



А.В.Подлазов

**Выделение основного кластера на
диаграмме рассеяния электоральных
данных**

Рекомендуемая форма библиографической ссылки

Подлазов А.В. Выделение основного кластера на диаграмме рассеяния электоральных данных // Проектирование будущего. Проблемы цифровой реальности: труды 5-й Международной конференции (3-4 февраля 2022 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2022. — С. 193-204. — <https://keldysh.ru/future/2022/17.pdf>
<https://doi.org/10.20948/future-2022-17>

Размещено также видео выступления

Выделение основного кластера на диаграмме рассеяния электоральных данных

А.В. Подлазов

Институт прикладной математики им. М.В. Келдыша РАН

Аннотация. Описан и формализован метод реконструкции фальсифицированных итогов выборов, основанный на анализе диаграммы рассеяния избирательных участков в пространстве электоральных характеристик. Результаты применения данного метода сравниваются с реконструкцией, основанной на анализе распределения избирателей по электоральным характеристикам. Это позволяет не только выявить и исключить специфические ситуации, когда один из методов оказывается неэффективен, но и получить независимое подтверждение предположений, лежащих в основе каждого из них.

Ключевые слова: выборы, фальсификации, электоральная статистика, большие данные, реконструкция, обратные задачи, диаграмма рассеяния, основной кластер

Identification of the main cluster in the scatter diagram of electoral data

A.V. Podlazov

RAS Keldysh Institute of Applied Mathematics

Abstract. I describe and formalize a method for reconstructing falsified election results based on the analysis of a scatter diagram of polling stations in the space of electoral characteristics. I compare the results of applying this method with the reconstruction based on the analysis of the distribution of voters by electoral characteristics. This allows not only to identify and eliminate specific situations where one of the methods is ineffective, but also to obtain independent confirmation of the assumptions underlying each of them.

Keywords: elections, fraud, electoral statistics, big data, reconstruction, inverse problems, scatterplot, main cluster

1. Постановка задачи

Для России XXI в. характерно уникальное сочетание массовых фальсификаций итогов выборов с доступностью детализированных официаль-

ных данных об этих итогах. В настоящее время электоральная статистика достигла того уровня развития, когда наличие некоторых видов фальсификаций может быть строго доказано [1,2,3,4,5,6,7,8,9]. Однако успехи в реконструкции истинных итогов пока не столь велики.

В работе [10] автором был предложен и апробирован интегральный метод Шпилькина – формальный подход к реконструкции, опирающийся на гипотезу о взаимной независимости общей явки и результата партии/кандидата власти. Восстановление истинных значений этих величин выполняется единообразно на основе простой математической процедуры, не использующей дополнительных предположений и свободных параметров. Этот инструментарий возник как развитие классического подхода [11,12,13,14], связанного с анализом распределения избирателей по их политической активности (поддержка власти или оппозиции, участие или неучастие в голосовании).

Вместе с тем, из работ [11,12,13,14] известен принципиально иной способ реконструкции, использующий диаграмму рассеяния избирательных участков в координатах «явка – результат». Основной кластер точек на такой диаграмме (см. рис. 1 далее) соответствует избирательным участкам с честно подведенными итогами, а тянущийся вправо и вверх/вниз (для результата власти/оппозиции) хвост – с фальсифицированными. Положение центра основного кластера дает независимую оценку истинных величин общей явки и результата участника выборов.

Критическим недостатком данного подхода являлось наличие свободного параметра, задающего долю точек в искомом кластере. В упомянутых работах она произвольно полагалась равной половине. Однако выбор размера кластера может весьма заметно влиять на положение его центра (см. рис. 2 далее), в силу чего подобный произвол порождает закономерные сомнения в надежности выполненных реконструкций.

Далее явным образом описывается методика выделения основного кластера и формулируется критерий для определения оптимальной доли точек в нем.

2. Описание метода

Хотя весь анализ в настоящей работе выполняется для двумерной постановки в координатах «явка – результат», он без потери общности распространяется и на случай большего числа электоральных характеристик. Поэтому метод здесь описывается для произвольной размерности.

