



А.В. Подлазов

**Формальные методы выявления
масштабных электоральных
фальсификаций**

Рекомендуемая форма библиографической ссылки

Подлазов А.В. Формальные методы выявления масштабных электоральных фальсификаций // Проектирование будущего. Проблемы цифровой реальности: труды 2-й Международной конференции (7-8 февраля 2019 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2019. — С. 120-137. — URL: <https://keldysh.ru/future/2019/12.pdf> doi:[10.20948/future-2019-12](https://doi.org/10.20948/future-2019-12)

Формальные методы выявления масштабных электоральных фальсификаций

А.В. Подлазов

Институт прикладной математики им. М.В. Келдыша РАН

Аннотация. Описан набор тестов, позволяющих на уровне субъектов федерации обнаруживать масштабные фальсификации результатов выборов и оценивать их общий объем. Эти тесты основываются только на официальных данных о результатах выборов и не используют каких-либо предположений о поведении избирателей или эталонных результатов.

Ключевые слова: электоральная статистика, результаты выборов, фальсификация результатов, психологически привлекательные числа, жадное голосование, сгустки результатов, недействительные бюллетени

Formal methods of identifying large-scale election fraud

A.V. Podlazov

RAS Keldysh Institute of Applied Mathematics

Abstract. I describe a number of tests to detect large-scale fraud of the election results at the level of the subjects of the federation and estimate the total size of fraud. These tests base only on official data on election results and do not use any assumptions about voter behavior or reference results.

Keywords: electoral statistics, election results, falsification of results, psychologically attractive numbers, greedy voting, clots of results, invalid ballots

Введение

В России существует драматический разрыв в уровне развития общества и его институтов. А современные технологии, обеспечивая прозрачность, сами по себе не гарантируют публичности. И чтобы наблюдение происходящего выливалось в обсуждение и осознанные изменения, необходимы инструменты, позволяющие достоверно интерпретировать, что именно мы наблюдаем.

4. Математические модели цифрового мира

Выборы органов власти, будучи крупнейшим соцопросом, могут наиболее полно репрезентировать общественное мнение. К сожалению, в нашей стране результаты выборов систематически фальсифицируются, что не позволяет опираться на них в социологических исследованиях без предварительной реконструкции. Первым шагом к ее осуществлению должно быть определение и исключение тех регионов, где результаты фальсифицированы столь сильно, что исчезает сама основа для реконструкции.

ЦИК РФ публикует официальные результаты всех выборов с детализацией до избирательных участков. Здесь рассматриваются результаты 10 выборов федерального уровня: парламентских 1999, 2003, 2007, 2011 и 2016 гг. и президентских 2000, 2004, 2008, 2012 и 2018 гг. Данные за 1999 и 2000 гг. неполны и охватывают лишь 41% и 97% списочного состава избирателей соответственно, поэтому к результатам анализа за эти годы следует относиться с определенной осторожностью. Также надо помнить, что в 1999 г. у партии власти было два независимых крыла: движение «Единство» и партия «Отечество – Вся Россия».

Проверка на наличие фальсификаций осуществляется на уровне субъектов федерации по причине существенной разницы между ними как в масштабах фальсификаций, так и в предпочитаемых способах их осуществления.

1. Рисование результатов

Самой тяжелой формой фальсификации результатов является их выдумывание, когда числа в избирательных протоколах не находятся в какой-либо связи с содержимым урн. При столь грубом подходе в «результатах» преобладают *психологически привлекательные числа*. Для целочисленных электоральных характеристик таковыми являются круглые числа, а для процентных – значения без десятых долей (т.е. круглым оказывается количество промилле).

У целочисленной случайной величины с разбросом во многие десятки и даже сотни единиц последняя цифра должна практически равновероятно принимать все возможные значения. Если же встречаемость цифры «0» в конце числа превышает 10%, можно предполагать наличие выдуманных результатов. Проверка этого предположения сводится к проверке статистической гипотезы о естественном возникновении превышения. И если ее *уровень значимости α* (вероятность отвергнуть данную гипотезу, когда она верна) окажется мал, то наличие фальсификаций следует считать подтвержденным.

Регистрация на избирательном участке некоторой электоральной характеристики представляет собой испытание Бернулли с вероятностями *успеха* (круглое число) $p = 1/10$ и *неудачи* (некруглое число) $q = 9/10$.

Количество успехов описывается биномиальным распределением, для которого вероятность зарегистрировать не менее k успехов в n испытаниях $\alpha = \sum_{i=k}^n C_n^i p^i q^{n-i}$. Это – вероятность того, что, хотя встречаемость k/n и может казаться аномально высокой, ее подъем над уровнем p есть лишь результат стечения обстоятельств.

Поскольку в случае масштабных фальсификаций уровень значимости α чрезвычайно мал, далее для удобства вместо него всюду приводится его десятичный показатель $p\alpha = -\lg \alpha$, увеличение которого на единицу соответствует уменьшению на порядок вероятности отсутствия фальсификаций.

Российская Федерация состоит из без малого 100 субъектов. Поэтому на конкретных выборах для любой электоральной значимости $p\alpha$ начиная примерно с 2 следует считать *подозрительными*, с 3 – *исключительными*, а с 4 – *невероятными*. А при рассмотрении всего массива из 10 федеральных выборов эти классификационные пороги увеличиваются на 1 – до 3, 4 и 5 соответственно.

Чаще всего объектом фальсификации становятся следующие электоральные характеристики:

- количество избирателей, *принявших участие в выборах*;
- *явка избирателей* – доля зарегистрированных избирателей, принявших участие в выборах (получивших бюллетень);
- *результат власти* – доля избирателей, поддержавших партию/кандидата власти, измеряемая от числа принявших участие в голосовании.

Две последние величины, являющиеся дробями, при анализе превращаются в целые числа путем умножения на 1 000 и округления до ближайшего целого. Чтобы гарантировать значительную ширину разброса рассматриваемых характеристик, учитываются только те участки, на которых участие в выборах приняли не менее 100 человек. Это, кроме того, позволяет сделать пренебрежимо малой вероятность концентрации явки и результата на дробях с малыми знаменателями. Наконец, для явки исключаются из рассмотрения участки, где участие в выборах приняли все без исключения зарегистрированные избиратели, т.к. это может быть следствием отсутствия заранее составленных их списков, что делает явку фиктивной величиной.

