the nature of fraud. If t^* appears to be high that would mean that in each fraudulent region magnitude of ballot stuffing was huge as substantial share of spoiled precincts ended up in the very right tail of the reported turnout distribution. Alternatively, reasonably low t^* means that fraud in a given precinct was not extreme though total number of fraudulent regions could still be substantial. Indeed, there could be the cases when sharp changes in the value of β are not observed at all (for instance when elections are clean). In this case the only way to define t^* is to make some reasonable, yet ad-hoc choice, for instance, some number between 0.5 and 0.8.

It is important to notice that particular choice of threshold value can affect the corresponding estimate of ballot stuffing magnitude. Recall, that the estimate described above is the difference between actual number of votes for the incumbent and the number votes that incumbent should have received in order just not to reject the hypothesis of no ballot stuffing at the desired confidence level. Since higher value of threshold effectively means that higher fraction of data is considered as clean, and thus lower share of data is considered as potentially fraudulent, the fraud estimate will generally be a decreasing function of threshold. Thus, this estimate should be though of as a lower bound of ballot stuffing rather than its measure. Though generally direct comparison of such estimates across different elections would not be entirely correct, under certain circumstances it might be still useful for getting an idea about relative extent of ballot stuffing. See Section 5 for further discussion.

Noticeable growth or decline of coefficient β starting from some turnout threshold value itself signals about fraud as in the absence of ballot stuffing the coefficient should not sharply change. Thus, testing for the broken trend in β as a function of t^* could be another and probably simpler way to test for the presence of ballot stuffing. Alternatively, because the suggested methodology is based on the observation that turnout in clean elections should follow approximately a normal distribution, one can simply test for the distribution symmetry. However, the suggested methodology has a number of advantages over these alternatives. First, the method would indicate the direction of ballot stuffing if it exists. Depending on whether observed value of the statistic falls to the left or right tail of its distribution under the null hypothesis, one can always say whether ballot stuffing is in favor of the incumbent or challenger. Second and most important, in contrast to the symmetry and broken trend tests, the suggested method provides some information on ballot stuffing magnitude.

Finally, it is important to note that ballot stuffing is just one of the technologies for rigging elections, while the whole range of the techniques is wide (Lehoucq, 2003). As a result, the suggested methodology tests for the presence and provide and an estimate of just a particular fraud activity, which however is very widespread and popular, and which accounts for substantial share of voting fraud due to its obvious cost effectiveness. Moreover, technically the suggested methodology is intended to detect any activity that leads to simultaneous increase in reported turnout and vote share of a candidate, and ballot stuffing indeed, is not the only rigging technique that leads to this. Such activities as multiple voting and vote buying also result in increase of turnout and incumbent's vote share in a regions where they occur, and thus the suggested methodology is fully appropriate for tracing out their steps in electoral data.

4 Testing Fairness of Elections

In this section I apply the described methodology of detecting ballot stuffing to several distinct datasets. I first generate artificial clean electoral data and then impute fraud into them. I apply the test to original clean data and then to the fraudulent data to show that the test raises a red flag in case of fraudulent data only. Then, I apply the methodology to real data which came from the 2006 Finnish presidential elections. As the integrity of these elections was never subject to debate, I consider them as an example of presumably clean elections and show that the test cannot reject a null hypothesis of no ballot stuffing for this data, but it does reject it once fraud is artificially imputed.

4.1 Artificial Data

First, I show that the method is able to detect electoral fraud of a reasonable magnitude in artificial data. For this purpose I create a dataset that consists of turnout and candidates' vote shares. Specifically, I generate 1000 observations for turnout t that follow a normal distribution with 0.5 mean and 0.1 standard deviation. Each observation represents data for a precinct. Then I generate vote shares for the incumbent allowing for a natural correlation between vote shares and turnout as well as noise drawn from normal distribution $N(0, 0.05)$:

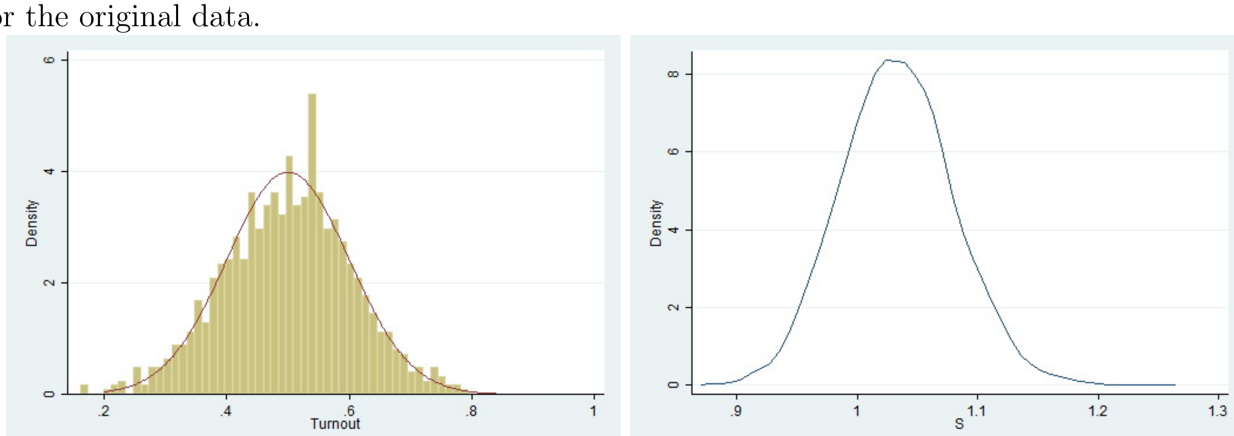
$$V_{Ii} = 0.05t_i + \epsilon_i. \tag{5}$$

I then apply the methodology described above to the simulated data. First, I need to choose a threshold value of turnout to define left and right tails. There is no clear trend break in coefficient on turnout from regression (1) as a function of turnout percentile where threshold value is evaluated, and the variation in the coefficient is not substantial (See Figure 2). I choose threshold value to be at the 61st percentile, which implies that the left tail contains precincts with turnout below 0.556.

Then I estimate the relationship between the incumbent's vote share in the left tail of the turnout distribution (i.e., in precincts where turnout is less than 0.556) by running a regression of V_I on t. Next, I choose the parameters of the normal distribution such that it fits the left tail of the observed turnout distribution as precisely as possible. Having the regression coefficients, their standard errors and turnout distribution parameters, I predict incumbent vote share in the right tail of the turnout distribution (i.e., in precincts where $t \geq 0.556$) allowing for variation in coefficients (coefficients for prediction are randomly drawn from distributions consistent with the estimated coefficient means and standard errors) and noise. Noise is added in a way that variances of incumbent's predicted and original vote shares are the same for precincts with $t < 0.556$.