Итак, пусть имеется p точек $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$ в d -мерном пространстве, из которых, как ожидается, некоторые $q \leq p$ точек достоверны, а остальные считаются аутлаерами, которые необходимо исключить из рассмотрения. Среднее положение достоверных точек $\mathbf{m} = \langle \mathbf{x} \rangle$, как уже было сказано, и будет результатом реконструкции.

Для выявления достоверных точек выполняется минимизация детерминанта ковариационной матрицы [15], опирающаяся на понятие о расстоянии Махаланобиса. Оно служит мерой удаленности между векторами случайных величин, учитывающей корреляции между ними. Для точки \mathbf{x} расстояние $h(\mathbf{x}) = (\xi^T \cdot S^{-1} \cdot \xi)^{1/2}$, где $\xi = \mathbf{x} - \mathbf{m}$ – центрированное положение точки, а $S = \langle \xi \cdot \xi^T \rangle$ – ковариационная матрица множества достоверных точек. Очевидно, что введенная таким образом метрика инварианта к линейным преобразованиям пространства. И выбрав преобразование, превращающее ковариационную матрицу в единичную, мы немедленно получаем простую нормировку $\langle h^2(\mathbf{x}) \rangle = d$, которая используется далее.

Детерминант ковариационной матрицы $|S|$ минимизируется посредством следующего итерационного алгоритма. В начале очередного его шага каждая точка имеет какую-то оценку расстояния Махаланобиса до центра кластера (для первого шага – случайную). Далее из всего массива точек выбираются q точек, характеризующихся наименьшими значениями $h(\mathbf{x})$. И для множества этих точек рассчитываются среднее положение \mathbf{m} и ковариационная матрица S , на основе которых получают новые оценки расстояний. После чего шаг повторяется – до тех пор, пока величина $|S|$ не перестанет убывать.

Результат работы алгоритма может зависеть от выбора случайных расстояний, использовавшихся на первом шаге. Поэтому вся описанная процедура выполняется многократно (обычно достаточно нескольких десятков попыток) и сохраняется результат с наименьшим $|S|$.

Описанный алгоритм позволяет на выходе получить эллипсоид наименьшего объема $V \sim |S|^{1/2}$, содержащий q достоверных точек. Соответственно, на входе алгоритм предполагает эллиптически симметричное унимодальное распределение точек с плотностью $u(\mathbf{x}) = g(h^2(\mathbf{x})) / V$, где g – некоторая монотонно убывающая функция (в частности, если g – экспонента, то распределение $u(\mathbf{x})$ – гауссово). Следует отметить, что унимодальность является в некотором смысле более сильным условием, чем симметричность, поскольку вблизи моды почти любое распределение является приблизительно симметричным.

Симметричность распределения точек в выделенном кластере требует проверки, которая и положена в основу критерия для определения оптимального числа достоверных точек. Асимметрию их множества уместно характеризовать, сравнивая взвешенное среднее $\mathbf{w} = \langle h^2(\mathbf{x}) \cdot \mathbf{x} \rangle / d$ с обычным средним $\mathbf{m} = \langle \mathbf{x} \rangle$. Для симметричного распределения эти вектора должны совпадать, поэтому при реконструкции следует использовать значение параметра $q^* = \arg \min \{h(\mathbf{w})\}$.

Электоральные фальсификации носят в той или иной мере организованный характер, в силу чего на диаграмме рассеяния могут присутствовать небольшие плотные кластеры точек, соответствующих участкам с фальсифицированными итогами выборов. В результате при $q \ll p$ возмож-

но возникновение ситуации, когда алгоритм находит центр кластера фальсифицированных, а не достоверных точек. И поскольку невозможно гарантировать, что такой кластер не окажется более симметричным, чем кластер достоверных точек, имеется риск ошибочной реконструкции. Чтобы его исключить, применяется следующая процедура.