Таблица. 1. Показатель $p\alpha$ для участия избирателей в выборах

Субъект федерации	1999	2000	2003	2004	2007	2008	2011	2012	2016	2018
Дагестан	1,4	17,8	34,8	26,3	16,7	31,8	103,0	27,6	20,5	8,1
Татарстан	0,8	0,6	0,2	1,8	8,6	23,5	1,2	2,0	2,3	4,3
Чечня	—	—	13,3	15,2	1,4	0,2	0,3	0,0	0,6	0,4
Башкортостан	0,9	3,9	0,3	2,9	11,4	7,9	3,2	4,2	2,0	1,9

4. Математические модели цифрового мира

Субъект федерации	1999	2000	2003	2004	2007	2008	2011	2012	2016	2018
Северная Осетия	0,9	1,1	1,5	4,4	0,5	3,6	1,7	2,9	8,5	7,0
Краснодарский кр.	1,1	1,8	1,6	1,0	1,8	6,9	0,7	1,1	3,4	4,0
Кемеровская обл.	0,8	0,9	0,1	2,4	4,1	7,0	0,4	1,6	5,7	2,9
Крым	—	—	—	—	—	—	—	—	0,4	5,2
Карачаево-Черкесия	1,1	1,4	0,7	3,9	4,0	5,0	0,6	2,4	1,1	3,8
Тюменская обл.	0,2	0,3	0,3	1,1	0,9	2,0	1,4	0,1	5,0	1,6
Хабаровский кр.	0,5	0,4	0,1	0,2	0,8	1,1	0,8	1,2	0,3	4,9
Кабардино-Балкария	1,0	2,3	4,5	1,5	0,8	2,8	2,1	2,5	2,5	0,5
Липецкая обл.	0,6	1,2	0,3	0,3	0,3	5,0	1,0	1,1	1,5	1,6
Самарская обл.	0,2	0,3	0,2	0,9	0,1	4,4	0,1	0,0	0,6	0,7
Пермская обл.	1,2	0,6	3,8	2,7	—	—	—	—	—	—
Ярославская обл.	0,8	0,4	1,9	0,2	3,7	1,5	0,8	0,1	0,7	0,2
Мордовия	0,5	0,4	0,2	0,6	3,2	2,0	1,6	0,9	0,8	0,2
Алтайский кр.	0,2	0,0	0,2	0,1	0,2	1,3	0,8	0,5	2,2	3,2
Ставропольский кр.	0,3	0,1	0,2	1,0	0,4	0,2	0,8	2,8	0,7	3,1
Пензенская обл.	0,3	0,6	0,4	1,7	3,0	1,3	0,3	0,6	0,3	0,4

Таблица. 2. Показатель $\rho\alpha$ для общей явки

Субъект федерации	1999	2000	2003	2004	2007	2008	2011	2012	2016	2018
Татарстан	0,9	0,2	0,3	0,5	0,8	6,5	11,8	16,3	49,2	27,3
Дагестан	1,5	1,0	1,1	17,2	5,8	8,8	1,4	8,3	20,3	30,2
Краснодарский кр.	0,4	0,1	0,7	0,6	1,5	3,1	1,5	1,9	1,7	26,7
Башкортостан	0,9	0,6	0,0	17,7	8,9	18,1	11,9	11,7	23,7	18,5
Кабардино-Балкария	0,7	2,0	5,3	4,6	4,6	22,1	0,2	9,7	1,3	0,4
Кемеровская обл.	0,2	0,4	0,4	3,1	2,7	6,4	0,9	4,1	17,4	17,0
Карачаево-Черкесия	0,4	0,7	0,1	0,5	1,1	9,7	3,0	2,0	1,0	3,9
Саратовская обл.	0,4	0,3	0,9	0,4	1,2	1,6	0,8	1,1	4,1	7,8
Северная Осетия	0,1	0,5	0,7	4,4	3,6	7,8	0,1	0,4	4,6	2,5
Мордовия	0,3	0,0	1,1	1,8	3,8	2,0	2,2	1,8	6,9	1,3
Ямало-Ненецкий а/о	0,1	0,0	0,0	0,6	0,3	2,0	0,2	6,9	2,1	0,5
Москва	0,2	0,5	0,3	2,1	0,5	5,2	6,0	0,5	1,5	0,3
Ставропольский кр.	0,1	0,2	0,2	0,2	1,1	0,2	0,1	0,1	0,5	5,3
Ростовская обл.	0,0	0,0	0,1	0,9	1,0	2,5	0,2	2,8	4,7	2,5
Ингушетия	0,3	0,0	0,5	0,4	1,6	0,9	0,2	4,2	0,5	0,3
Чечня	—	—	3,9	4,0	0,0	1,0	0,0	0,0	0,1	0,4
Московская обл.	0,5	0,4	0,4	0,8	3,1	3,6	0,9	0,7	0,7	2,5
Санкт-Петербург	0,4	0,3	0,4	0,7	0,5	3,3	0,1	0,7	0,5	0,4

Таблица. 3. Показатель $\rho\alpha$ для результата власти

Субъект федерации	1999	2000	2003	2004	2007	2008	2011	2012	2016	2018
Башкортостан	0,7	0,5	0,2	77,4	16,1	43,5	9,6	7,5	24,0	12,7
Татарстан	0,5	1,1	8,9	31,8	25,7	59,2	31,2	28,6	48,8	21,4
Дагестан	0,1	5,9	11,4	31,1	12,2	13,9	20,4	17,0	37,6	19,3
Кабардино-Балкария	0,4	2,5	3,2	4,7	7,4	26,7	4,9	1,9	1,9	0,2
Кемеровская обл.	0,9	0,5	0,1	1,3	3,4	16,1	3,2	7,0	3,2	6,7
Краснодарский кр.	1,6	0,2	0,2	0,4	0,0	12,7	3,3	3,4	1,1	5,0
Мордовия	0,2	0,2	2,4	12,2	12,4	10,1	12,2	1,4	10,9	2,2

Субъект федерации	1999	2000	2003	2004	2007	2008	2011	2012	2016	2018
Москва	0,4	0,1	0,1	0,5	0,4	1,3	10,1	0,3	0,9	0,4
Карачаево-Черкесия	0,1	0,2	1,2	0,3	9,9	8,0	1,0	0,2	0,0	1,8
Ростовская обл.	0,5	0,9	0,2	2,6	4,8	9,3	1,6	0,6	3,8	0,8
Чечня	—	—	3,2	5,6	0,8	0,5	0,0	8,6	0,0	0,7
Саратовская обл.	1,8	2,3	0,5	6,7	0,1	4,1	2,4	8,4	1,4	2,2
Ингушетия	0,6	0,3	0,7	0,9	6,5	0,2	0,3	0,4	0,6	0,3
Северная Осетия	0,4	0,9	0,0	1,6	2,2	2,7	0,3	0,8	5,2	0,6
Ставропольский кр.	0,7	0,2	0,2	0,9	0,7	0,1	2,6	1,4	1,8	4,6
Тюменская обл.	0,1	0,0	1,6	2,7	2,6	4,2	3,0	1,3	3,1	4,5
Тыва	0,4	0,8	0,1	0,2	1,3	2,5	1,6	2,7	0,3	4,3
Кировская обл.	0,1	0,0	0,7	1,8	0,3	3,2	0,8	0,0	0,1	2,2

В табл. 1, 2 и 3 приведены данные по всем субъектам, где указанные характеристики хотя бы раз оказывались круглыми на подозрительно большой доле участков. Субъекты отсортированы по убыванию максимального $\rho\alpha$.