Once the right tail data are constructed, the test statistic is calculated. Then I repeat the procedure of the right tail prediction 5000 times, calculate the statistic on each simulation, get the distribution of the statistic under the null hypothesis of no ballot stuffing, tabulate critical values and compare them to the value of the statistic from the observed data. The value of the statistic from the observed data is 1.059, while the 90% critical value is 1.093 and the 10% critical value is 0.973. Thus, the test cannot reject the null hypothesis.

Figure 1: True turnout distribution and distribution of the statistic under the null hypothesis for the original data.



Then, I impute fraud in the data. I randomly choose 150 precincts (15%). In each of them I give the incumbent additional votes: in every spoiled region, the incumbent receives additional number of votes f_i proportional to the size of the district E_i. Then, if we denote $\theta = f_i/E_i$ where E is the size of the district, after fraud turnout $\hat{t_i}$ and incumbent's vote share $\hat{V_{Ii}}$ can be expressed in terms of before fraud turnout t_i and the vote share V_{Ii} as follows:

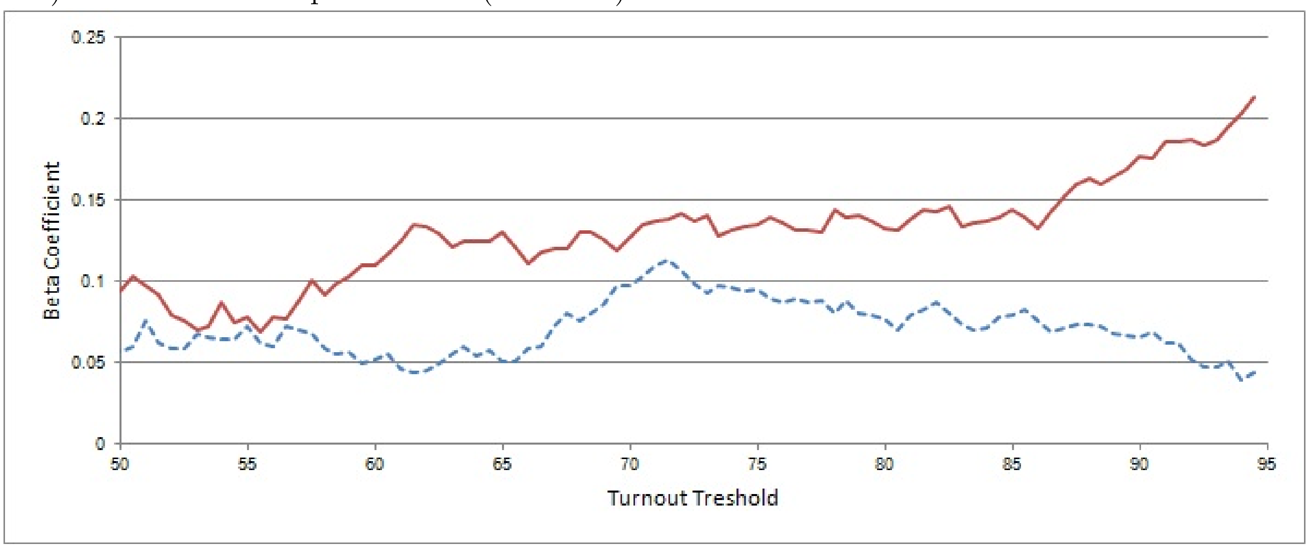
$$\hat{t_i} = t_i + \theta. \tag{6}$$

$$\hat{V_{Ii}} = \frac{V_{Ii}t_i + \theta}{t_i + \theta}. \tag{7}$$

Indeed, the higher θ the easier for the test to reject the hypothesis of no ballot stuffing. Thus, I choose the smallest value of θ such that the null hypothesis is rejected at the 99% confidence level. To guarantee 99% confidence rejection, θ should approximately be 0.18, which gives the incumbent about extra 1.9% of fraudulent votes on aggregate level measured as the difference between his before and after fraud vote shares. Figure 3 shows after fraud distributions of turnout. One can see that fraud results in a thicker right distribution tail.

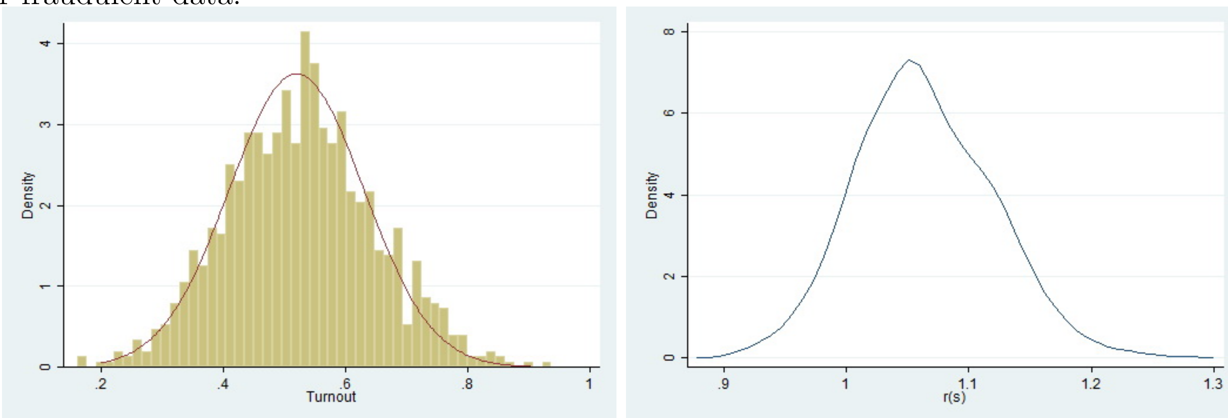
Once fraudulent data are generated I apply the detecting procedure described in Section 3. For this test threshold value is chosen to be at 55.5th percentile, where according to Figure 2 coefficient on turnout starts persistently growing.

Figure 2: Coefficient on turnout as a function of threshold value for clean artificial data (dashed line) and data with imputed fraud (solid line).



With $\theta = 0.18$, the value of the statistic is 1.215 and the 99% critical value is 1.185. Figure 3 presents the distribution of the statistic under the null hypothesis of no ballot stuffing. One can see that this distribution is not exactly as the one in Figure 1. This is because the distribution on Figure 3 is obtained using the after fraud data, and, as it was discussed above, the left tail of turnout distribution can still contain fraudulent precincts, which fully account for the observed difference.

Figure 3: True turnout distribution and distribution of the statistic under the null hypothesis for fraudulent data.



The exercise shows that the suggested methodology does not reject the hypothesis of fair elections even if there is natural positive correlation between turnout and share of votes for the incumbent but successfully detects with 99% confidence less than 2% fraud.

