Measuring Regional Electoral Fraud

Abstract

This article introduces a novel quantitative electoral fraud index, designed to measure electoral fraud at the sub-national level. The index is constructed from precinct-level electoral data, combining two distinct statistical footprints of manipulation: (1) the Pearson correlation coefficient between voter turnout and the incumbent's vote share, indicative of ballot-box stuffing, and (2) the anomalous share of precincts reporting results as round percentage points, suggestive of protocol falsification. This paper outlines the methodology used in constructing the index. We apply this methodology to Russian elections in 2000-2018. We describe the features of the resulting dataset, which covers nine federal elections across Russian regions. The index is validated through correlation with external rankings of institutional quality. The statistically significant correlation with some of the rankings demonstrates the index's capacity to capture meaningful aspects of institutional quality. Finally, we describe the potential research utility of the index for studying political and institutional dynamics in countries that conduct fraudulent elections.

Keywords: electoral fraud, electoral integrity, sub-national politics

1. Introduction

Many authoritarian countries regularly conduct elections. Even though the results of such elections do not necessarily provide a clear understanding of citizens' political preferences, they nonetheless offer some information that can be used. Namely, electoral data can be statistically analysed to get insights into the process by which it was created.

Existing approaches to measuring electoral fraud mostly focus on the national level, with the goal to assess whether elections are conducted fairly. Applying statistical methods to precinct-disaggregated datasets allows us to measure malpractice at the sub-national level. In this article, we use the electoral data to construct a quantitative metric that captures the intensity of electoral fraud in regions. Systematic electoral falsification leaves statistical footprints in disaggregated election results. By focusing on these footprints, we can construct an objective, comparable measure of electoral malpractice.

Our index combines two well-documented statistical anomalies associated with electoral fraud: the correlation between turnout and incumbent vote share (Kiesling, 2004; Kobak et al., 2012) and the excessive reporting of round-number results (Kalinin & Mebane, 2011; Hillson, 2011). We argue that the prevalence of these irregularities serves as a quantifiable indicator of the extent to which regional administrations engage in, or permit, electoral manipulation.

We use nine elections in Russia in 2000-2018 to test our approach, leveraging a readily available, extensive dataset.

This paper proceeds as follows: First, we provide a brief review of the literature on electoral forensics and the specific types of fraud that our index aims to capture. Second, we detail the data sources and the step-by-step methodology for constructing the electoral fraud index, including its two sub-components. Third, we describe the key features and summary statistics of the resulting dataset. Fourth, we present a validation of the index by correlating it with an external measure of regional administrative behavior. Finally, we discuss the broader implications of the index and suggest avenues for future research, including how the conditions enabling such electoral fraud may reflect on wider aspects of regional governance. This index offers a valuable new tool for scholars of authoritarian politics, as well as researchers from other social science fields who may wish to explore the interaction between authoritarianism and electoral fraud with economic, societal, and cultural processes.

2. Detecting Electoral Fraud through Statistics

The usage of statistical methods to identify electoral fraud has become increasingly popular in political science since at least 1996, when J.B. Kiesling, an electoral observer in ODIHR's mission to Armenia, noticed that in a fraudulent election, the distribution of votes cast in precincts with different turnout was not normal: it had an unusually thick right tail, and the voter behaviour within this tail was different from that behaviour on other precincts, consistent with reports of ballot box stuffing and other types of fraud (Kiesling, 2004).

The methods have been developing since Kiesling's work. New approaches are being proposed and critiqued in statistics and political science papers, and several summary works systematise methods, including Hicken and Mebane (2017) and Shpilkin (2017). Additionally, forensic statistics methods are described in online laboratories for analysing electoral data (Electoral'naya Grafika [Electoral Graphics], 2013), (Election Forensics Toolkit, 2015).

Our index focuses on two common types of manipulation visible in precinct-level results: ballot box stuffing and protocol falsification.

- Protocol Falsification and Round Numbers: Falsifying final protocols to achieve predetermined results can also leave statistical traces. One such trace is an unusually high frequency of precincts reporting results as round percentage points. This can occur due to behavioral biases (ease of using round numbers), communicational simplicity (transmitting target figures), or political signaling by precinct officials attempting to demonstrate desired outcomes (Kalinin and Mebane, 2011), (Hillson, 2011), (Kobak et al., 2016), (Rozenas, 2017), (Kobak et al., 2020).
- Ballot Box Stuffing and Turnout-Winner Correlation: Ballot box stuffing, the illicit addition of ballots for a favored candidate, typically inflates both the turnout and the vote share for that candidate. This creates an artificial positive correlation between turnout and the winner's vote share in precincts where stuffing occurs (Kobak et al., 2012). Some researchers employ a Monte Carlo method to simulate various types of ballot box stuffing and then compare the resulting turnout and winner distributions with those of actual elections in Russia and Finland (Vorobyev, 2011) and Russia and Uganda (Klimek et al., 2012).

3. Constructing the index

The electoral fraud index is constructed as a panel dataset covering Russian regions for nine federal elections between 2000 and 2018.

3.1. Data Sources

The primary data for constructing the index are precinct-level results from Russian federal elections. We utilize datasets scraped by Sergey Shpilkin from the Russian Central Electoral Commission (CEC) website and subsequently published and maintained by Dmitry Kobak (Kobak & Shpilkin, 2020). These datasets include presidential elections (2000, 2004, 2008, 2012, 2018) and proportional representation components of parliamentary elections (2003, 2007, 2011, 2016). Data for the 2016 and 2018 elections contains datapoints for occupied Crimea and Sevastopol. The inclusion of this data in our analysis reflects the reporting practices of Russian electoral authorities and is retained for completeness.

Key variables extracted for each precinct include the number of registered voters, the number of valid ballots, the number of invalid ballots, and the number of votes for the ruling candidate or party.

3.2. Sub-Component 1: The Correlation Index

For each region and election year, we calculate the Pearson correlation coefficient between precinct turnout and the incumbent's share of valid votes at the precinct level.

Correlation coefficients smaller than 0.05 are treated as 0, as small correlations are not robustly indicative of this type of fraud.