Для каждой рассматриваемой электоральной характеристики x_i , которая увеличивается (уменьшается) в результате фальсификаций, имеется такое значение q_i , при котором ее реконструированное значение оказывается минимально (максимально). Требуется выполнение дополнительного условия $q^* \geq \max\{q_i\}$. Иначе говоря, если добавление точек в кластер приводит к более агрессивной (указывающей на более масштабные фальсификации) реконструкции хотя бы одной электоральной характеристики, значит, мы имеем дело с кластером фальсифицированных точек и параметр q необходимо увеличить.

Далее, как и в работах С.А. Шпилькина, рассматривается только частный случай с $d = 2$. При этом компонентами вектора $\mathbf{x} = (t, r)^T$ являются общая явка избирателей t и результат партии/кандидата власти r , увеличивающиеся в результате фальсификаций. В двумерной постановке матрица S легко обращается аналитически. Для расстояния Махаланобиса получаем простую формулу $h^2(\mathbf{x}) \cdot |S| = D_r \cdot \tau^2 - 2C \cdot \tau\rho + D_t \cdot \rho^2$, где $\tau = t - \langle t \rangle$ и $\rho = r - \langle r \rangle$ – центрированные значения координат, $D_t = \langle \tau^2 \rangle$, $D_r = \langle \rho^2 \rangle$ и $C = \langle \tau\rho \rangle$ – их дисперсии и ковариация, а $|S| = D_t \cdot D_r - C^2$ – детерминант ковариационной матрицы, подлежащий минимизации.

3. Примеры реконструкции

Как и в предыдущих исследованиях [6,7,8,9,10], автор здесь ограничивает анализ голосованиями только федерального уровня, проводившимися в XXI в., которых на настоящий момент набралось уже 11 штук: президентские выборы 2000, 2004, 2008, 2012 и 2018 гг., парламентские выборы по федеральному округу 2003, 2007, 2011, 2016 и 2021 гг., а также плебисцит по конституционным поправкам 2020 г.

Рис. 1 демонстрирует типичный вид диаграмм рассеяния избирательных участков с выделенным основным кластером. Вероятно, он содержит не все достоверные участки, а лишь те из них, которые расположены наиболее плотным и симметричным образом. Многие точки, образующие ореол основного кластера, тоже могут быть достоверными. Однако расположены эти точки несимметрично, в силу чего не учитываются при реконструкции. Наконец, хвост диаграммы, простирающийся до ее правого верхнего угла, представляет участки, итоги выборов на которых, скорее всего, были сфальсифицированы.