Здесь и далее для выборов 1999 г. в качестве результата партии власти берется сумма результатов обеих ее крыльев. Для нее ожидаемо не наблюдается круглых процентов. Однако раздельное рассмотрение результатов создало бы ложное впечатление об их выдумывании как о методе фальсификаций на тех выборах, в силу специфики которых в некоторых субъектах на ряде участков то «Единство», то «Отчество – Вся Россия» вообще не имели ни одного голоса.

Невероятные показатели хотя бы по разу продемонстрировали 17 субъектов (Башкортостан, Дагестан, Ингушетия, Кабардино-Балкария, Карачаево-Черкесия, Крым, Мордовия, Северная Осетия, Татарстан, Чечня, Москва, а также Ямало-Ненецкий а/о, Краснодарский и Ставропольский кр., Кемеровская, Ростовская и Саратовская обл.), а исключительные – 22 субъекта (добавляются еще Тыва, Хабаровский кр., Липецкая, Самарская и Тюменская обл.).

Хотя фальсификации имеют тенденцию концентрироваться в одних и тех же субъектах федерации, то, что субъект прошел данный тест, еще не означает, что результаты выборов там совсем не рисовались. Если, например, в разных избирательных комиссиях субъекта сосредотачиваются на выдумывании различных электоральных характеристик, встречаемость круглых чисел может быть ограничена. Поэтому целесообразно обратиться к их интегральным мерам.

В качестве одной такой меры выбрано число субъектов, по участкам которых наиболее часто в конце рассматриваемых характеристик стоит «0» (если при этом другие цифры встречаются столь же часто, субъект учитывается с пропорционально меньшим весом), отнесенное к общему числу субъектов в рассматриваемом году N . Другой мерой круглости служит число субъектов с $\alpha > 1/N$ (в среднем должен встречаться 1 субъект с такой значимостью).

4. Математические модели цифрового мира

Изменение этих мер во времени, представленное на рис. 1, показывает, что по распространенности фальсификаций данного типа выборы 2008 г. держат абсолютное лидерство, а выборы 2018 г. ни для одной электоральной характеристики не выходят из первой тройки. Распространенное в обществе ощущение рекордных фальсификаций на выборах 2011 г. ошибочно. По-видимому, оно обусловлено тем, что тогда результат власти массово фальсифицировался даже в Москве, чего не наблюдалось в иные годы (см. табл. 3). И хотя по рисованию результата власти эти выборы держатся в первой тройке, другие электоральные характеристики тогда рисовались довольно умеренно.

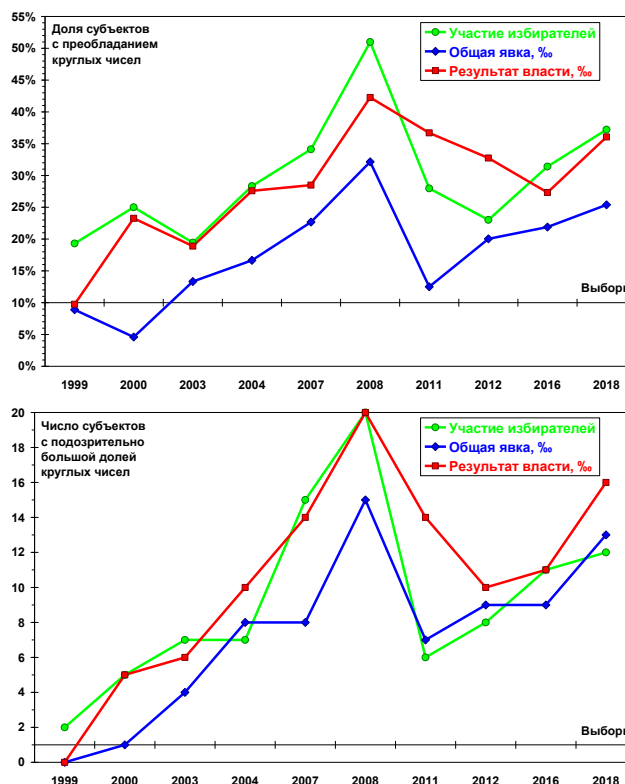


Рис. 1. Динамика рисования результатов

2. Сгустки явки и результатов

Возможным источником круглых промилле может быть и координация фальсификаторов из единого центра, если он по небрежности устанавливает целые *целевые проценты*. Однако ничто не мешает ему выбрать и нецелые, что позволит субъекту пройти тест на психологически привлекательные числа. Но в любом случае подобные фальсификации приводят к концентрации электоральных характеристик на узких интервалах их значений. Появляются *сгустки*, вероятность возникновения которых поддается оценке.

Для поиска и анализа сгустков рассматриваемой электоральной характеристики используется гистограмма ее распределения с карманами равной ширины. На ней выделяется *тестовый* карман и по s *опорных* карманов, расположенных с обеих сторон от тестового с пропуском карманов, непосредственно примыкающих к нему (см. рис. 2).

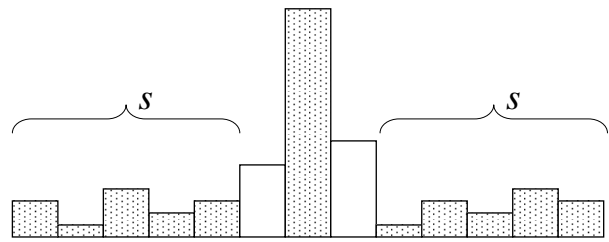


Рис. 2. Карманы гистограммы

Необходимость пропуска карманов, лежащих между тестовым и опорными, обусловлена тем, что сгустки иногда немного расплываются из-за дискретности избирателей, препятствующей точному попаданию в целевые проценты.

При условии, что система карманов достаточно узка, чтобы на ее протяжении плотность распределения рассматриваемой характеристики изменялась линейно, ее значения, попавшие в тестовый или опорные карманы, оказываются именно в тестовом с вероятностью $p = 1/(2s + 1)$. Если отношение количества участков в тестовом кармане k к их суммарному количеству в опорных и тестовом карманах n превышает p , можно предполагать наличие сгустка. При этом значимость гипотезы α о естественном возникновении превышения дается приведенной ранее формулой для испытаний Бернулли.

Субъект характеризуется величиной α_{\min} – минимальной из значимостей, зафиксированных при всех возможных положениях тестового кармана.

К сожалению, из-за того, что все рассматриваемые характеристики возникают как частное двух целых чисел, непосредственное применение описанного подхода приводит к появлению ложных сгустков на дробях с малыми знаменателями. Чтобы не допустить этого, необходимо предварительно размыть данные, для чего все дроби вычисляются m раз с добавлением к числителю случайной величины, равномерно распределенной на интервале $(-0,5; +0,5)$.

Кратный счет компенсируется делением на m количества успехов k и попыток n , что делает эти величины нецелыми, более не позволяя вычислять значимость как сумму биномиальных коэффициентов. Теперь следует использовать общую формулу $\alpha = I_p(k/m, (n - k + m)/m)$, где $I_p(x, y) = B_p(x, y) / B(x, y)$ – регуляризованная неполная бета-функция, которая, увы, не реализована в некоторых популярных пакетах обработки данных, что затрудняет анализ.