4.2 Finnish 2006 Presidential Elections

In this section the methodology is applied to real data that came from a presumably clean first round of the 2006 presidential elections in Finland. These elections were chosen as an example of direct executive elections whose integrity can be hardly put in doubt.[2]

Another reason why Finnish presidential elections were chosen to test the ballot stuffing detection methodology is that these elections were in some sense close to the Russian presidential elections, which are extensively analyzed in the next section. Though it is hard to believe that Russian and Finnish elections are truly comparable in any dimension, this is probably the best match one could do: the dates of the elections were not too far apart (I analyze Russian presidential elections of 2000, 2004, 2008 and 2012), the electoral systems in both countries are close, and the importance of elections is in some sense similar: directly elected presidents of Russia and Finland both have an executive power in contrast to the majority of European countries. In fact, only few countries in Europe have a directly elected president as an executive

[2]OSCE Office for Democratic Institutions and Human Rights (ODIHR), the largest and probably the most experienced organization that deploys elections observation missions in Europe, was requested to observe the Finnish Parliamentary elections of 2007 held one year after the presidential race. In their report OSCE analysts recommended that no OSCE/ODIHR election observation or assessment activity shall be undertaken in connection with the 18 March 2007 parliamentary elections. A tradition of democratic elections in Finland is accompanied by a commensurate level of public trust. All interlocutors expressed their overall confidence in the electoral process, and no immediate issues were brought to the attention of the Needs Assessment Mission that would necessitate OSCE/ODIHR involvement. Republic of Finland. Parliamentary Elections 18 March 2007, OSCE/ODIHR Needs Assessment Mission Report. Page 4. Available at www.osce.org/odihr/elections/finland/24126 (retrieved 01.10.2012).

(Armenia, Azerbaijan, Belarus, Bulgaria, Cyprus, Finland, France, Georgia, Lithuania, Poland, Romania, Slovakia, Ukraine) and Finland seems to be the best choice from this sample if one would like to have an example of a country which has as many as possible similarities with Russia in terms of electoral environment and, what is the most important for this paper, the highest confidence in electoral transparency and integrity.

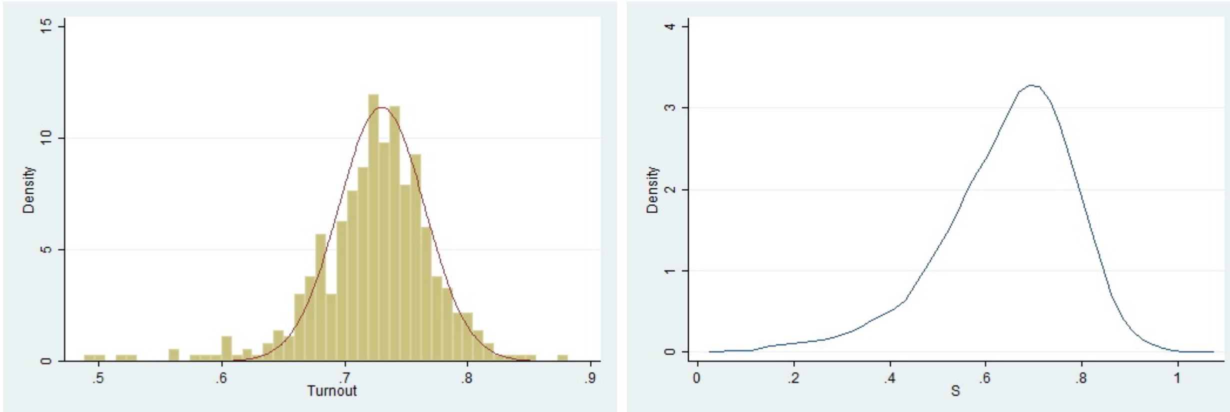
For the analysis I use the data for the first round of the elections. The main reason for such a choice is again an intention to make Finnish elections as comparable as possible to the Russian elections analyzed further. Because since 1996 the second round in Russian presidential elections has never been held due to the winning of one of the candidates in the first round, only first round Russian data are available for the analysis. Thus, it necessary to analyze first round data also for Finland as the ballot stuffing detection procedure for more than two candidate elections slightly differs from the one applied to the artificial data in the previous section. As it was discussed in Section 2, in such case the method requires analysis of the correlation between turnout and vote shares not only for incumbent but also for the challenger.

The dataset consists of 461 municipality-level (lowest available level) data observations which came from the Finnish public authority Statistics Finland[3]. To perform the analysis I first need to choose turnout threshold value. Following the approach suggested in Section 3 I draw coefficient on turnout form regression (2) as a function of threshold value. It can be seen from Figure 5 that there is no clear break of the trend which would suggest a choice of the threshold. So I choose two different threshold values at the 56.5th percentile where there is small growth of the coefficient and at the 88th percentile where one can see a small decline in the graph.

In both cases the test does not reject the hypothesis of no ballot stuffing: with turnout threshold value chosen at the 56.5th percentile the statistic is 0.685, while the 90% and the 10% critical values are 0.801 and 0.467 respectively, and for the 88th percentile threshold the value of statistic is 0.601, while the 90% and the 10% critical values are 0.733 and 0.422 respectively. Note that a critical value below 1 suggests that natural correlation between turnout and votes for the incumbent is lower than between turnout and votes for runner-up.

[3]www.tilastokeskus.fi

Figure 4: Turnout distribution and distribution of the statistic under H0 for the original data and 56.5 threshold.



Then, as in the previous case, I randomly choose 69 precincts (15%), artificially spoil them by adding additional 20% of votes in favor of the incumbent which gives him approximately extra 2.5% of votes, and conduct the test once again. Now the threshold value is chosen at the 81th percentile. With these 2.5% fraudulent data, the test rejects the null hypothesis of fair elections at the 99% confidence level: statistic 1.018, 99% critical value 0.928.

Figure 5: Coefficient on turnout as a function of threshold value for clean Finnish data (dashed line) and data with imputed fraud (solid line).