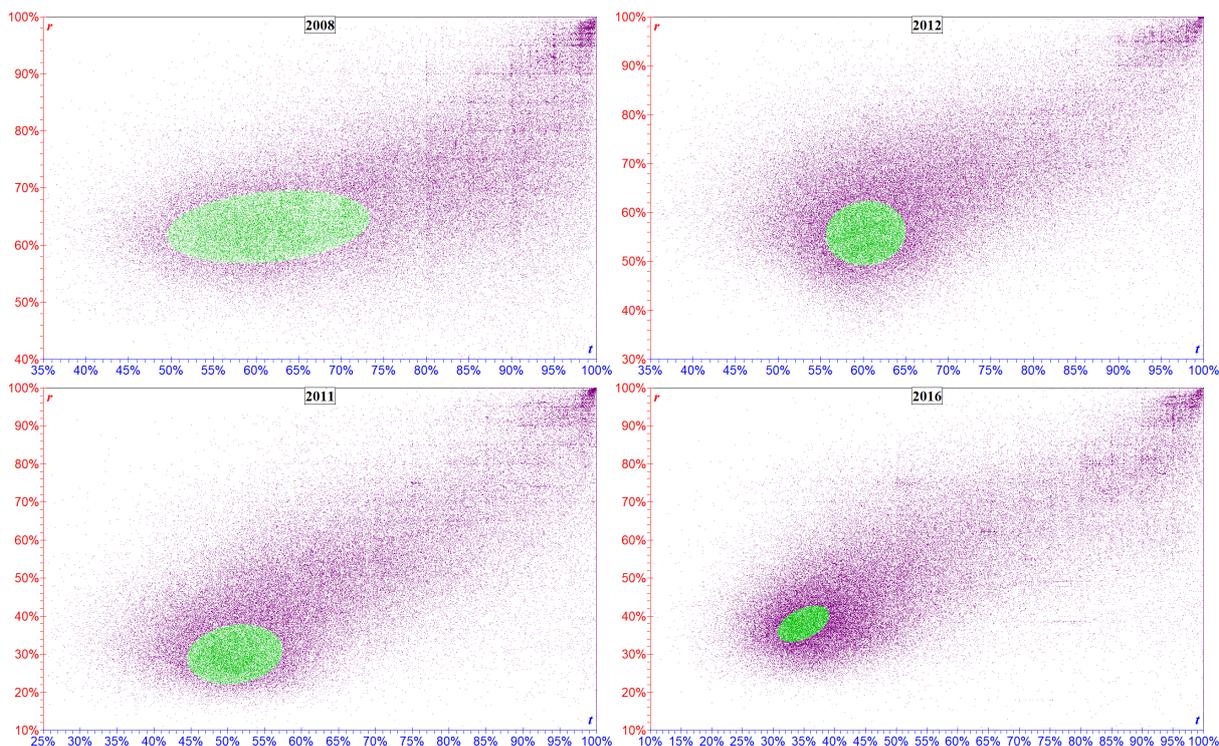


Рис. 1. Диаграммы рассеяния в координатах «явка – результат» с выделенным основным кластером для президентских выборов 2008 и 2012 гг. и парламентских выборов 2011 и 2016 гг.

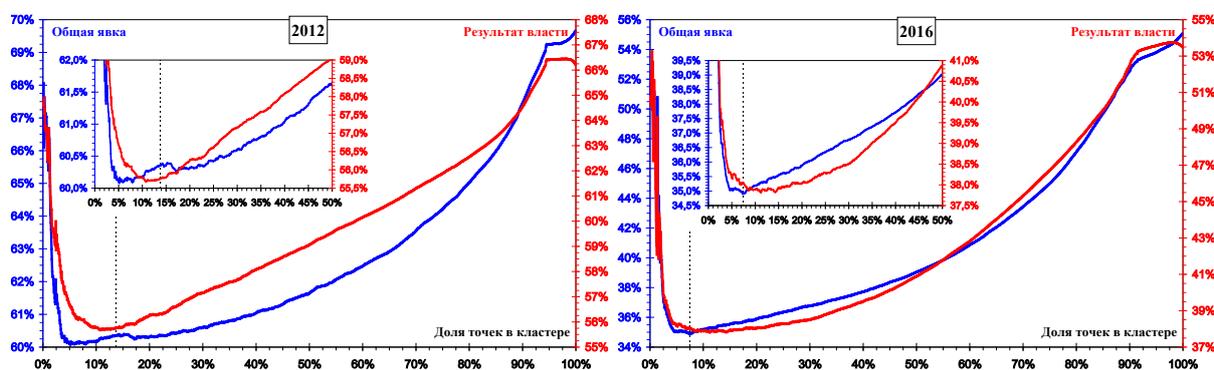


Рис. 2. Зависимость положения центра основного кластера диаграммы «явка – результат» от доли входящих в него точек для выборов 2012 и 2016 гг.

Для выполненных здесь реконструкций доля q/p лежит в диапазоне $7 \div 21\%$. Примеры зависимости от нее значений реконструируемых характеристик приведены на рис. 2 (вертикальным пунктиром показана величина q/p). Все графики оказываются в целом выпуклыми вниз. Их левые части соответствуют кластеризации тех участков, где итоги голосования фальсифицировались к целевым показателям, определявшимся централизованно. Поскольку разброс точек в целевых кластерах обуславливается лишь ошибками округления, такие кластеры всегда плотнее кластера достоверных участков. Однако организованные фальсификации затрагивают

сравнительно небольшое число участков, поэтому, начиная с некоторого их количества, наиболее плотный кластер отслеживает уже достоверные участки. При этом общая явка и результат власти, определяемые положением его центра, могут опускаться ниже истинных значений этих характеристик. Дальнейшее добавление точек в кластер приводит к учету и фальсифицированных участков, из-за чего реконструируемые величины вновь начинают возрастать.