При определении параметров алгоритма здесь выбран простейший подход. Тестовые карманы шириной 1% располагаются симметрично

4. Математические модели цифрового мира

вокруг целых промилле рассматриваемых характеристик. Кратность размытия $m=10^3$ берется равной общему числу карманов. Суммарная ширина опорных карманов принимается равной 1%, что соответствует $s=5$. Такая система карманов представляется уже достаточно широкой для опоры на нее при сравнении сгустка и его окружения, но еще достаточно узкой для линейного изменения плотности распределения в большинстве субъектов, если ограничиться явкой и результатами лишь двух-трех сильнейших партий/кандидатов. Особо отметим, что возможность анализа результатов, показанных конкурентами власти, является важнейшим преимуществом данного метода.

Сложнее всего качественно классифицировать минимальную значимость α_{\min} . Для этого необходимо оценить эффективное количество попыток ее вычисления в некотором субъекте, что удается сделать лишь приблизительно.

Значения любой электоральной характеристики для львиной доли участков сосредотачиваются в диапазоне, не превышающем 30÷50% пунктов. И только в его пределах возможны флуктуации содержимого карманов, достаточно сильные, чтобы значимость α могла претендовать на роль минимума.

Кроме того, чтобы иметь хоть какие-то шансы оказаться местоположением сгустка, тестовый карман должен содержать больше участков, чем в среднем содержат опорные карманы. Это выполняется с вероятностью p .

Наконец, многократное стохастическое размытие данных приводит к некоторой переоценке значимости α из-за того, что один и тот же участок может одновременно оказываться и внутри тестового кармана и вне его, что нарушает независимость испытаний. Для оценки этого эффекта определим величину $\tilde{\alpha}$ как геометрическое среднее значимостей, получаемых при однократном размытии. Среднее (по всем выборкам) занижение показателя $p\alpha_{\min}$ как функция от его минимального учитываемого значения показано на рис. 3. Как можно видеть, переход от области несомненного отсутствия сгустков к области их вероятного наличия происходит при $p\tilde{\alpha}_{\min} - p\alpha_{\min} \approx 0,3 \div 0,5$.

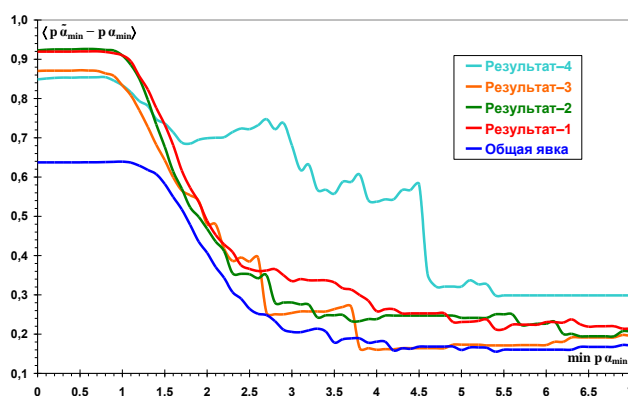


Рис. 3. Эффект кратного размытия

Подводя итог, можно сказать, что для каждого субъекта совершается примерно 10 попыток образовать сгусток естественным путем. Поэтому пороги качественной интерпретации для $p\alpha_{\min}$ следует взять на 1 больше

использованных для $p\alpha$: начиная примерно с 3(4) – подозрительно, с 4(5) – исключительно, а с 5(6) – невероятно для конкретных выборов (всего массива из 10 выборов).

Примечательно, что порог $p\alpha_{\min} = 3(4)$ соответствует на рис. 3 замедлению падения графиков (их выходу на стационар), что косвенно подтверждает адекватность полученных оценок. Этот же рисунок показывает, что при выбранных параметрах алгоритма не следует рассматривать сгустки результатов четвертого по силе участника (их распределение может быть слишком узким для линейного изменения плотности на масштабе системы карманов).

В 1999 г. надпороговых сгустков не было. Для остальных лет данные по всем субъектам, где общая явка или результаты трех лидирующих (в общефедеральном зачете) партий/кандидатов, хотя бы раз образовывали подозрительные сгустки, приведены в табл. 4. Выборы, для которых тест на сгустки по сравнению с тестом на круглые числа повышает качественную оценку показателей на одну ступень, отмечены зеленой заливкой, а на две и более – желтой.

Обращают на себя внимание несколько обстоятельств:

1) в Башкортостане, Дагестане, Коми, Мордовии, Татарстане, Москве, Санкт-Петербурге, а также Кемеровской и Ростовской обл. сгустки, по-видимому, образуются лишь на целых процентах, из-за чего данный тест не говорит о фальсификациях в этих субъектах ничего нового;

2) в Северной Осетии фальсификаторам удается (за счет варьирования целевых процентов) избежать сгустков результата власти, но при этом упускаются из вида результаты ее соперников и общая явка;

3) согласованно сверхвысокие показатели для всех характеристик на выборах 2016 г. в Саратовской обл. отмечают фальсификацию, уникальную даже по российским меркам (соответствующие гистограммы приведены на рис. 4).

Таблица. 4. Показатель $p\alpha_{\min}$ для общей явки и результатов первой тройки

Субъект	Величина	2000	2003	2004	2007	2008	2011	2012	2016	2018
Башкортостан	Общая явка	1,7	2,2	8,9	5,0	7,5	13,5	5,1	9,9	3,4
	Результат–1	1,8	1,8	7,5	5,4	9,7	6,5	3,0	18,6	6,2
	Результат–2	1,9	1,6	1,3	2,2	2,7	2,5	3,1	21,7	1,6
	Результат–3	1,9	1,2	1,0	2,0	2,7	2,2	2,4	19,4	1,8
Дагестан	Общая явка	1,8	1,9	10,2	4,9	23,8	2,5	10,3	13,1	11,4
	Результат–1	2,5	6,2	3,8	4,8	7,7	9,1	8,6	6,5	11,7
	Результат–2	1,6	6,8	2,4	8,8	10,7	11,4	8,4	6,1	2,6
Ингушетия	Общая явка	1,7	1,7	2,3	6,3	1,9	2,1	10,2	8,0	3,2
	Результат–1	1,4	1,5	1,5	9,8	10,4	6,4	2,7	11,1	6,8
	Результат–2	1,5	0,8	0,5	0,0	13,8	5,9	6,1	15,2	4,9
	Результат–3	1,2	1,0	0,1	0,4	5,7	6,2	5,2	18,3	8,3
Кабардино-	Общая явка	4,4	2,5	7,0	2,9	10,0	2,1	5,0	1,6	1,4