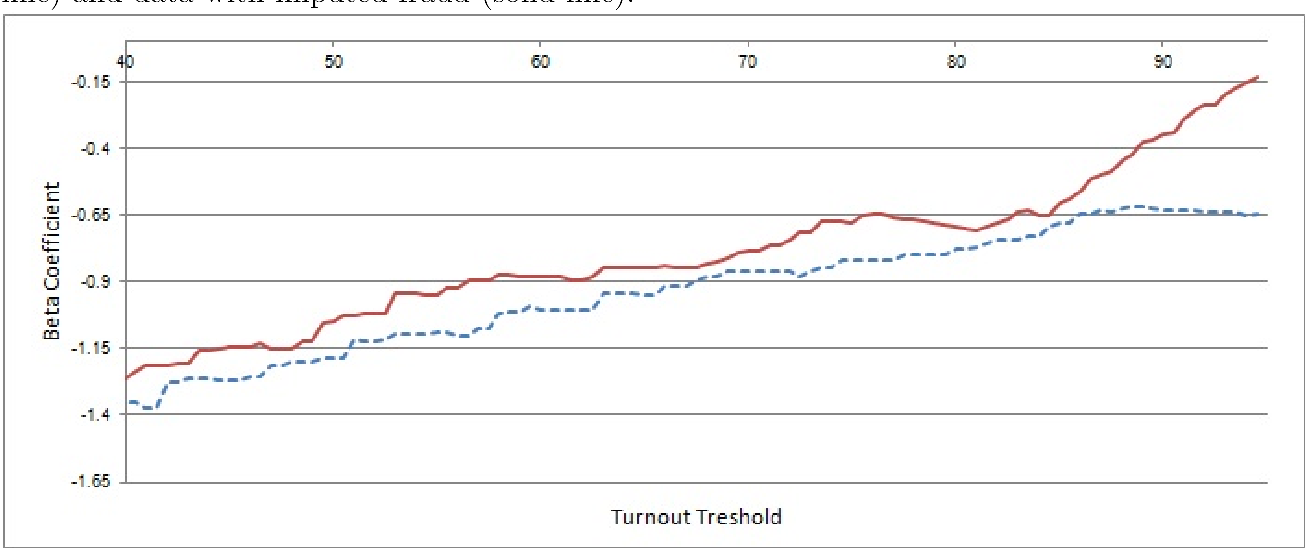
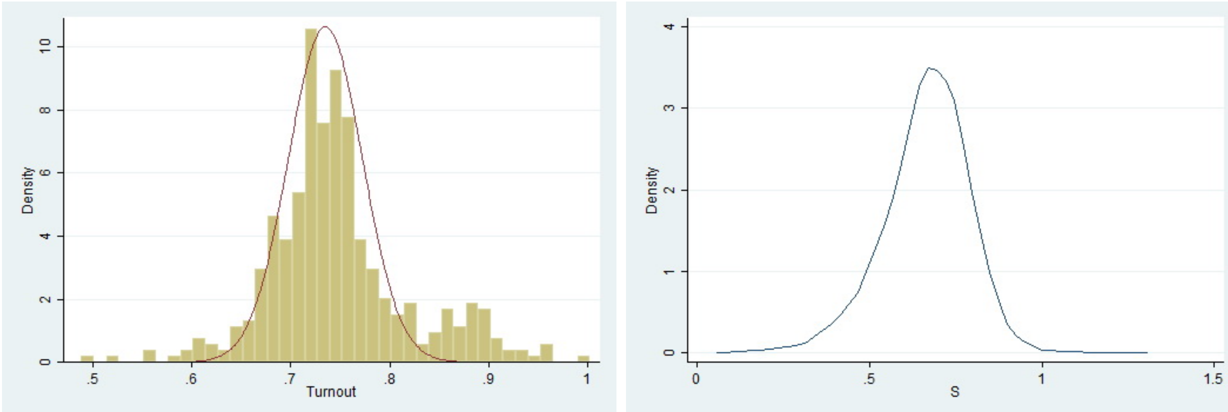


Figure 6: After fraud turnout distribution and the distribution of the statistic under H0 for fraudulent data.



Since fraud was artificially imputed in the data, it is easy to test the ability of methodology to provide an estimate of the ballot stuffing magnitude or, as it was discussed in Section 3, an estimate of the ballot stuffing lower bound. Recall that the estimate is obtained as the difference between actual number of votes for the incumbent and the number votes that incumbent should have received in order just not to reject the hypothesis of no ballot stuffing at the desired confidence level.

Imputed 20% of votes in 69 precincts corresponds in this exercise to approximately 100 200 extra votes for the incumbent. These votes give the incumbent 24.74% reported victory margin, while without fraud the incumbent would win at 22.24% margin. With turnout threshold value at 81th percentile the methodology provides the following estimates for various confidence levels:

Confidence Level	Counterfactual Victory Margin	Stuffed Votes	Underestimated Fraud
90	22.77%	79 500	20 700 (20.6%)
95	23.22%	62 000	38 200 (38.1%)
99	24.14%	24 600	75 600 (75.4%)

Table 1: Estimates of Ballot Stuffing Magnitude

According to Table 1 with 99% confidence the number of ballots stuffed in favor of the incumbent is at least 24 600, and victory margin should not exceed 24.14%, while the observed margin is 24.74%. For 90% confidence level these numbers are 79 500 and 22.77% respectively, which is quite close to actual number of stuffed ballots (100 200) and true margin (22.24%). Given that the methodology performs good on artificial and artificially fraudulent data in terms of both testing data for presence of ballot stuffing and estimating ballot stuffing magnitude, the next natural step is to apply it to real presumably fraudulent data.

5 Russian Presidential Elections

In this section I apply the methodology to the Russian presidential elections of 2000, 2004, 2008 and 2012. The fairness and transparency of Russian elections are often questioned, and evidence of electoral misconduct regularly appear in academic research (see, for example, Treisman (2009), Myagkov and Ordeshook (2008), Sakwa (2005)), reports of international observers[4], press, etc.

In all the four election the incumbents won with an overwhelming advantage in the first round by receiving more than 50% of the votes. The officially reported results of the elections are summarized in the following table.

Year	Turnout	Winner's Votes	Runner-Up's Votes	Victory Margin
2000	68.64%	39 740 467 (52.99%)	21 928 468 (29.24%)	23.75%
2004	64.38%	49 565 238 (71.31%)	9 513 313 (13.69%)	57.62%
2008	69.70%	52 530 712 (70.28%)	13 243 550 (17.72%)	52.56%
2012	65.34%	45 602 075 (63.60%)	12 318 353 (17.18%)	46.42%

Table 2: Official Results of the Russian Presidential Elections of 2000-2012.

Using the described fraud detection methodology I analyze these four consecutive presidential elections in Russia in order to test for the presence of ballot stuffing and obtain some comparable measures of its magnitude.

5.1 Country-Level Analysis

First, I follow the same approach as in the previous sections to analyze polling station level data obtained from the central elections commission of Russia, that contain information about the number of registered voters, turnout and votes cast for candidates at each polling station in 2000, 2004, 2008 and 2012 elections. The data contain approximately 95 000 observations for each elections.

The results of the analysis are presented in Table 3. The results provide clear evidence of persistent growth of ballot stuffing in Russian elections between 2000 and 2012. The estimates of ballot stuffing are provided for the 95% and 99% confidence levels. The percentages in the columns "Ballot Stuffing" are the differences between officially reported incumbent's victory margin and counterfactual, corrected for ballot stuffing of the corresponding level of confidence, victory margin. One may notice that the difference between the 95% and 99% confidence estimates are very small while in the example of Finland (see Table 1) it was relatively large. The main reason for this is the size of the data used for the analysis: large Russian data allow for

[4]See, for example, reports of OSCE on various federal-level elections in Russia: www.osce.org/odihr/elections/russia (retrieved 20.03.2013)

more precise analysis. Finally, it is important to note that though the estimates of the absolute numbers of stuffed ballots are substantial, ballot stuffing was not pivotal for determining the outcome of any of these four elections: without these ballots incumbent's would still win with overwhelming advantage in the first round. Even though the suggested methodology underestimates true fraud magnitude (see Section 3 for the discussion), underestimated fraud is unlikely to be large enough to alter any of the outcomes.