Иная картина возникает в случае очень масштабных фальсификаций, которые имели место на плебисците 2020 г. Для него графики оказываются выпуклыми вверх, поскольку при любом отношении q/p основной кластер отслеживает не достоверные, а фальсифицированные участки, т.к. их слишком много. В этом случае результаты работы алгоритма бессмысленны, однако его можно форсировать, искусственно исключив из рассмотрения как заведомо недостоверные все участки с очень большими значениями явки и результата. Критерием того, что исключено достаточное количество участков, служит выгибание графика вниз. Точные значения порогов отсека недостоверных участков в настоящей работе не устанавливались. Для указанного голосования достаточно не принимать во внимание участки, где и общая явка, и результат власти превысили 75%.

4. Сравнение реконструкций

Все реконструированные значения представлены на рис. 4. Осветленной заливкой столбиков выделены те реконструкции, которые при дальнейшем анализе будут признаны ненадежными.

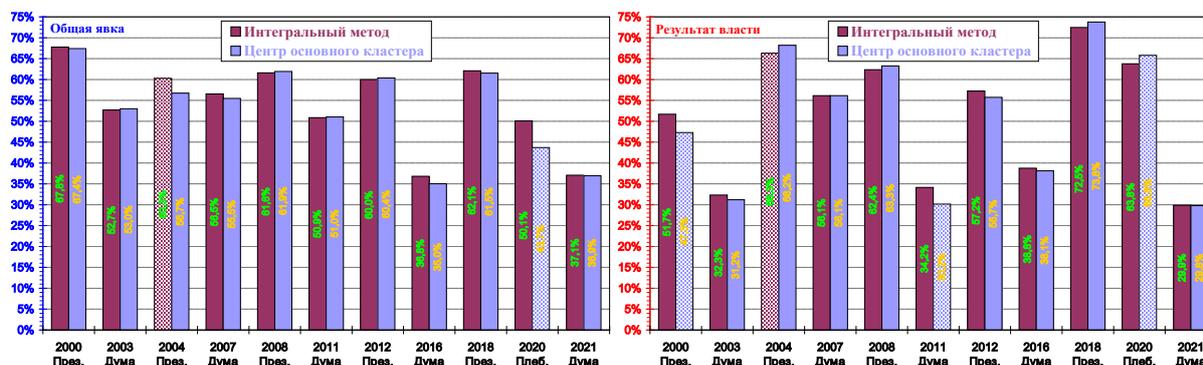


Рис. 4. Сравнение реконструкций истинных значений общей явки и результата власти, полученных с помощью интегрального метода Шпилькина и путем выделения основного кластера диаграммы рассеяния

Рис. 5 показывает, насколько реконструкция с помощью диаграммы рассеяния участков оказывается агрессивней реконструкции с помощью анализа распределения избирателей. Здесь добавлена третья электоральная характеристика – явка за власть. Она с точностью до поправки на унесенные избирателями бюллетени совпадает с произведением общей явки на

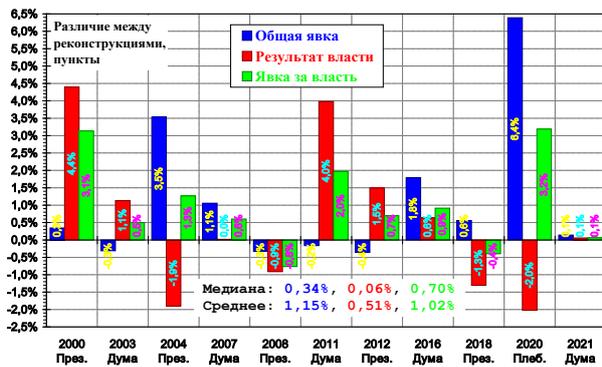


Рис. 5. Разность между результатами реконструкций, выполненных в поиске центра основного кластера и с помощью интегрального метода

выходит сразу за 3% пункта. Рассмотрим эти случаи подробно.