4. Математические модели цифрового мира

Субъект	Величина	2000	2003	2004	2007	2008	2011	2012	2016	2018
Балкария	Результат–1	3,0	4,1	5,3	2,9	10,2	6,4	2,6	6,2	3,0
	Результат–2	1,9	1,5	2,7	1,0	3,1	6,3	5,6	4,1	7,7
	Результат–3	2,0	2,0	0,8	1,0	7,1	1,5	4,3	1,0	13,2
Карачаево-Черкесия	Общая явка	1,8	2,0	2,5	2,3	4,4	2,4	1,6	3,2	2,8
	Результат–1	1,7	1,7	1,7	1,8	7,5	3,6	1,5	12,1	8,9
	Результат–2	1,3	1,4	1,7	2,1	3,3	8,5	1,7	24,6	10,9
	Результат–3	0,9	0,8	1,2	1,1	1,9	0,9	4,0	13,9	6,8
Коми	Общая явка	1,9	1,7	2,0	2,2	12,4	1,7	2,2	2,9	2,1
Мордовия	Результат–1	2,1	1,8	4,2	5,5	2,9	2,7	1,6	2,6	2,9
Северная Осетия	Общая явка	1,6	2,9	2,1	11,3	6,8	1,7	2,7	3,9	1,9
	Результат–2	2,0	1,5	1,7	1,2	1,4	7,0	7,5	15,8	2,5
	Результат–3	1,3	1,0	0,6	0,9	1,1	17,3	49,7	1,5	1,7
Татарстан	Общая явка	2,4	2,0	2,7	2,4	5,2	5,3	6,8	14,8	7,5
	Результат–1	1,9	2,1	4,0	3,9	24,7	4,3	4,9	11,2	3,5
	Результат–2	1,7	1,4	2,1	2,8	3,7	2,1	3,4	5,7	3,7
	Результат–3	1,5	1,1	1,8	1,4	3,5	1,6	2,0	4,8	3,7
Чечня	Общая явка	—	3,4	5,5	2,5	1,7	15,5	0,5	2,8	1,2
	Результат–1	—	2,0	2,3	8,3	1,3	39,6	0,5	7,8	2,4
Москва	Общая явка	2,7	1,7	2,2	2,0	4,0	3,0	2,2	2,2	3,0
	Результат–1	2,1	2,2	2,8	2,3	4,2	1,9	2,6	2,2	2,4
Санкт-Петербург	Результат–1	2,0	1,6	1,5	2,3	2,2	1,8	4,1	2,0	1,4
Краснодарский кр.	Общая явка	1,8	1,9	2,0	2,4	1,9	1,8	2,6	2,0	10,6
	Результат–1	2,1	1,4	1,6	2,1	8,1	5,9	2,3	2,4	5,0
	Результат–2	1,6	1,8	2,0	1,9	1,3	1,6	2,2	2,1	5,4
Ставропольский кр.	Общая явка	4,4	1,5	1,8	2,2	2,6	1,6	2,3	2,4	13,4
	Результат–1	2,2	1,2	1,6	2,0	1,7	1,9	2,1	1,6	6,5
Ханты-Манс. а/о	Результат–1	1,9	2,0	2,3	2,4	1,5	1,9	1,6	5,6	5,5
Кемеровская обл.	Общая явка	1,9	2,1	3,5	2,0	13,5	1,8	2,5	4,1	3,3
	Результат–1	1,2	1,4	1,5	1,7	9,0	3,1	5,3	2,2	7,1
	Результат–2	1,7	1,1	1,6	1,9	7,9	1,3	1,7	2,5	1,1
	Результат–3	1,4	1,0	1,5	1,6	11,0	1,6	1,6	6,2	1,2
Московская обл.	Общая явка	1,6	2,3	2,1	2,4	4,3	2,2	2,2	2,9	2,3
Ростовская обл.	Результат–1	1,6	1,7	3,4	5,1	6,6	2,4	1,6	2,4	2,1
Саратовская обл.	Общая явка	2,1	2,1	1,9	1,6	2,0	2,4	2,2	39,0	2,2
	Результат–1	1,9	1,8	2,6	1,4	1,8	2,3	2,9	58,5	1,7
	Результат–2	1,8	1,6	1,2	1,7	1,5	1,5	2,0	76,4	2,3
	Результат–3	2,0	1,4	1,3	1,1	1,4	1,6	2,0	82,2	1,1
Тюменская обл.	Результат–1	1,2	1,2	1,6	2,0	2,3	2,1	1,9	13,4	3,5
	Результат–2	1,5	1,4	1,4	1,5	2,2	1,4	2,0	27,9	2,4
	Результат–3	1,9	1,1	1,1	1,4	1,5	1,3	1,5	13,4	4,0

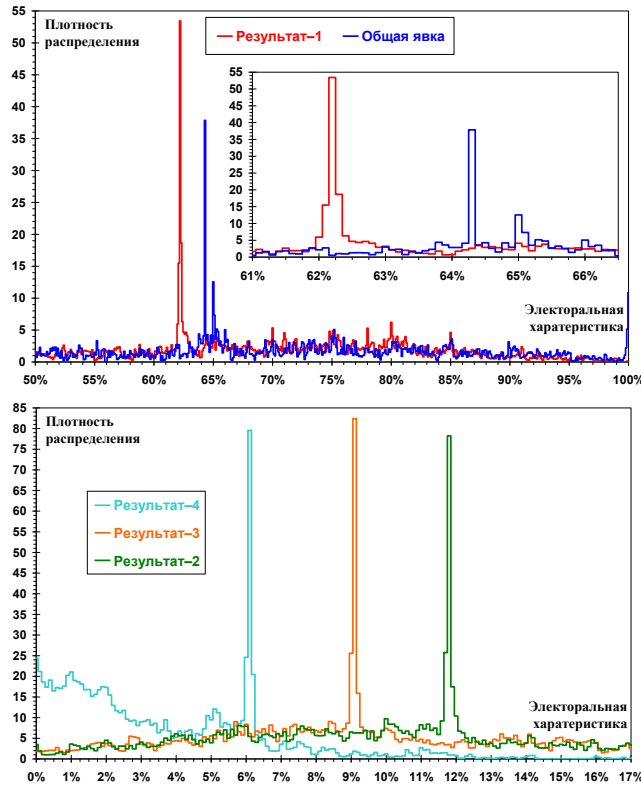


Рис. 4. Сгустки на выборах 2016 г. в Саратовской обл.

3. Жадное голосование

В табл. 5 дано общее количество случаев рисования с разными показателями значимости. Приведенные числа возрастают по строкам,

Таблица. 5. Сводка нарисованных результатов

Характеристика	$pa > 2$	$pa > 3$	$pa > 4$	$pa > 5$
Участие избирателей	89	45	33	21
Общая явка, %	73	51	42	33
Результат власти, %	101	64	54	46

что коррелирует с предполагаемой ценностью соответствующих электоральных характеристик для фальсификаторов. В самом деле, приоритетен результат власти, чуть меньше внимания уделяется явке, а количество избирателей уже второстепенно. Такой порядок нарушается лишь во втором столбце: подозрительная встречаемость круглых чисел для участия возникает в большем числе случаев, чем для явки.