To validate the assumption that ballot box stuffing generates such correlations, we perform Monte Carlo simulations, which confirm that fraudulent regions exhibit significantly higher correlations than fairly run regions (see more in Appendix 2).

$$r_{t,\mathrm{reg}} = \frac{\sum_{i=1}^{n}\left(\frac{\mathrm{votes\_for\_incumbent}_{i,t,\mathrm{reg}}}{\mathrm{turnout\_abs}_{i,t,\mathrm{reg}}} - \overline{\frac{\mathrm{votes\_for\_incumbent}_{t,\mathrm{reg}}}{\mathrm{turnout\_abs}_{t,\mathrm{reg}}}}\right)\left(\frac{\mathrm{turnout\_abs}_{i,t,\mathrm{reg}}}{\mathrm{registered\_voters}_{i,t,\mathrm{reg}}} - \overline{\frac{\mathrm{turnout\_abs}_{t,\mathrm{reg}}}{\mathrm{registered\_voters}_{t,\mathrm{reg}}}}\right)}{\sqrt{\sum_{i=1}^{n}\left(\frac{\mathrm{votes\_for\_incumbent}_{i,t,\mathrm{reg}}}{\mathrm{turnout\_abs}_{i,t,\mathrm{reg}}} - \overline{\frac{\mathrm{votes\_for\_incumbent}_{t,\mathrm{reg}}}{\mathrm{turnout\_abs}_{t,\mathrm{reg}}}}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(\frac{\mathrm{turnout\_abs}_{i,t,\mathrm{reg}}}{\mathrm{registered\_voters}_{i,t,\mathrm{reg}}} - \overline{\frac{\mathrm{turnout\_abs}_{t,\mathrm{reg}}}{\mathrm{registered\_voters}_{t,\mathrm{reg}}}}\right)^2}}$$

$$correlation\_index = \begin{cases} 0, if \quad r_{t,reg} < 0.05 \\ r_{t,reg}, if \quad r_{t,reg} \geq 0.05 \end{cases}$$

3.3. Sub-Component 2: The Roundness Index

This index captures the share of precincts reporting suspiciously round numbers for the incumbent's vote share and turnout.

1. Expected Roundness:

For each precinct, we calculate the probability of the incumbent's vote share (votes for incumbent/absolute turnout) and turnout (absolute turnout / registered voters) being a "round number" (defined as an integer percentage, or its floor/ceiling equivalent for non-integer products) by chance. We make four indicators, corresponding to whether the incumbent's vote share or turnout is divisible by 1% or 5%: (a) 001-incumbent, (b) 005-incumbent, (c) 001-turnout, and (d) 005-turnout. These indicators reflect whether the reported percentage falls exactly on a multiple of 1 or 5, such as 81.00% or 75.00%.

Below is an example of calculating the 001-incumbent probability for a precinct:

$$k \in \{0.01, 0.02, 0.03, \ldots, 1\}$$

$$\mathrm{round\_outcomes}_{\mathrm{abs\_turnout}} = \sum_{i=1}^{100} \begin{cases} 1, & \text{if abs\_turnout} \times k_i \text{ is an integer} \\ 2, & \text{otherwise} \end{cases}$$

$$p_{\mathrm{abs\_turnout}} = \frac{\mathrm{round\_outcomes}_{\mathrm{abs\_turnout}}}{\mathrm{abs\_turnout}}$$

$$p_{\text{precinct}} = p_{\text{abs\_turnout}}\ (for\ the\ absolute\ turnout\ at\ the\ precinct).$$

The 005-incumbent, 001-turnout, and 005-turnout probabilities are calculated similarly.

2. Observed vs. Expected:

For each region-year, we sum these probabilities to get the expected number of precincts with round results, and compare this sum with the actual number of such precincts.

$$\text{round\_expected}_{t,\text{reg}} = \sum_{i=1}^{n} p_i, (n\ is\ the\ number\ of\ precincts\ in\ the\ region)$$

$$\text{unround\_expected}_{t,\text{reg}} = \text{number\_of\_precincts}_{t,\text{reg}} - \text{round\_expected}_{t,\text{reg}}$$

3. Statistical Significance:

A Chi-squared test is performed to determine if the two numbers differ significantly. Precincts with <200 or >3000 voters are omitted from this calculation to avoid biases from very small (where every result would be round) or "special" (where the number of registered voters is statutorily the same as the turnout) precincts.

$$\chi^2 = \frac{(\text{round\_observed}_{t,\text{reg}} - \text{round\_expected}_{t,\text{reg}})^2}{\text{round\_expected}_{t,\text{reg}}} + \frac{(\text{unround\_observed}_{t,\text{reg}} - \text{unround\_expected}_{t,\text{reg}})^2}{\text{unround\_expected}_{t,\text{reg}}}$$

4. Index Calculation:

If the p-value of the Chi-squared test for all four indicators is greater than 0.05 (i.e., the null hypothesis is not rejected), the roundness index is 0. If the p-value is ≤ 0.05 for at least one indicator, the index is defined as the share of precincts in the region that reported a round result for that indicator. If multiple indicators have p-values ≤ 0.05, the final index is the maximum of these shares, taken from the indicators where the null hypothesis is rejected.

$$\text{round\_index}_{t,\text{reg}} = \begin{cases} 0, & \text{if } p\text{-stat}_{t,\text{reg}} > 0.05 \\ \frac{\text{round\_observed}_{t,\text{reg}}}{\text{precinct\_number}_{t,\text{reg}}}, & \text{if } p\text{-stat}_{t,\text{reg}} \leq 0.05 \end{cases}$$

3.4. The Composite

The final index for each region-year is the arithmetic mean of the two sub-components:

$$\text{fraud\_index}_{t,\text{reg}} = \frac{\text{correlation\_index}_{t,\text{reg}} + \text{round\_index}_{t,\text{reg}}}{2}$$

4. Features of the Russian elections

The resulting dataset provides a fraud index value for each Russian region for nine federal elections.

Summary statistics reveal several key features:

- Temporal Variation: There is a general trend of increasing electoral fraud, as measured by the index, from 2000, reaching a peak in the 2016 parliamentary elections, with slight decreases in the 2012 and 2018 presidential elections. This broadly aligns with narratives of increasing authoritarian consolidation in Russia during this period.

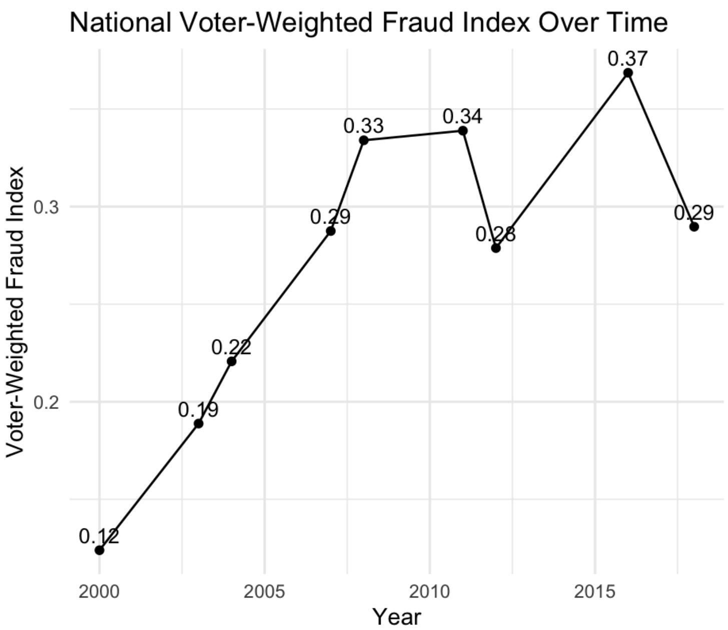

Figure 1. Fraud index, weighed by turnout, by election year.