Президентские выборы 2000 г. пришлись на момент перестройки российской политической системы. Еще не минуло и 100 дней с парламентских выборов 1999 г., на которых две ветви будущей партии власти организовывали фальсификации каждая в свою пользу в регионах, контролируемых их представителями. А в каких-то регионах оставалась в силе и основная оппозиционная партия, в пользу лидера которой там фальсифицировались результаты президентских выборов 1996 г. [10]. Из-за неравновесности процесса перестройки распределение кандидатов по поддержке избирателей в 2000 г. не было унимодальным (см. рис. 6 вверху слева). В результате центр основного кластера, отслеживающий самую сильную моду распределения, занижает результат власти, правильную реконструкцию которого дает интегральный метод.

К 2004 г. президентские выборы уже стали неконкурентными, т.е. насущной необходимости фальсифицировать результат власти больше не было. Однако еще сохранялся 50%-порог явки, преодоление которого требовалось для того, чтобы выборы были признаны состоявшимися. И к перетаскиванию явки через этот порог фальсификаторы прикладывали значительные усилия, следы которых можно без труда увидеть на рис. 6 вверху справа. При таком экзотическом целеполагании фальсификации явки могут существенно преобладать над фальсификациями результата, из-за чего реконструкция явки, основанная на анализе распределения избирателей по политической активности, оказывается ненадежной. И в данном случае следует доверять положению центра основного кластера диаграммы рассеяния.

Чем более лояльны по отношению к власти избиратели конкретного участка, тем на нем, как правило, хуже контроль за чистотой проведения голосования и выше готовность членов участковой комиссии фальсифицировать его итоги. Поэтому и общая явка, и результат власти чаще завыша-

результат власти и, как правило, более стабильна. Но, вместе с тем, для нее существенно выше медианное значение разницы, т.е. анализ рассеяния участков дает в целом более агрессивную реконструкцию по сравнению с анализом распределения избирателей.

Сразу бросается в глаза то, что для каждой из рассмотренных электоральных характеристик в 9 случаях различие между реконструкциями не превышает 2% пункта, зато в оставшихся 2 случаях оно