Year	Threshold	Statistics	95%		99%	
			Critical Value	Ballot Stuffing	Critical Value	Ballot Stuffing
2000	44.5	1.084	1.048	0.58% (559 600)	1.055	0.47% (455 500)
2004	73.5	2.170	0.945	3.98% (5 927 700)	0.948	3.96% (5 355 600)
2008	71.0	2.379	1.221	4.35% (6 236 200)	1.226	4.33% (6 204 800)
2012	40.0	1.525	0.937	9.14% (10 258 700)	0.941	9.08% (10 295 200)

Table 3: Russian presidential elections 2000-2012.

Since, I want to not only test the data for the presence of fraud but also credibly compare its magnitude across these four elections, I perform an additional analysis where I use the same threshold value for all four tests - 60th percentile value of the turnout distribution. As discussed in Section 3, particular choice of threshold value affects the estimate of ballot stuffing magnitude, and thus comparison of the estimates across different elections obtained with different thresholds may not be fully correct. In fact, this estimate should be thought of as a lower bound of ballot stuffing magnitude, and thus it may still provide some information about relative fairness of different elections. Given that threshold value is chosen at the level of 60th percentile for all four elections, higher estimate of the lower bound would signal, though imperfectly, about higher magnitude of ballot stuffing. The results of the analysis are summarized in Table 4.

Year	Statistics	95%		99%	
		Critical Value	Ballot Stuffing	Critical Value	Ballot Stuffing
2000	1.184	1.011	1.50% (1 433 600)	1.014	1.47% (1 408 900)
2004	1.614	0.947	4.28% (6 343 800)	0.951	4.26% (6 311 800)
2008	2.128	1.189	5.89% (8 199 100)	1.194	5.86% (8 156 900)
2012	1.983	1.019	7.34% (8 562 500)	1.022	7.31% (8 538 300)

Table 4: Russian presidential elections 2000-2012. Common threshold.

The results of the analysis with common threshold are fully consistent with the results presented in Table 3 though the numbers slightly differ: ballot stuffing magnitude has been persistently growing between 2000 and 2012. Again, the estimated ballot stuffing was not pivotal: even without it the incumbent would win the first round with an overwhelming advantage in all four elections.

5.2 Regional-level Analysis

As it was discussed in Section 3, one of the crucial assumptions that underlies the described fraud detection approach is a certain degree of homogeneity between electoral precincts. When there are systematic differences in voters' behavior across precincts the method might produce not fully correct conclusions. The case of Russia is an example of such situation. Russian regions are very different in various aspects (economic and social conditions, demography, cultural and historical peculiarities, etc) which might affect voters' behavior. If the ballot stuffing detection methodology is applied to a dataset that contains electoral information on precincts with very distinct voting patterns, then the results and conclusions regarding fairness of the elections and especially estimates of fraud magnitude might be biased.

One way to deal with this issue is to split the data set on more homogenous subsets, apply the method to each subset separately, obtain fraud estimates for all the subsets and then aggregate them. In case of Russia splitting the country-level dataset on subsets by regions seems to be the most natural. Indeed, dividing data on large number of relatively homogenous subsets requires original dataset to be sufficiently large, and such detailed data are not always available. Fortunately, Russian central election commission provides such kind of data: the lowest level available datasets (polling station level) contain about 95 000 observations for each elections, and for the regions the number of observations varies between 700 and 3000 with several exceptions of very small regions, which is sufficient to perform the analysis.

I perform the analysis separately for each region for the elections of 2000, 2004 and 2008. In 2008 there were 83 regions in Russia. Complete analysis for all three elections was possible for 62 out of 83 regions. For several small regions there were to few observations (e.g. Chukotskiy and Yamalo-Nenetskiy autonomous discticts, the republic of Yakutiya) and in some regions fraud was so extensive that the implementation of the methodology was not possible (see the details below). As in the previous section I use in the analysis the same turnout threshold at 60th percentile of the turnout distribution in order to make fraud estimates comparable across both different elections and different regions. The results of the regional analysis are summarized in Tables 8-14 of the Appendix. The regions are organized by federal districts[5] of Russia. For each of the three elections the tables contain incumbent's reported victory margin (VM) and ballot stuffing estimates (Fraud) both in percentages and absolute number of stuffed ballots. Recall once again that ballot stuffing estimates are lower bounds of actual fraud and may substantially underestimate the real level of falsifications (see Section 3 for the discussion).

The elections of 2000 were relatively clean: fraud is detected in 13 regions and its magnitude is usually moderate with an average of about 0.8%, measured as the difference between actual

[5]Federal districts are 8 geography-based groupings of federal subjects (regions) of Russia for the convenience of operation and governing.

and 99% confidence fraud-corrected victory margins of the incumbent. In the elections of 2004 fraud is detected in 22 regions with an average of 1.32%. Finally, in 2008 fraud presents in 45 regions from the sample with an average of 2.29%. The elections of 2008 are the most fraudulent in terms of both number of fraudulent regions and estimated fraud magnitude. The highest level of ballot stuffing in 2008 is found in the republics of Tatarstan and Chuvashiya, Belgorodskaya, Voronezhskaya, Moskovskaya, Orlovskaya, Penzenskaya, Rostovskaya, Samarskaya and Tymenskaya oblasts, Primorskiy kray, as well as at the city of Moscow. It is important to notice that there are almost no exits from the pool of fraudulent regions over time: once fraud appears in a region it stays their almost for sure. This finding is consistent with Myagkov and Ordeshook (2008) who argue that ballot stuffing and some other forms of fraud in the mid-1990s presented only in a few ethnic Russian regions but then spread to the other regions with noticeable acceleration during the 2000s. There are jut 12 regions out of 62 in the sample where fraud is not detected at 99% confidence in all three elections: Astrakhanskaya, Kirovskaya, Kurganskaya, Leningradskaya, Lipetskaya, Magadanskaya, Tambovskaya and Tulskaya oblasts, the republics of Hakasiya and Komi as well as Yugra autonomous district.