- Regional Variation: The index reveals significant variation across Russian regions, with some consistently exhibiting higher fraud levels than others (e.g., Tatarstan, Bashkortostan, and the North Caucasian republics), which aligns with qualitative and quantitative expert assessments (Kobak et al., 2012; Gorbachev-Fond, 2012).

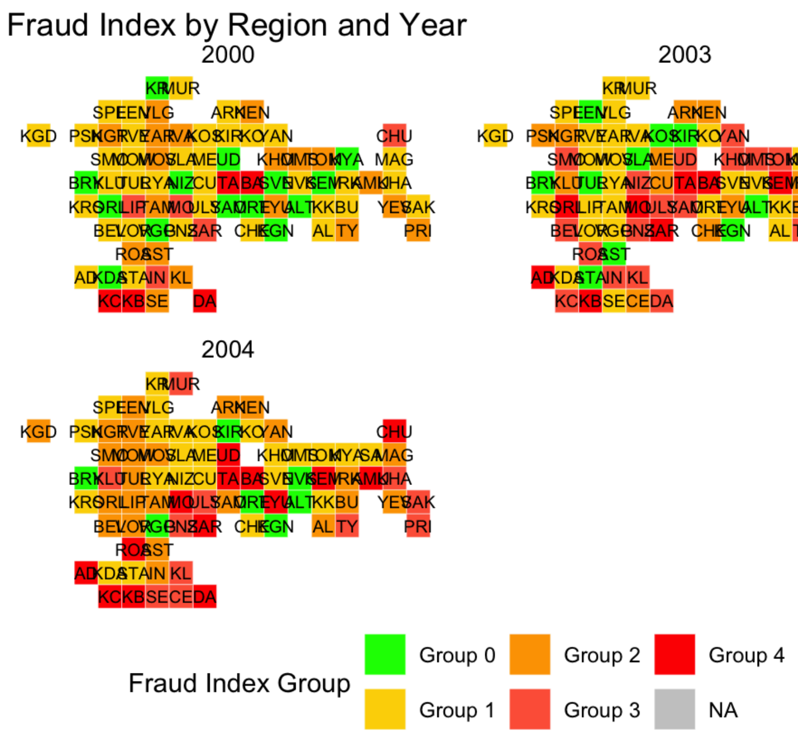

Figure 2a. Fraud index, by region, by election year (2000, 2003, 2004). Group 0 encompasses regions with a fraud index of 0; Groups 1 to 4 are based on quartiles of non-zero fraud index values across all region-year pairs.

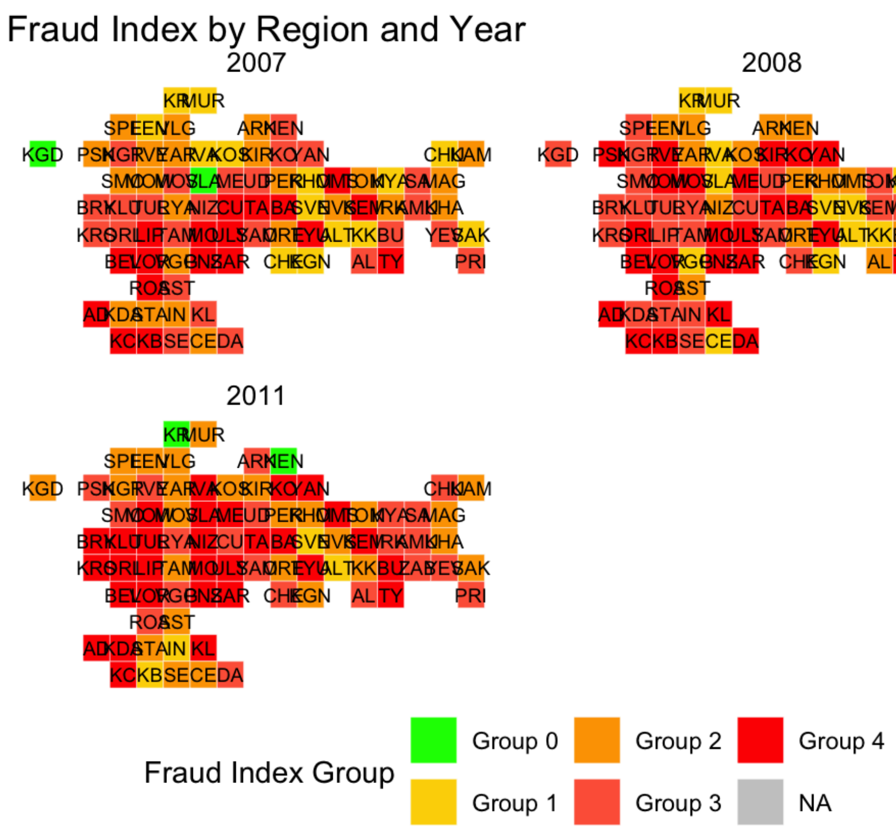

Figure 2b. Fraud index, by region, by election year (2007, 2008, 2011). Group 0 encompasses regions with a fraud index of 0; Groups 1 to 4 are based on quartiles of non-zero fraud index values across all region-year pairs.