Дело в том, что помимо выдумывания чисел из головы существует еще один механизм фальсификации, способствующий возникновению круглого количества избирателей, принявших участие в выборах. Количество избирательных бюллетеней, полученных участковой избирательной комиссией, зачастую само оканчивается на «0», т.к. их отсчитывают десятками. И если фальсификаторы приравнивают количество пришедших избирателей количеству имеющихся бюллетеней, это тоже может приводить к появлению круглых чисел.

4. Математические модели цифрового мира

Введем представление о *жадном голосовании*, понимая под ним ситуацию, когда количество избирателей, принявших участие в выборах, в точности совпадает с числом бюллетеней, полученных участковой избирательной комиссией, но при этом оказывается меньше числа избирателей, зарегистрированных в ее списках. Последнее ограничение необходимо, чтобы исключить «герметичные» участки, образованные в труднодоступных местах, где и в самом деле могут проголосовать все зарегистрированные избиратели, что требует получения такого же количества бюллетеней. Поэтому вполне возможно, что при полной явке и впрямь были использованы все бюллетени, однако крайне подозрительно, если это происходит при неполной.

Помимо участков с полной явкой из рассмотрения следует исключить участки, оборудованные комплексами электронного голосования. При их использовании избирательные протоколы зачастую (примерно в $\frac{3}{8}$ случаев) заполняют неправильно, указывая в качестве числа полученных бюллетеней не число одноразовых карточек для голосования, а просто число проголосовавших избирателей, что ошибочно может восприниматься как жадность.

Несмотря на то что жадное голосование на конкретном участке почти наверняка означает фальсификацию на нем результатов выборов, ее масштаб, как правило, сравнительно невелик. И ограничивает его как раз нехватка избирательных бюллетеней. Там, где их имеется в достатке, фальсификаторы, дойдя до желаемых результата власти и/или общей явки, могут и остановиться. А вот дефицит бюллетеней вынуждает использовать их все до единого.

То, что фальсификаторы имеют возможность и причины жадничать, в лучшем случае говорит об отсутствии эффективного контроля над процедурой голосования и подведением его итогов, а в худшем – о наличии спущенных сверху целевых показателей. Иначе говоря, жадное голосование на одних участках субъекта федерации, скорее всего, свидетельствует о фальсификациях и на многих других его участках, более изобильных бюллетенями. И чем выше распространенность жадного голосования, тем весомее это свидетельство.

Определим *жадность* субъекта g как долю его участков, где имело место жадное голосование. Чтобы преобразовать эту количественную меру в качественные категории, воспользуемся техникой ранг-размерного анализа. Для этого отсортируем все зарегистрированные значения жадности по убыванию и назовем номер значения в этом списке его *рангом* r . Ранг самого жадного субъекта r_0 , с которого начинается нумерация, является подгоночным параметром. При правильном его подборе для небольших r можно ожидать гиперболической зависимости $g(r) \sim 1/r$. В идеальном случае она описывает ранжировку чисел, обратных

к равномерно распределенным случайным величинам. Однако и в случае масштабных фальсификаций, когда на величину жадности уже сильнее влияют вариации не числа жадных участков, а их общего числа в субъекте, зависимость ранг–размер должна оставаться примерно гиперболической.

Анализ, подтверждающий эти рассуждения, представлен на рис. 5. Произведение $g(r) \cdot r$ практически не зависит от r в области малых рангов, логарифмически убывает в области средних и обрушивается в области больших. Зависимость в области малых и средних рангов в полулогарифмическом масштабе аппроксимирована непрерывной кусочно-линейной функцией с константным первым участком. Параметр r_0 , а также точки перелома и завершения обработки данных подбираются так, чтобы максимизировать сумму коэффициентов детерминации зависимости для обоих участков. Чтобы обеспечить одинаковую представительность всех частей оси абсцисс, данные учитываются с весом $1/r$.

Проведенный анализ позволяет классифицировать фальсификации как *незначительные*, *умеренные* и *значительные* с соответствующими порогами жадности в 0,31% и 1,08%. Привести данные обо всех случаях умеренных или хотя бы значительных фальсификаций, выявленных таким способом, не представляется возможным, поэтому в табл. 6 представлены сведения о жадности голосования лишь в тех 26 субъектах, где она хотя бы

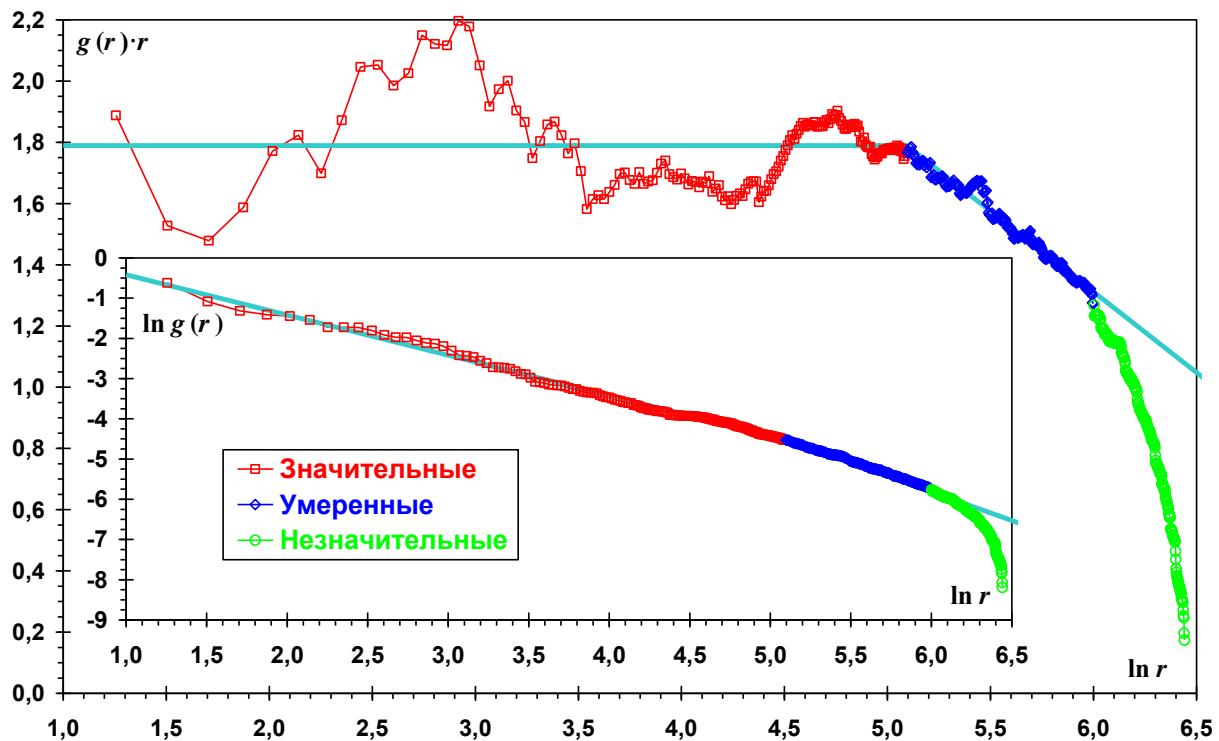


Рис. 5. Классификаций фальсификаций по жадности голосования

однажды превысила 2,5%.