Table 5 contains the results of the regional-level analysis aggregated at the level of federal districts. The table demonstrates that in the Russian elections held in the 2000s the most severe fraud took place in southern and Caucasian regions, central part of Russia (mainly because of the capital) as well as Volga regions among which there are a lot of ethnic republics (Bashkortostan, Chuvashiya, Mariy El, Mordoviya, Tatarstan, Udmurtiya). In contrast, Siberia, Ural and Northwestern are the most "clean" regions in all the elections. Note that the numbers in Table 5 are aggregates of numbers from the regions for which the analysis was possible. The analysis was not possible for some very small regions and regions with very extensive fraud. Since such extensive fraud regions are mainly from the South (all Caucasian republics, Krasnodarskiy kray) and Volga area (Bashkortostan, Mordovia), the aggregated results substantially underestimates the amount of fraud in Volga, Southern and North Caucasian federal districts in comparison to the other districts.

As discussed above, the suggested methodology is based on the number of assumptions, one of most crucial of which is that fraud must be implemented in a small number of precincts. If the majority of precincts in a given region are fraudulent, then the analysis is not able to detect fraud as it treats some share (60% in case of this particular analysis) of data as clean. Some Russian regions are an excellent example of the cases when the number of fraudulent precincts is so large, that the suggested methodology fails to detect falsification at all. Figures 7 and 8 depicts turnout distributions in the republic of Bashkortostan and Saratovskaya oblast. Both regions shows extremely skewed distributions with extremely high mean in all the three elections which signals extensive falsifications. The suggested fraud detection methodology cannot be applied

Federal District	2000		2004		2008	
	VM (%)	Fraud (%)	VM (%)	Fraud (%)	VM (%)	Fraud (%)
Central	19.46	0.18	52.27	0.91	48.02	3.08
Ural	30.35	0	61.85	0.15	53.16	1.51
Volga	24.29	0.86	55.42	1.34	49.69	2.98
Northwestern	44.11	0.12	61.65	0.11	51.62	1.48
Southern and North Caucasian	20.52	0.86	50.17	3.18	51.12	2.36
Far Eastern	13.85	0.49	47.37	1.62	44.42	1.81
Siberian	8.58	0.04	50.07	0.46	44.75	0.61

Table 5: Russian presidential elections 2000-2008. Regional-level analysis. Federal districts aggregates.

to such kind of data. Similar picture can be observed in the republics of Adygeya, Chechnya, Dagestan, Ingushetiya, Kabardino-Balkariya, Kalmykiya, Karachevo-Cherkessiya, Mordoviya, North Osetiya and Tyva.

Figure 7: Distribution of turnout across polling stations in the republic of Bashkortostan in the elections of 2000, 2004 and 2008.

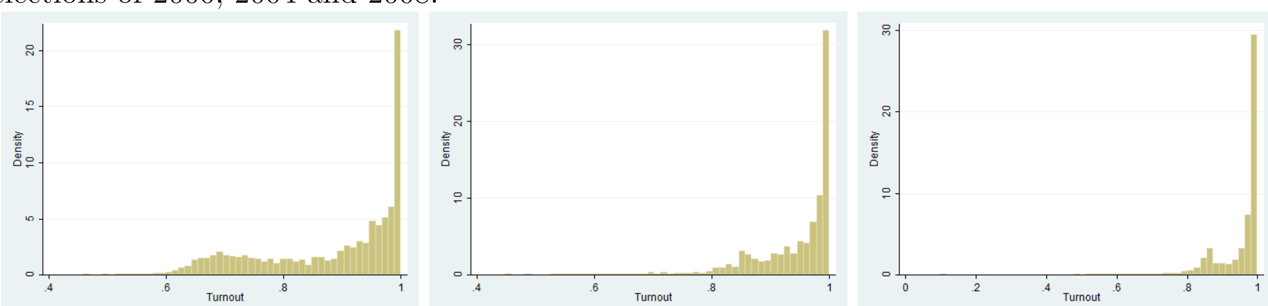
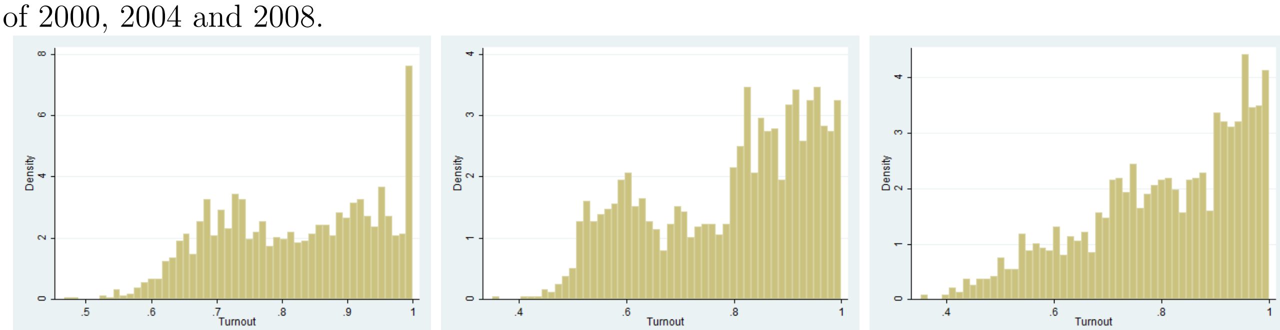
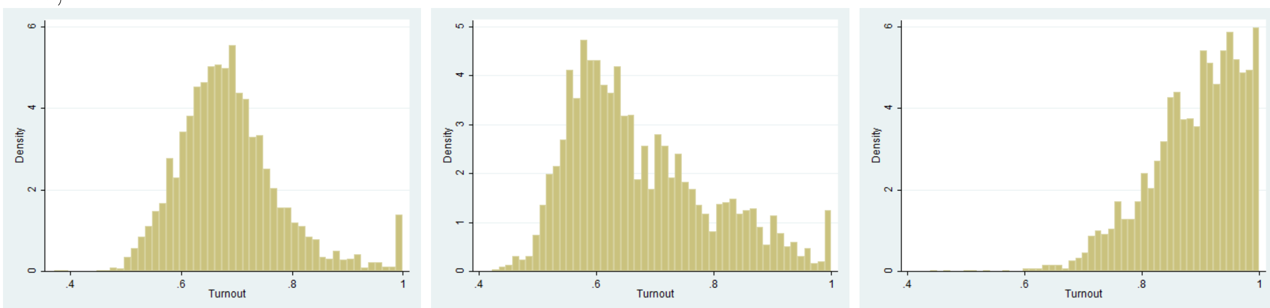


Figure 8: Distribution of turnout across polling stations in Saratovskaya oblast in the elections of 2000, 2004 and 2008.



In several regions such extremely skewed distributions are observed in 2008 or/and 2008 only, while in 2000 the turnout distributions have reasonable shape, and the data allows for performing the analysis. Such situation can be observed in Krasnodarskiy kray (see Figure 9) as well as in the republics of Buryatiya and Mariy El.

Figure 9: Distribution of turnout across polling stations in Krasnodarkiy kray in the elections of 2000, 2004 and 2008.