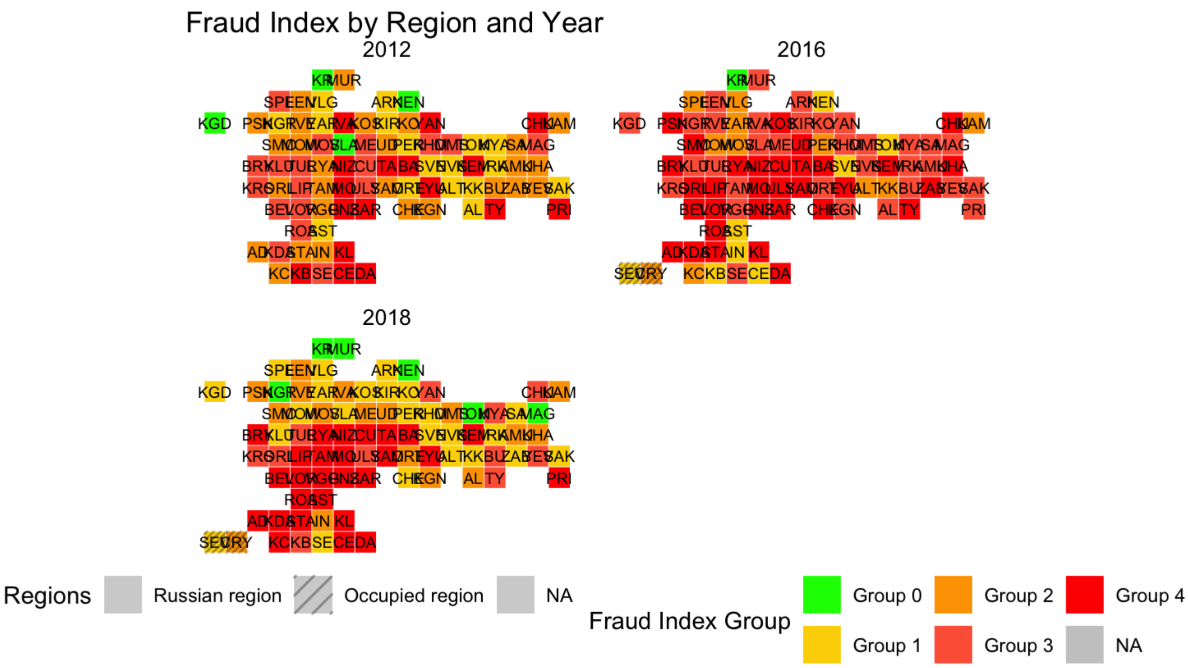

Figure 2c. Fraud index, by region, by election year (2012, 2016, 2018). Group 0 encompasses regions with a fraud index of 0; Groups 1 to 4 are based on quartiles of non-zero fraud index values across all region-year pairs.

- Component Correlation: The two subcomponents of the fraud index, the correlation index and the round index, are positively correlated across region-year observations, indicating some degree of overlap in what they capture. A pooled correlation across all years shows a statistically significant but modest relationship ($r = 0.21$, $p < 0.001$), suggesting that regions with higher anomalies in one dimension tend to exhibit irregularities in the other as well. This indicates that the two subcomponents capture related yet distinct aspects of electoral malpractice: while they often co-occur, they are not redundant, and offer complementary insights into the mechanics of electoral irregularities.

Correlation Between Fraud Index Subcomponents by Year

Pearson correlation between correlation_index and round_index

Year	Correlation	p-value		N Regions
2000	0.145	0.182		87
2003	0.201	0.058		89
2004	0.247	0.018	*	91
2007	0.026	0.809		86
2008	0.289	0.007	**	85
2011	0.021	0.851		84
2012	0.246	0.023	*	85
2016	0.152	0.164		85
2018	0.373	< 0.001	***	87

*Significance codes: *** < 0.001, ** < 0.01, * < 0.05*

Table 1: Correlation between the subcomponents of the Fraud Index by year.
No filling: parliament elections, blue filling: presidential elections.

Notably, the correlation between the two indices appears stronger and more consistently significant during presidential election years. In 2004, 2008, 2012, and 2018, the indices are significantly aligned, with the strongest association observed in 2018 ($r = 0.37$, $p < 0.001$). In contrast, parliamentary elections, such as those in 2007 and 2011, exhibit little to no correlation between the indices. This pattern suggests that presidential elections may be characterized by more systematic or centrally coordinated forms of electoral manipulation that affect multiple indicators simultaneously, while irregularities in parliamentary contests may be more fragmented or localized.

5. Validation of the fraud index

To assess whether the fraud index captures meaningful aspects of regional administrative behavior, we correlate it with external rankings.

A comprehensive review of different scorings utilised to measure the quality of institutions in Russian regions was conducted by Baranov et al. (2015). The authors review 20 different rankings that were available to them, from the "effectiveness of the executive authorities" by the Ministry of Regional Development to the number of businessmen murdered in business-related crimes. They identify the key problems with existing rankings. First, the lack of a systematic approach to computing ratings:
"The problem is that most institutional indices for Russian regions are calculated either once or with significant time intervals, and additionally, the list of regions for which calculations are made may vary from time to time" (Baranov et al., p. 22).
Second, the inability of ratings to comprehensively capture and summarise the phenomenon in question. The existing rankings show a weak covariation:
"Firstly, paired correlations are significantly different from zero for only slightly more than 40% of all possible pairwise combinations of institutional indices. Secondly, among the correlations that are significantly different from zero, only 70% have a positive sign, which would be expected in the case of a monotonic relationship between individual indicators and the overall state of regional institutions" (Baranov et al., p. 18).

One of the easily available metrics that the authors propose to use is the share of informally employed workers, which is published by the Federal State Statistics Service (Rosstat).

Several of the rankings referenced by Baranov et al. (2015) could not be located in public repositories or archives. Three sources were obtained:

- "Administrative Pressure – 2019" ranking, published by the Presidential Commissioner for Entrepreneurs' Rights (2019),

- "Index of Conditions for the development of small and medium-sized businesses in the manufacturing sector: Ranking of 39 Russian regions", published by the All-Russian public organisation for small and medium-sized enterprises "Opora Russia" (2012),
- The share of informally employed workers for the years 2001, 2004, 2007, 2009, 2012, 2013, 2015, and 2017, published by Rosstat (*Federal State Statistics Service, 2002, 2006, 2008, 2010, 2013, 2014, 2016, 2018*).

We compare these rankings with the rankings of regions by fraud index in the nearest elections: the 2012 elections for Opora's 2012 ranking, and the 2018 elections for the 2019 Administrative Pressure ranking.
We construct the fraud rank separately for each comparison by keeping only the regions included in the target ranking and use Spearman's correlation to compare the resulting rankings.

Correlation Between Institutional Rankings and Electoral Fraud				
Comparison	Spearman's ρ	p-value		N Regions
Opora 2012 vs Fraud 2012	-0.117	0.475		39
Administrative Pressure 2019 vs Fraud 2018	0.388	< 0.001	***	80
*Significance codes: *** < 0.001, ** < 0.01, * < 0.05*				

Table 2: Correlation between available institutional rankings and electoral fraud.

The fraud index for the 2018 election shows a statistically significant positive correlation with the "Administrative Pressure – 2019" ranking from the Presidential Commissioner for Entrepreneurs' Rights. This suggests that regions with higher observed electoral fraud also tend to be those where businesses experience greater administrative pressure. The non-significant correlation with the 2012 Opora's ranking may reflect the weak concordance of the rankings, as reflected by Baranov.