4. Математические модели цифрового мира

Рис. 6 демонстрирует схожесть динамики долей субъектов с незначительными фальсификациями и совсем без них, с одной стороны, и со значительными и умеренными фальсификациями – с другой. Таким образом, можно полагать, что умеренные фальсификации – это, скорее, система, чем единичные проявления, а незначительные – скорее, единичные проявления, чем система.

Таблица. 6. Наиболее жадные голосования

Субъект федерации	1999	2000	2003	2004	2007	2008	2011	2012	2016	2018
Дагестан	2,7%	8,5%	11,8%	5,5%	13,9%	24,4%	53,8%	11,2%	7,3%	3,5%
Чечня	—	—	33,9%	23,6%	0,0%	0,0%	0,4%	0,4%	0,4%	0,0%
Карачаево-Черкесия	0,0%	0,4%	0,8%	4,1%	26,8%	17,8%	1,2%	2,4%	0,4%	6,0%
Ингушетия	10,0%	0,0%	1,7%	1,7%	21,4%	0,0%	0,0%	0,8%	1,5%	0,7%
Агинский Бурятский а/о	0,0%	0,0%	0,0%	0,0%	17,9%	—	—	—	—	—
Усть-Ордынский Бур. а/о	0,0%	2,8%	0,0%	0,5%	17,8%	—	—	—	—	—
Татарстан	1,8%	2,3%	2,8%	3,6%	14,0%	16,4%	1,6%	1,1%	2,2%	2,1%
Мордовия	0,3%	3,1%	1,3%	4,3%	14,7%	8,8%	3,5%	0,7%	0,7%	4,2%
Кабардино-Балкария	0,0%	1,5%	1,2%	3,9%	12,8%	0,6%	2,0%	0,3%	0,6%	0,3%
Тыва	0,0%	0,5%	0,0%	2,2%	12,1%	6,6%	3,8%	3,8%	2,2%	6,6%
Башкортостан	1,2%	3,4%	1,0%	5,1%	8,9%	6,5%	1,5%	0,5%	0,4%	1,1%
Ямало-Ненецкий а/о	2,0%	1,0%	0,0%	1,0%	4,5%	7,8%	1,0%	2,4%	0,0%	0,4%
Северная Осетия	0,0%	1,3%	0,0%	6,3%	1,1%	1,9%	0,5%	0,3%	2,9%	3,5%
Липецкая обл.	0,0%	0,0%	0,1%	0,2%	1,1%	5,6%	0,3%	0,3%	1,0%	2,9%
Иркутская обл.	0,1%	0,9%	0,1%	0,4%	0,6%	4,6%	0,0%	0,0%	0,0%	0,3%
Чукотский а/о	0,0%	0,0%	0,0%	4,5%	1,7%	0,0%	0,0%	0,0%	0,0%	0,0%
Кемеровская обл.	0,7%	0,4%	0,5%	2,0%	3,1%	4,3%	0,7%	1,3%	2,0%	2,3%
Орловская обл.	0,5%	2,3%	2,0%	3,7%	0,9%	4,2%	0,1%	0,1%	0,0%	0,3%
Тюменская обл.	0,0%	0,7%	0,4%	2,4%	3,3%	1,9%	0,6%	0,4%	1,0%	1,2%
Бурятия	1,4%	0,3%	0,3%	1,0%	0,6%	2,0%	2,2%	1,9%	0,8%	3,2%
Адыгея	0,0%	0,0%	0,0%	2,3%	1,5%	1,5%	0,0%	0,4%	1,9%	3,0%
Таймырский а/о	0,0%	2,9%	0,0%	0,0%	—	—	—	—	—	—
Пензенская обл.	0,5%	0,6%	0,1%	0,4%	2,2%	2,7%	0,3%	0,3%	0,4%	0,3%
Якутия	0,5%	—	0,3%	0,8%	2,6%	1,6%	0,6%	0,4%	0,2%	0,5%
Белгородская обл.	0,0%	0,3%	0,3%	0,3%	1,7%	2,6%	0,6%	0,2%	0,7%	0,4%
Тамбовская обл.	0,0%	0,5%	0,0%	1,2%	1,2%	2,5%	0,4%	0,7%	0,2%	1,1%

Среди всех критериев выявления масштабных фальсификаций жадность наилучшим образом подходит для оценки их общего объема. Чтобы получить ее, субъекты сортируются по возрастанию жадности, после чего отбрасываются по одному с конца списка. Для его сохраненной части вычисляется охват избирателей, а также доля тех из них, кто проголосовал за власть. Домножение последней величины на общее число избирателей дает оценку истинного числа голосов, поданных в поддержку власти. Как видно из рис. 7, гипотетическое завышение поддержки власти практически не зависит от охвата избирателей для его значений в диапазоне 25÷50%. При меньшем охвате уже существенна недостаточная

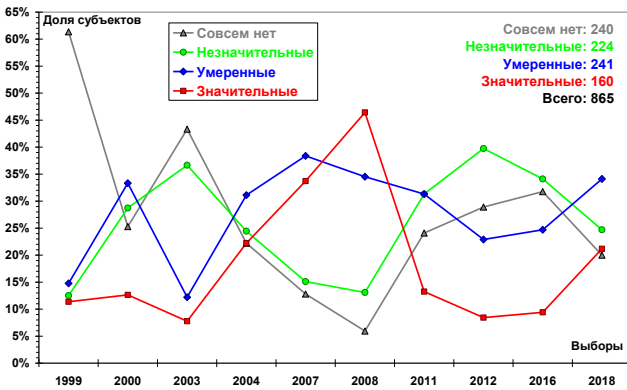


Рис. 6. Жадные фальсификации

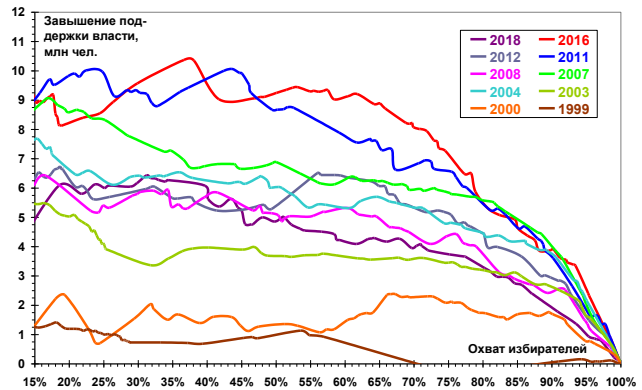


Рис. 7. Объем фальсификаций

представительность анализируемой выборки, а при большем – в ней еще сохраняются субъекты с масштабными фальсификациями.

Для выборов 1999 г. в качестве результатов партии власти здесь взяты суммарные результаты ее конкурировавших крыльев, из которых безусловным лидером по фальсификациям было движение «Отечество – Вся Россия», успешно кравшее голоса даже у «Единства», не говоря уже про прочие партии.