One way to deal with such situation is to apply the method in a usual way to the first elections, and then use the same underlying turnout distribution for all the following elections. Indeed, the results produced with such an approach may not be fully precise, but can still provide some idea about the extent of ballot stuffing and its dynamics over the period. The estimates are presented in Table 6. For each of the three elections the tables contain incumbent's reported victory margin (VM) and ballot stuffing (fraud) estimates both in percentages and absolute number of stuffed ballots. Again, the results are consistent with the general trend: fraud in these regions has been growing since 2000, both in absolute number of stuffed ballots and extra victory margin (except Bashkortostan). One may notice that in comparison to other regions (see Tables 8-14 of the Appendix) fraud estimates are not as large as they can be expected given extreme turnout distributions. The reason for such a result is that in the analysis I again use 60th turnout percentile threshold, while if fraud is very extensive a lower value of the threshold should be used (see Section 3 for the discussion). Table 7 contains fraud estimates obtained using the same analysis but with 50th turnout percentile value as the threshold. The estimates are much higher in all three cases, making the regions to be among the most fraudulent regions in Russia.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Bashkortostan	32.17	4.47 (142 100)	87.82	1.25 (238 800)	80.08	1.84 (242 700)
Krasnodarskiy kray	14.10	0	48.10	0	58.24	1.32 (95 000)
Saratovskaya	29.63	0.92 (19 753)	53.76	1.43 (43 200)	59.37	3.06 (102 600)

Table 6: Russian presidential elections 2000-2008. Regions with extreme turnout distributions. High turnout threshold.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Bashkortostan	32.17	6.75 (203 700)	87.82	1.76 (321 500)	80.08	2.93 (347 400)
Krasnodarskiy kray	14.10	0	48.10	0	58.24	2.15 (142 800)
Saratovskaya	29.63	1.85 (39 000)	53.76	2.58 (76 000)	59.37	4.32 (167 900)

Table 7: Russian presidential elections 2000-2008. Regions with extreme turnout distributions. Low turnout threshold.

To summarize, the main findings of the regional-level analysis are the following. First, ballot stuffing had been substantially growing between 2000 and 2008. This is a result of both increasing fraud magnitude and expanding of electoral falsification across the regions of Russia. This finding could serve as an additional argument and motivation for the idea presented in the second chapter of this dissertation where I argue that fraud has a tendency to grow over a lifetime of non-democratic regimes. Second, ballot stuffing is very persistent: once it appears in a region, it stays there in the following elections. Third, in some regions, primarily ethnic republic and southern regions, ballot stuffing is so extensive that the fraud detecting methodology suggested in the paper cannot be applied. For the regions where analysis is possible, the most severe fraud is detected again in ethnic republics, southern regions and the capital.

6 Conclusion

This paper suggests a simple statistical method for testing for the presence of ballot stuffing using official detailed electoral data. The method is based on the observation that ballot stuffing increases both turnout and the incumbent's vote share in precincts where it occurs. Hence, precincts with relatively low reported turnout are more likely to be clean. Using the information on relatively clean precincts, it is possible to simulate counterfactual data for infected precincts and compare them with the observed data.

The method is first piloted on artificial data and artificially fraudulent real data, and subsequently applied to test the fairness of the Russian executive elections in 2000, 2004, 2008 and 2012, whose transparency and integrity are dubious. Results strongly reject the hypothesis of no ballot stuffing in all four elections, while the estimates of ballot stuffing magnitudes suggest that fraud has been persistently growing over time. However in none of the elections ballot stuffing was sufficiently large to alter the winner. Finally, regional-level analysis of Russian electoral data shows that the most severe ballot stuffing takes place primarily in ethnic republics and southern regions of Russia.

Appendix

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Belgorodoskaya	7.76	0.33 (3000)	27.2	2.37 (26265)	47.33	4.12 (67153)
Bryanskaya	-2.93	0	40.29	0	34.72	3.04 (26818)
Ivanovskaya	22.87	0.33 (2716)	50.84	0.14 (1368)	44.77	1.08 (8902)
Kaluzhskaya	17.18	0	55.43	0	43.72	1.43 (13320)
Kostromskaya	33.45	0	50.68	0	39.71	0
Kurskaya	10.3	0.69 (5487)	44.16	0	42.44	0
Lipetskaya	-6.55	0	42.37	0	44.05	0
Moscow City	27.05	0	61.21	1.87 (191213)	55.05	4.62 (428352)
Moskovskaya	19.97	0.30 (13141)	60.47	0.74 (58525)	52.41	3.78 (277217)
Orlovskaya	1.22	0	37.63	0	43.61	3.26 (27362)
Ryazanskaya	12.30	0	59.52	0	36.59	1.38 (12708)
Smolenskaya	17.70	0	44.03	0	34.73	2.25 (18053)
Tambovskaya	6.78	0	39.31	0	53.18	0
Tverskaya	29.82	0	55.20	0	48.39	1.64 (23661)
Tulskaya	11.45	0	47.03	0		0
Vladimirskaya	22.43	1.46 (15372)	53.37	0.43 (6393)	42.11	1.28 (15769)
Voronezhskaya	24.79	0	43.33	2.05 (41977)	43.69	6.46 (137818)
Yaroslavskaya	43.14	0.44 (6021)	58.64	0.11 (1634)	42.93	1.80 (19516)

Table 8: Russian presidential elections 2000-2008. Central Federal District.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Chelyabinskaya	16.54	0	56.01	0	45.41	1.51 (49701)
Kurganskaya	11.98	0	47.78	0	44.29	0
Sverdlovskaya	46.65	0	68.56	0	55.80	0.91 (41771)
Tyumenskaya	25.35	0	62.37	1.14 (22129)	69.45	3.94 (98036)
Yugra	39.53	0	67.40	0	52.2	0

Table 9: Russian presidential elections 2000-2008. Ural Federal District.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Chuvashiya	1.4	0	45.59	0.77 (8867)	43.91	3.91 (45689)
Kirovskaya	31.33	0	45.91	0	62.13	0
Nizhegorodskaya	21.00	0	48.66	0	37.91	0.72 (21099)
Orenburgskaya	2.73	0	34.22	0	34.50	0.21 (3364)
Penzenskaya	11.00	0	44.40	0.53 (7012)	52.36	5.68 (88866)
Permskiy	40.83	0	62.66	0	50.60	1.24 (28646)
Samarskaya	12.45	0	44.16	0	41.45	4.22 (98080)
Tatarstan	48.81	3.49 (137965)	75.98	3.26 (272004)	66.30	4.90 (299358)
Udmurtiya	36.18	0	66.62	0.05 (995)	54.38	1.75 (27212)
Uljanovskaya	9.18	0	46.65	0	45.59	3.30 (37777)