For the informality, we employ three different approaches to align the datasets. First, we check the correlation between informality and the fraud index only for the intersection of years, i.e., years where both Rosstat published informality data and elections were held (2004, 2007, and 2012). Second, we use the linear interpolation of informal employment data to estimate values for all election years. Third, we match each election to the closest available informality dataset. All three approaches yield similar results: a weak but statistically significant correlation. This suggests a connection between weak institutions (as evidenced by higher informality) and higher electoral fraud.

Correlation Between Informal Employment and Electoral Fraud			
Method	Correlation	p-value	
Matched years (2004, 2007, 2012)	0.137	0.028	*
Interpolated	0.116	0.005	**
Closest-year	0.177	< 0.001	***
*Significance codes: *** < 0.001, ** < 0.01, * < 0.05*			

Table 3: Correlation between informal employment and electoral fraud.

These validation results support the interpretation that the fraud index partially captures meaningful and observable institutional weaknesses.

6. Discussion

The fraud index offers a new tool for measuring electoral fraud at the sub-national level in countries that conduct unfair elections and publish their results. Its construction is transparent, and its components are based on well-established statistical indicators of electoral manipulation.

The capacity of a regional administration to orchestrate electoral fraud, quantified by this index, likely requires a significant degree of control over institutions (electoral commissions) and a disregard for legal norms. While this index directly measures electoral phenomena, these underlying administrative characteristics may have broader implications for regional governance. For instance, an administration that manipulates elections might also be one less constrained by legal norms in its dealings with businesses and civil society. This potential link to the general state of the rule of law or governance quality in these regions requires further investigation.

The fraud index can be used in a variety of future research for different countries, provided that they publish precinct-level data:

- Exploring the political determinants of sub-national electoral fraud.
- Investigating the relationship between electoral fraud and other socio-political outcomes like protest activity, public trust, or intergovernmental fiscal relations, investment and procurement.
- Researching its connection to sub-national measures of corruption, judicial independence, or property rights enforcement.
- Investigating the careers of bureaucrats that work in regions with low or high electoral fraud.
- Researching the relations between electoral fraud and media landscape or education.

A key limitation is that this metric is applicable only in contexts where (a) fraudulent elections are conducted, and (b) detailed precinct-level results are published.
Changes in electoral administration post-2018 in Russia (namely, the proliferation of e-voting without precinct disaggregation) may complicate the application of this methodology to more recent elections.

7. Data Availability

The fraud index dataset, covering Russian regions for nine federal elections between 2000 and 2018, along with the R files used for its construction and the analyses presented in this paper, are available at Github: https://github.com/iv-div/fraud_index_paper.

References

All-Russian public organization for small and medium-sized enterprises "OPORA Russia." (2012). *Entrepreneurial climate in Russia: OPORA Index-2012*. https://opora.ru/site/assets/files/49533/3__indeks_opory_2012.pdf

Baranov, A., Malkov, E., Polishchuk, L., Rochlitz, M., & Syunyaev, G. (2015). Izmerenie institutov v rossiiskikh regionakh: metodologiia, istochniki dannykh, analiz [Measuring Institutions in Russian regions: Methodology, Sources of data, Analysis]. *Voprosy Ekonomiki*, *2*, 69-103. https://doi.org/10.32609/0042-8736-2015-2-69-103

Election Forensics Toolkit. (2015). Center for Political Studies, University of Michigan, Institute for Social Research. Retrieved July 15, 2024, from https://electionforensics.cps.isr.umich.edu/

Electoral'naya grafika [Electoral Graphics]. (2013). Roman Udot. Retrieved July 15, 2024, from https://www.electoral.graphics

Federal State Statistics Service (Rosstat). (2002). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2002* [Labour force, employment and unemployment in Russia: 2002]. https://rosstat.gov.ru/bgd/regl/B02_61/Main.htm

Federal State Statistics Service (Rosstat). (2006). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2006* [Labour force, employment and unemployment in Russia: 2006]. https://rosstat.gov.ru/bgd/regl/B06_61/Main.htm

Federal State Statistics Service (Rosstat). (2008). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2008* [Labour force, employment and unemployment in Russia: 2008]. https://rosstat.gov.ru/bgd/regl/B08_61/Main.htm

Federal State Statistics Service (Rosstat). (2010). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2010* [Labour force, employment and unemployment in Russia: 2010]. https://rosstat.gov.ru/bgd/regl/B10_61/Main.htm

Federal State Statistics Service (Rosstat). (2013). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2013* [Labour force, employment and unemployment in Russia: 2013]. https://rosstat.gov.ru/bgd/regl/B12_61/Main.htm

Federal State Statistics Service (Rosstat). (2014). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2014* [Labour force, employment and unemployment in Russia: 2014]. https://rosstat.gov.ru/bgd/regl/B14_61/Main.htm

Federal State Statistics Service (Rosstat). (2016). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2016* [Labour force, employment and unemployment in Russia: 2016]. https://rosstat.gov.ru/bgd/regl/B16_61/Main.htm

Federal State Statistics Service (Rosstat). (2018). *Rabochaia sila, zaniatost' i bezrabotitsa v Rossii: 2018* [Labour force, employment and unemployment in Russia: 2018]. https://rosstat.gov.ru/bgd/regl/B18_61/Main.htm

Gorbachev-Fond (Mezhdunarodnyi fond sotsial'no-ekonomicheskikh i politologicheskikh issledovanii). (2012). K novoi modeli rossiiskogo federalizma: Vzgliad ikh regionov. Doklady proekta – 2012 [Toward a new model of Russian federalism: A view from the regions. Project reports – 2012]. Moskva: Gorbachev-Fond.

Hicken, A., & Mebane, W. R., Jr. (2017). *A Guide to Elections Forensics*. https://pdf.usaid.gov/pdf_docs/PA00MXR7.pdf

Hillson, R. (2011). *Statistical Methods for Detecting Anomalous Voting Patterns: A Case Study*. Naval Research Laboratory.