В 1999-2007 гг. объем выявляемых фальсификаций монотонно возрастает со временем. Однако далее эта тенденция нарушается, и парламентские выборы начинают обгонять идущие за ними президентские по объему фальсификаций. В паре 2007/08 гг. отрыв еще невелик, но в 2011/12 и 2016/18 гг. он уже становится подавляющим. Президентские выборы 2004-18 гг. застыли на объеме масштабных фальсификаций, выявляемых жадностью, около 6 млн голосов, тогда как потолок для парламентских выборов, составляющий примерно 10 млн голосов, был достигнут лишь в 2010-е гг.

Можно предположить, что естественно высокая поддержка кандидата власти оставляет меньше возможностей для подтасовок в его пользу. Кроме того, мажоритарный характер президентских выборов, сам по себе амплифицирующий фальсификации, не требует их гигантского масштаба для достижения целей фальсификаторов. С другой стороны, сравнительно невысокая поддержка партии власти в сочетании с пропорциональной системой парламентских выборов не только допускают, но и стимулируют фальсификации.

4. Недействительные бюллетени

Весьма простым и надежным показателем недостоверности результатов служит низкая доля *недействительных бюллетеней*. При фальсификациях она может лишь уменьшаться, т.к. нет политического субъекта, защищающего волеизъявление тех, кто испортил бюллетень.

4. Математические модели цифрового мира

Умышленная порча бюллетеней как форма протестного голосования начала складываться в России после 2006 г., когда была отменена графа «против всех». 2007-08 гг. стали временем адаптации избирателя к этой новации, так что здесь анализ ограничен 2011-18 гг.

Выборы разных лет различаются по силе протестных настроений вообще и по доле недействительных бюллетеней в частности. Причем имеет значение и тип выборов. Для парламентских выборов доля недействительных бюллетеней в целом выше, чем для президентских. Если для первых ее медиана, рассчитанная по всем участкам страны, учитываемым с весом, равным числу избирателей, варьируется от 1,00% до 1,95%, то для вторых – от 0,69% до 1,19%.

В силу указанной вариативности субъекты на выборах характеризуются не самой долей недействительных бюллетеней, а местом в отсортированном списке ее значений. В табл. 7 приведены данные о субъектах, хотя бы раз попадавших в верхнюю десятку из 85 субъектов (Крым и Севастополь исключены из рассмотрения для обеспечения сравнимости выборов разных лет). Для каждого выборов первая колонка – место, вторая – доля недействительных бюллетеней. В заголовке в скобках приведены ее медианные значения. Отсортированы субъекты по наивысшему занятому ими месту.

Все перечисленные в таблице субъекты хотя бы по разу (а большинство – по многу раз) проваливали описанные выше тесты. Вместе с тем, для некоторых субъектов были и выборы, на которых не удалось выявить наличие масштабных фальсификаций ни одним из этих тестов. Однако стабильно низкая доля недействительных бюллетеней в этих субъектах позволяет рассчитывать, что по мере пополнения системы тестов фальсификации будут доказаны и там.

Таблица. 7. Доля недействительных бюллетеней

Субъект федерации	2011 (1,49%)		2012 (1,13%)		2016 (1,95%)		2018 (1,06%)	
Кабардино-Балкария	1	0,04%	1	0,11%	1	0,01%	1	0,13%
Чечня	2	0,06%	2	0,14%	2	0,04%	2	0,38%
Карачаево-Черкесия	4	0,23%	3	0,22%	3	0,21%	4	0,43%
Ингушетия	6	0,85%	4	0,26%	4	0,32%	3	0,40%
Дагестан	3	0,18%	5	0,36%	8	0,53%	5	0,44%
Мордовия	5	0,48%	7	0,65%	6	0,49%	8	0,75%
Кемеровская обл.	44	1,55%	21	0,97%	5	0,46%	6	0,64%
Тыва	8	0,96%	6	0,58%	13	1,09%	7	0,71%
Татарстан	7	0,91%	8	0,73%	9	0,55%	12	0,80%
Северная Осетия	17	1,24%	20	0,97%	7	0,49%	9	0,75%
Башкортостан	9	1,01%	9	0,76%	12	0,89%	13	0,81%
Ямало-Ненецкий а/о	10	1,08%	10	0,80%	15	1,22%	10	0,77%
Тюменская обл.	12	1,19%	11	0,83%	10	0,67%	15	0,85%

Заключение

Для выявления массовых фальсификаций предложена батарея тестов, различающихся как по степени обоснованности и надежности, так и по существу анализируемых признаков и уровню избирательности. Сочетание этих тестов дает разностороннее представление о формах и масштабах фальсификаций.

Тест на преобладание круглых чисел является вполне строго обоснованным и при этом достаточно простым для использования любым желающим. Существование лишь одного подобного теста позволяет делать далеко идущие политологические выводы. Система избирательной власти, не готовая выявлять и пресекать несомненное и очевидное рисование результатов, в принципе непригодна для организации выборов, а судебная система, не способная давать оценку этой ситуации, – для разрешения конфликтов в электоральной сфере.

Тест на сгустки несколько сложнее в реализации, допускает варьирование параметров и с трудом позволяет интерпретировать пограничные ситуации. Вместе с тем он достаточно надежен, чтобы основанные на нем экспертные мнения могли быть основой для управленческих и даже судебных решений.

Тест на жадность голосования наглядно демонстрирует исключительно широкую распространенность фальсификаций и позволяет оценить снизу их общий объем. Однако качественные суждения о величине жадности опираются на полуэмпирические соображения, что снижает валидность данного теста.

Наконец, анализ доли недействительных бюллетеней показывает, что пока описаны далеко не все признаки массовых фальсификаций и еще остается простор для дальнейших исследований в этой области.

Простейшим путем здесь представляется проверка на круглые числа многомерных электоральных характеристик (вроде пар «явка–результат»), а также поиск сгустков на гистограмме с карманами адаптирующейся ширины.

Другим возможным подходом является сопоставление результатов последовательных выборов, проводимых на одних и тех же территориях. Политическая активность и симпатии избирателей меняются медленно, причем их изменения в значительной мере синхронны. Нарушение корреляции или рассогласование динамики электоральных характеристик могут служить признаками фальсификаций. Развитие этого подхода пока сдерживается сложностями формального учета изменений в электоральной географии регионов.

Третьим перспективным направлением может стать анализ ширин распределений электоральных характеристик. При ограниченных фальсификациях, затрагивающих лишь часть участков, ширина распределения неизбежно возрастает, а при тотальных, затрагивающих

4. Математические модели цифрового мира

практически все участки, она обычно уменьшается. Поэтому основная трудность тут – в необходимости формализации априорной оценки масштаба фальсификаций в рассматриваемом регионе.

Как бы то ни было, прогресс, уже достигнутый в деле выявления фальсификаций, позволяет надеяться на успех и анонсированных методов.

Работа выполнена при поддержке РФФИ (проект 18-01-00619).