Table 10: Russian presidential elections 2000-2008. Volga Federal District.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Arkhangelskaya	39.54	1.21 (14047)	68	0.24 (4696)	48.21	0.54 (6394)
Kaliningradskaya	36.69	0	58.07	0	38.85	3.15 (21427)
Kareliya	47.23	0	64.02	0	49.97	1.42 (8653)
Komi	38.14	0	61.94	0	56.95	0
Leningradskaya	47.64	0	66.92	0	52.26	0
Murmanskaya	50.52	0.21 (2163)	59.68	0	47.03	1.73 (14601)
Novgorodskaya	43.44	0	57.84	0	45.62	0.92 (5290)
Pskovskaya	36.91	0	54.16	0	49.83	4.10 (30199)
St Petersburg City	45.42	0	67.73	0.28 (17873)	55.49	2.22 (108752)
Vologodskaya	47.91	0	64.27	0	52.62	2.57 (7740)

Table 11: Russian presidential elections 2000-2008. Northwestern Federal District.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Astrakhanskaya	34.32	0	47.8	0	57.99	0
Volgogradskaya	19.53	0	41.07	0	38.02	0.20 (3935)
Rostovskaya	20.79	2.05 (59462)	58.27	6.15 (304304)	61.67	3.64 (199793)
Stavropolskiy	15.73	0	43.22	0	41.63	1.89 (39977)

Table 12: Russian presidential elections 2000-2008. Southern and North Caucasian Federal Districts.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Amurskaya	15.89	0	46.13	0	43.85	1.32 (10380)
Magadanskaya	39.50	0	59.68	0	42.92	0
Primorskiy	4.15	0.42 (14432)	42.25	2.10 (34428)	44.04	4.28 (68709)
Sakhalinskaya	15.94	0	54.14	0.96 (4529)	42.26	0
Khabarovskiy	21.49	0	51.5	2.32 (33482)	45.99	0

Table 13: Russian presidential elections 2000-2008. Far Eastern Federal District.

Region	2000		2004		2008	
	VM	Fraud	VM	Fraud	VM	Fraud
Altayskiy	4.09	0	49.05	0	37.1	0.9 (15901)
Buryatiya	1.84	0	49.5	1.47 (12655)	52.37	0.64 (6449)
Chitinskaya	13.65	0	57.71	0	48.47	1.84 (19633)
Irkutskaya	17.07	0	45.19	0.3 (5160)	39.42	1.94 (37491)
Kemerovskaya	10.08	0	60.2	1.73 (60544)	62.05	0
Krasnoyarskiy	15.54	0.23 (3730)	49.21	0	41.85	0.54 (11869)
Novosibirskaya	1.62	0	41.51	0	37.34	0.63 (13196)
Omskaya	-4.74	0.09 (974)	49.62	0	41.18	0.12 (2578)
Tomskaya	27.18	0	53.80	0.02 (140)		0
Hakasiya	6.04	0	43.06	0	37.69	0

Table 14: Russian presidential elections 2000-2008. Siberian Federal District.

References

Alvarez, M., Hall, T., Hyde, S., 2008. Election Fraud: Detecting and Deterring Electoral Manipulation. *Brookings Institution Press, Washington, D.C.*

Benford, F., 1938. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78, 4, 551572.

Berber, B., Scacco, A., 2008. What the Numbers Say: a Digit Based Test for Election Fraud Using New Data from Nigeria. *Presented at the Annual Meeting of the APSA.*.

Becker, R., Volinsky, C., Wilks, A., 2010. Fraud Detection in Telecommunications: History and Lessons Learned. *Technometrics*, 52(1), 2033.

Brady, H., 2005. Comments on Benford's Law and the Venezuelan Election. *Mimeo*.

Breuning, C., Goerres, A., 2011. Searching for Electoral Irregularities in an Established Democracy: Applying Benfords Law Tests to Bundestag Elections in Unified Germany. *Political Analysis*, 19(4), 409-433.

Cantu, F., Saiegh, S., 2011. Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning. *Political Analysis*, 19(4), 409-433.

Decker, J., Myagkov, M., Ordeshook, P., 2011. Benfords Law and the Detection of Election Fraud. *Political Analysis*, 19(3), 245-268.

Jacob, B., Levitt, S., 2003. Rotten Apples: An Investigation of The Prevalence and Predictors of Teacher Cheating. *The Quarterly Journal of Economics*, August, 843-877

Lehoucq, F., 2003. Electoral Fraud: Causes, Types and Consequences. *Annual Review of Political Science*, 6, 233-256.

Levin, I.,Cohn, G., Ordeshook, P., Alvarez, R., 2009. Detecting Voter Fraud in an Electronic Voting Context: An Analysis of the Unlimited Reelection Vote in Venezuela. *Electronic Voting Technology Workshop / Workshop on Trustworthy Elections*, Online Proceedings.

Li, J., Huang, K., Jin, J., Shi, J., 2008. A Survey on Statistical Methods for Health Care Fraud Detection. *Health Care Management Science*, 11(3), 275-287.

Lukinova, E., Myagkov, M., Ordeshook, P., 2011. Ukraine 2010: Were Tymoshenkos Cries of Fraud Anything More Than Smoke? *Post-Soviet Affairs*, 27(1), 37-63.

Mebane W., 2006. Election Forensics: Vote Counts and Benford's Law. *Working Paper*, University of Michigan.

Mebane W., Kalinin, K., 2009. Comparative Election Fraud Detection. *Presented at the Annual Meeting of the APSA.*

Myagkov, M., Ordershook, P., 2008. Russian Elections: An Oxymoron of Democracy. *VTP Working Paper.*

Myagkov, M., Ordeshook, P., Shakin, D., 2009. The Forensics of Election Fraud: Russia and Ukraine. *Cambridge University Press.*

Pandit, S., Chau, D., Wang, S., Faloutsos, C., 2007. NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. *Proceedings of the 16th international conference on WWW.*

Pericchi, L., Torres, D., 2004. La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela. *Reporte Tecnico No-Definitivo*, 2a.

Quash, J., Sriganesh, M., 2008. Real-Time Credit Card Fraud Detection Using Computational Intelligence. *Expert Systems with Applications*, 35(4), 17211732.

Roukema, B., 2009. Benfords Law Anomalies in the 2009 Iranian Presidential Election. *Mimeo.*

Sakwa, R., 2005. The 2003-2004 Russian Elections and Prospects for Democracy. *Europe-Asia Studies*, 57(2), 369-398.

Sudjianto, A.,Nair, S., Yuan, M., Zhang, A., Kern, D., Cela-Daz, F., 2010. Statistical Methods for Fighting Financial Crimes. *Technometrics*, 52(1), 5-19.

Treisman, D., 2009. Elections in Russia, 1991-2008. *Working Paper*, Moscow State University - Higher School of Economics.

Wolfers, J., 2006. Point Shaving: Corruption in NCAA Basketball. *American Economic Review*, 96(2), 279-283.