Kalinin, K., & Mebane, Jr., W. R. (2011). *Understanding Electoral Frauds through Evolution of Russian Federalism: from "Bargaining Loyalty" to "Signaling Loyalty."* Prepared for Presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, March 31-April 2, 2011. http://dx.doi.org/10.2139/ssrn.1668154

Kiesling, J. B. (2004). *Charting electoral fraud: Turnout distribution analysis as a tool for election assessment [Unpublished manuscript]*. https://www.academia.edu/2398547/Charting_Electoral_Fraud_Turnout_Distribution_Analysis_as_a_Tool_for_Election_Assessment

Klimek, P., Yegorov, Y., Hanel, R., & Thurner, S. (2012). Statistical detection of systematic election irregularities. *Proceedings of the National Academy of Sciences*, *109*(41), 16469-16473. https://doi.org/10.1073/pnas.1210722109

Kobak, D., & Shpilkin, S. (2020). *Polling Station Data for Russian Federal Elections* [Dataset]. https://github.com/dkobak/elections

Kobak, D., Shpilkin, S., & Pshenichnikov, M. S. (2012). Statistical anomalies in 2011-2012 Russian elections revealed by 2D correlation analysis. *arXiv Preprint arXiv:1205.0741*. https://arxiv.org/pdf/1205.0741

Kobak, D., Shpilkin, S., & Pshenichnikov, M. S. (2016). Integer percentages as electoral falsification fingerprints. *The Annals of Applied Statistics*, *10*(1). https://doi.org/10.1214/16-aoas904

Kobak, D., Shpilkin, S., & Pshenichnikov, M. S. (2020). Suspect peaks in Russia's "Referendum" results. *Significance*, *17*(5), 8-9. https://doi.org/10.1111/1740-9713.01438

Kobak, D. (2021, October 11). Russian elections once again had a suspiciously neat result. The Economist. https://www.economist.com/graphic-detail/2021/10/11/russian-elections-once-again-had-a-suspiciously-neat-result

Presidential Commissioner for Entrepreneurs' Rights. (2019). *Indeks "Administrativnoe davlenie – 2019": Kliuchevye problemy, reformy i predlozheniia v sfere kontrolia i nadzora. Prilozhenie k dokladu Prezidentu Rossiiskoi Federatsii 2019 [Index "Administrative Pressure – 2019": Key Problems, Reforms, and Proposals in the Sphere of Control and Supervision. Appendix to the Report to the President of the Russian Federation 2019].* http://doklad.ombudsmanbiz.ru/doklad_2019.html

Rozenas, A. (2017). Detecting Election Fraud from Irregularities in Vote-Share Distributions. *Political Analysis*, *25*(1), 41-56. https://doi.org/10.1017/pan.2016.9

Shpilkin, S. (2017). *Statistical analysis of elections*. Election Statistics Roundtable December 14, 2017.

Vorobyev, D. (2011). Towards detecting and measuring ballot stuffing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1965735

Appendix I

Fraud Index by Region and Year

Note: Data for the 2016 and 2018 elections contains datapoints for occupied Crimea and Sevastopol. The inclusion of this data in our analysis reflects the reporting practices of Russian electoral authorities and is retained for completeness.

Region	2000	2003	2004	2007	2008	2011	2012	2016	2018
Agin-Buryat Autonomous Okrug	0.141	0.127	0.271	0.258	NA	NA	NA	NA	NA
Altai Krai	0.000	0.000	0.000	0.075	0.031	0.108	0.058	0.207	0.149
Altai Republic	0.071	0.052	0.199	0.274	0.211	0.329	0.124	0.284	0.264
Amur Oblast	0.174	0.000	0.373	0.272	0.212	0.266	0.208	0.322	0.227
Arkhangelsk Oblast	0.106	0.225	0.163	0.172	0.195	0.272	0.158	0.276	0.100
Astrakhan Oblast	0.176	0.000	0.252	0.319	0.177	0.244	0.114	0.158	0.371
Baikonur City (Republic of Kazakhstan)	NA	NA	0.062	NA	0.071	NA	0.191	NA	0.087
Belgorod Oblast	0.101	0.325	0.232	0.430	0.545	0.406	0.346	0.584	0.409
Bryansk Oblast	0.000	0.000	0.000	0.304	0.349	0.376	0.291	0.560	0.513
Chechen Republic	NA	0.205	0.283	0.245	0.094	0.171	0.540	0.072	0.371
Chelyabinsk Oblast	0.114	0.248	0.136	0.031	0.310	0.320	0.199	0.467	0.115
Chita Oblast	0.000	0.102	0.055	0.100	NA	NA	NA	NA	NA
Chukotka Autonomous Okrug	0.329	0.338	0.369	0.158	0.335	0.274	0.424	0.429	0.295
Chuvash Republic	0.099	0.235	0.130	0.360	0.353	0.357	0.342	0.412	0.361
Evenk Autonomous Okrug	0.085	0.277	0.357	NA	NA	NA	NA	NA	NA
Irkutsk Oblast	0.113	0.206	0.166	0.191	0.290	0.285	0.114	0.274	0.056
Ivanovo Oblast	0.172	0.036	0.034	0.107	0.076	0.406	0.402	0.350	0.214
Jewish Autonomous Oblast	0.186	0.331	0.260	0.356	0.487	0.277	0.232	0.341	0.274
Kabardino-Balkarian Republic	0.421	0.474	0.418	0.485	0.419	0.123	0.366	0.115	0.288
Kaliningrad Oblast	0.061	0.070	0.189	0.000	0.358	0.184	0.000	0.298	0.138
Kaluga Oblast	0.091	0.256	0.286	0.324	0.350	0.377	0.292	0.322	0.134
Kamchatka Krai	NA	NA	NA	0.262	0.335	0.226	0.245	0.240	0.234

Kamchatka Oblast	0.258	0.355	0.269	NA	NA	NA	NA	NA	NA
Karachay-Cherkess Republic	0.377	0.323	0.427	0.481	0.502	0.385	0.209	0.180	0.447
Kemerovo Oblast	0.000	0.469	0.410	0.450	0.299	0.370	0.479	0.468	0.558
Khabarovsk Krai	0.079	0.285	0.267	0.163	0.169	0.240	0.211	0.365	0.199
Khanty-Mansi Autonomous Okrug – Yugra	0.169	0.317	0.155	0.144	0.183	0.261	0.267	0.309	0.149
Kirov Oblast	0.039	0.000	0.000	0.189	0.452	0.239	0.119	0.270	0.136
Komi-Permyak Autonomous Okrug	0.000	0.055	0.316	NA	NA	NA	NA	NA	NA
Koryak Autonomous Okrug	0.429	0.174	0.237	NA	NA	NA	NA	NA	NA
Kostroma Oblast	0.048	0.000	0.056	0.156	0.162	0.241	0.164	0.380	0.137
Krasnodar Krai	0.000	0.063	0.022	0.211	0.351	0.434	0.330	0.431	0.468
Krasnoyarsk Krai	0.000	0.074	0.083	0.146	0.102	0.304	0.133	0.332	0.293
Kurgan Oblast	0.000	0.000	0.000	0.127	0.126	0.244	0.195	0.322	0.168
Kursk Oblast	0.116	0.057	0.112	0.298	0.297	0.396	0.279	0.283	0.309
Leningrad Oblast	0.077	0.000	0.193	0.141	0.221	0.255	0.214	0.284	0.189
Lipetsk Oblast	0.285	0.129	0.215	0.388	0.329	0.377	0.327	0.407	0.590
Magadan Oblast	0.094	0.131	0.207	0.242	0.077	0.252	0.313	0.309	0.000
Moscow Oblast	0.215	0.137	0.196	0.340	0.407	0.218	0.276	0.247	0.255
Moscow city	0.014	0.017	0.206	0.224	0.391	0.388	0.212	0.186	0.110
Murmansk Oblast	0.082	0.131	0.333	0.059	0.156	0.252	0.215	0.295	0.000
Nenets Autonomous Okrug	0.218	0.212	0.184	0.296	0.253	0.000	0.000	0.118	0.000
Nizhny Novgorod Oblast	0.000	0.283	0.116	0.284	0.227	0.386	0.373	0.465	0.417
Novgorod Oblast	0.174	0.205	0.226	0.282	0.289	0.241	0.123	0.274	0.000
Novosibirsk Oblast	0.118	0.113	0.000	0.167	0.115	0.236	0.049	0.335	0.143
Omsk Oblast	0.198	0.271	0.150	0.373	0.207	0.365	0.277	0.341	0.233
Orenburg Oblast	0.000	0.035	0.000	0.258	0.167	0.252	0.147	0.332	0.175
Oryol Oblast	0.000	0.368	0.215	0.356	0.436	0.403	0.262	0.434	0.328
Penza Oblast	0.117	0.358	0.314	0.410	0.525	0.443	0.417	0.596	0.551

Perm Krai	NA	NA	NA	0.215	0.221	0.160	0.026	0.168	0.155
Perm Oblast	0.000	0.119	0.116	NA	NA	NA	NA	NA	NA
Primorsky Krai	0.178	0.318	0.359	0.296	0.363	0.268	0.445	0.322	0.375
Pskov Oblast	0.132	0.178	0.125	0.247	0.447	0.319	0.236	0.418	0.189
Republic of Adygea (Adygea)	0.028	0.387	0.404	0.418	0.590	0.381	0.165	0.501	0.372
Republic of Bashkortostan	0.485	0.462	0.583	0.593	0.598	0.575	0.583	0.461	0.501
Republic of Buryatia	0.115	0.165	0.261	0.299	0.258	0.362	0.263	0.305	0.300
Republic of Crimea	NA	NA	NA	NA	NA	NA	NA	0.181	0.180
Republic of Dagestan	0.425	0.351	0.397	0.293	0.384	0.279	0.483	0.480	0.467
Republic of Ingushetia	0.324	0.281	0.206	0.200	0.320	0.044	0.172	0.032	0.227
Republic of Kalmykia	0.227	0.348	0.277	0.357	0.426	0.417	0.389	0.414	0.371
Republic of Karelia	0.000	0.112	0.097	0.045	0.107	0.000	0.000	0.000	0.000
Republic of Khakassia	0.058	0.104	0.055	0.135	0.116	0.206	0.147	0.167	0.121
Republic of Komi	0.145	0.051	0.128	0.313	0.369	0.408	0.213	0.334	0.077
Republic of Mari El	0.118	0.217	0.096	0.338	0.594	0.430	0.352	0.274	0.170
Republic of Mordovia	0.325	0.610	0.613	0.634	0.653	0.647	0.590	0.692	0.648
Republic of North Ossetia – Alania	0.190	0.110	0.323	0.338	0.290	0.182	0.282	0.298	0.137
Republic of Sakha (Yakutia)	NA	0.000	0.046	0.288	0.129	0.350	0.258	0.282	0.045
Republic of Tatarstan (Tatarstan)	0.487	0.606	0.618	0.612	0.676	0.631	0.613	0.623	0.600
Republic of Tuva	0.164	0.308	0.322	0.496	0.562	0.498	0.500	0.557	0.358
Rostov Oblast	0.234	0.348	0.485	0.514	0.513	0.352	0.346	0.483	0.394
Ryazan Oblast	0.125	0.096	0.109	0.253	0.349	0.311	0.208	0.407	0.448
Saint Petersburg	0.013	0.013	0.049	0.225	0.320	0.229	0.296	0.229	0.156
Sakhalin Oblast	0.090	0.193	0.266	0.112	0.131	0.163	0.056	0.326	0.136
Samara Oblast	0.000	0.304	0.172	0.344	0.326	0.307	0.229	0.379	0.391
Saratov Oblast	0.350	0.409	0.495	0.363	0.537	0.579	0.512	0.589	0.561
Sevastopol	NA	NA	NA	NA	NA	NA	NA	0.034	0.126

Smolensk Oblast	0.076	0.293	0.166	0.260	0.328	0.358	0.232	0.359	0.189
Stavropol Krai	0.053	0.000	0.069	0.209	0.313	0.250	0.223	0.484	0.458
Sverdlovsk Oblast	0.000	0.019	0.095	0.051	0.093	0.103	0.032	0.110	0.049
Tambov Oblast	0.257	0.039	0.192	0.280	0.276	0.263	0.263	0.355	0.395
Taymyr (Dolgan-Nenets) Autonomous Okrug	0.345	0.387	0.406	NA	NA	NA	NA	NA	NA
Territory Outside the Russian Federation	NA	NA	0.399	0.286	0.312	0.314	0.309	NA	0.353
Tomsk Oblast	0.159	0.304	0.143	0.244	0.342	0.203	0.048	0.067	0.000
Tula Oblast	0.068	0.000	0.197	0.268	0.323	0.520	0.319	0.351	0.283
Tver Oblast	0.141	0.058	0.179	0.263	0.520	0.335	0.190	0.340	0.169
Tyumen Oblast	0.180	0.247	0.429	0.572	0.605	0.571	0.480	0.506	0.540
Udmurt Republic	0.000	0.331	0.363	0.278	0.339	0.306	0.247	0.595	0.196
Ulyanovsk Oblast	0.138	0.281	0.265	0.368	0.375	0.406	0.338	0.425	0.331
Ust-Orda Buryat Autonomous Okrug	0.116	0.023	0.134	0.305	NA	NA	NA	NA	NA
Vladimir Oblast	0.151	0.000	0.155	0.000	0.134	0.386	0.000	0.273	0.057
Volgograd Oblast	0.000	0.073	0.000	0.186	0.111	0.276	0.223	0.265	0.368
Vologda Oblast	0.204	0.028	0.137	0.184	0.227	0.201	0.128	0.200	0.085
Voronezh Oblast	0.070	0.155	0.245	0.405	0.389	0.457	0.354	0.637	0.481
Yamalo-Nenets Autonomous Okrug	0.135	0.333	0.228	0.316	0.402	0.467	0.425	0.317	0.308
Yaroslavl Oblast	0.232	0.119	0.126	0.198	0.182	0.217	0.071	0.254	0.068
Zabaykalsky Krai	NA	NA	NA	NA	0.187	0.278	0.162	0.412	0.105

Appendix II

In this Appendix, we will explain the method used to illustrate how ballot stuffing results in increasing the correlation between turnout and votes for the incumbent.

We construct a Monte Carlo simulation of 1,000 synthetic elections across 2,000 precincts. In each simulation, precincts are evenly divided between two hypothetical regions: Region 1, in which electoral fraud is artificially introduced, and Region 2, which serves as a control group without fraud.

For each precinct, we randomly generate the number of registered voters and simulate both turnout and share of votes for the incumbent candidate, using normally distributed values between 0 and 1. A subset of precincts (20%) in Region 1 is then subjected to ballot stuffing: fraudulent ballots are added both to the total turnout and the incumbent's vote count.

We then calculate the reported turnout percentage and incumbent share for each precinct and compute the correlation between these two indicators separately for each region. By repeating this process 1,000 times, we generate a distribution of correlation values for both the fraud-affected and control regions. The goal is to test whether the presence of ballot-stuffing induces a statistically significant positive correlation between the turnout and the incumbent share.

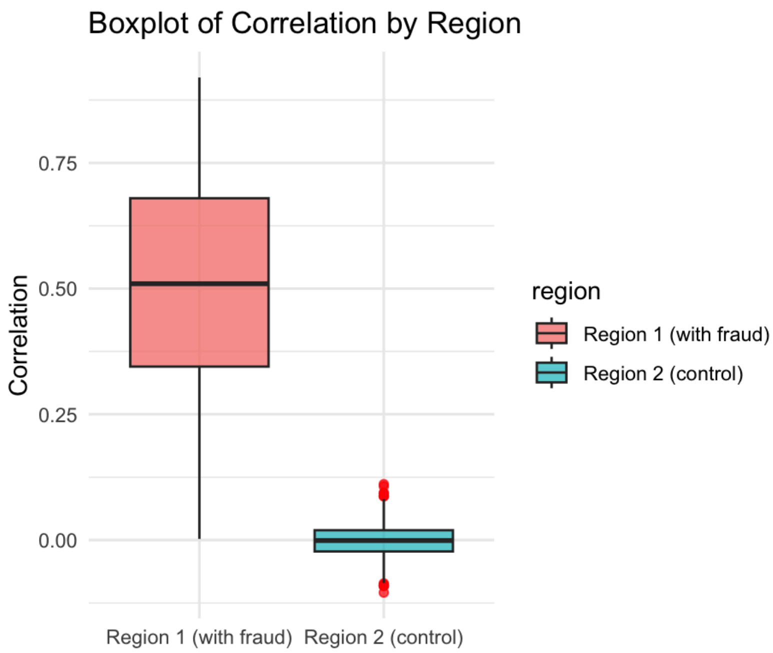

Figure 3. Distribution of turnout-incumbent share correlations across 1,000 simulations.

The results, visualized as a boxplot, show a clear upward shift in the distribution of correlations in the fraud-affected region compared to the control, thereby validating the use of turnout/incumbent correlation as a potential indicator of fraud.

Appendix III

Round_index visualisation.

A compelling visual representation of the fraudulent nature of Russian elections was presented by Dmitry Kobak (2021). A scatterplot of all precinct-level datapoints on the turnout and incumbent vote share axes reveals both types of fraud described in our article: a strong correlation between the two variables and a pronounced clustering of results on round-number values, forming a visible grid in the upper-right corner of the plot.

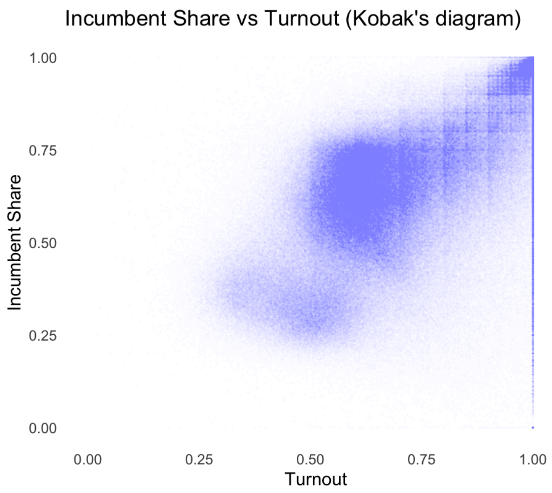

Figure 4. Reconstruction of the Kobak diagram with 2000-2018 election data.

We use this highly visible pattern to illustrate how well our round_index identifies regions with fraudulent elections. We split all data points into those corresponding to region-years with a zero round_index and those with a positive round_index.

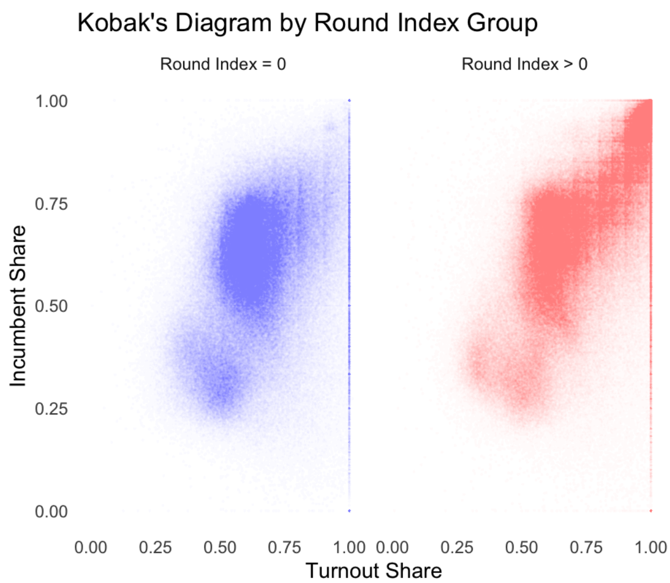

Figure 5. The Kobak diagram for regions with zero- (left) and non-zero (right) round_index.

The much weaker grid pattern among precincts from zero-index region-years supports the validity of the index in distinguishing between fraudulent and non-fraudulent regions.