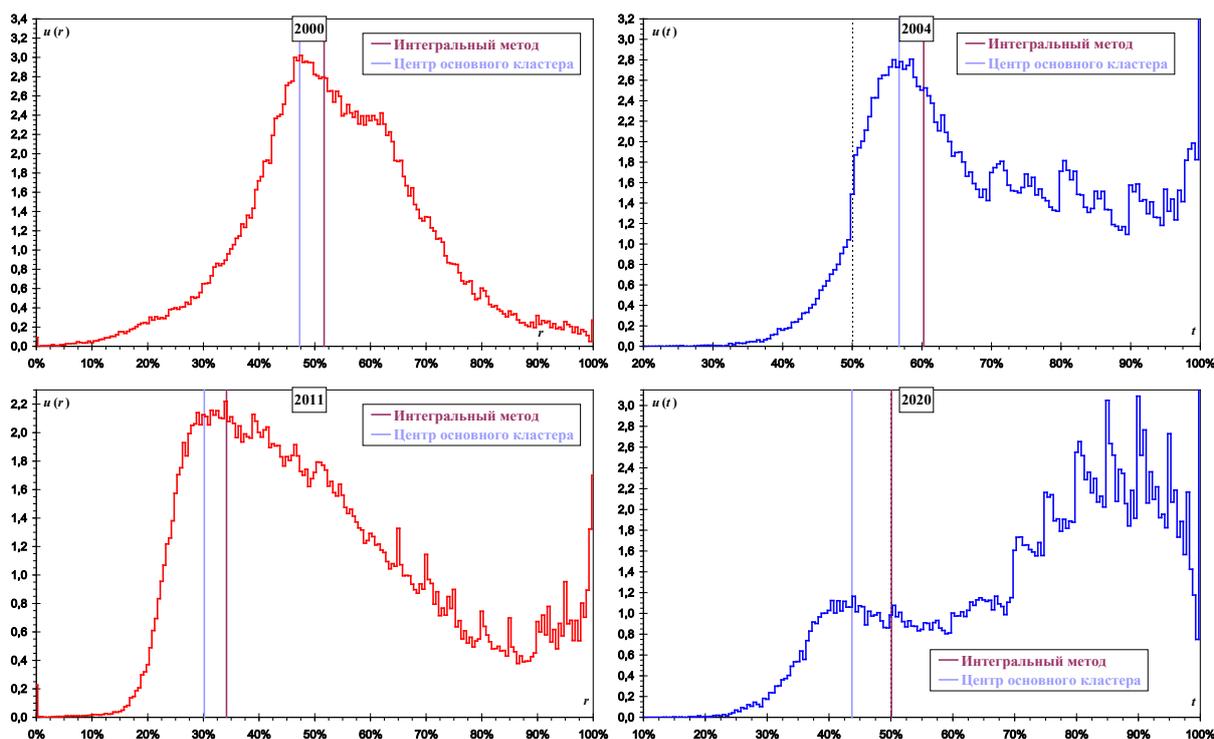


Рис. 6. Гистограммы распределения избирателей по результату власти на президентских выборах 2000 и парламентских выборах 2011 гг. и по общей явке на президентских выборах 2004 и плебисците 2020 гг.

ются там, где они и без того высоки. Но если фальсификации очень масштабны и затрагивают даже точки, находящиеся вблизи центра основного кластера диаграммы, то возникает опасность преимущественной потери им точек, лежащих немного правее и/или выше его центра, из-за чего тот обнаруживается алгоритмом левее и/или ниже своего истинного положения, т.е. реконструкция оказывается чрезмерно агрессивной. При этом доверять следует реконструкции, выполненной с помощью интегрального метода, использующего более надежные данные по участкам с низкими значениями электоральных характеристик.

Набор рассматриваемых голосований содержит лишь два примера подобного рода: парламентские выборы 2011 г. в части результата партии власти (рис. 6 внизу слева) и плебисцит 2020 г. в части явки (рис. 6 внизу справа). Несмотря на огромный масштаб фальсификаций на этих голосованиях описанный эффект затронул лишь по одной из двух электоральных характеристик – те, которые были критичны для власти и на которых, по видимому, было сосредоточено основное внимание организаторов фальсификаций. В 2011 г. чисто пропорциональная система формирования нижней палаты парламента уже не позволяла партии власти сохранить конституционное большинство, поэтому усилия по хотя бы максимизации своего представительства оказались для нее принципиально важными. А в 2020 г. из-за отсутствия политического актора, который мог проводить

агитацию против поправок, не было требовалось добиваться особенно высокого результата – он и без того был гарантирован. Зато власть тогда остро нуждалась в высокой явке, которая бы хоть как-то могла легитимировать итоги юридически невнятного голосования.

По сравнению с интегральным методом диаграммный метод в 2004 и 2020 гг. одновременно оказывается наиболее агрессивен при реконструкции общей явки и наиболее мягок при реконструкции результата власти. Это совпадение не случайно и связано с завышением числа активных избирателей, входящим и в числитель явки, и в знаменатель результата. И если мы выяснили, что в первом случае следует доверять реконструкции явки, полученной диаграммным методом, а во втором – интегральным, аналогичным образом приходится поступить и с реконструкцией результата, на которую может существенно влиять ошибочная реконструкция явки.

Таким образом, из 22 значений электоральных характеристик (11 пар величин общей явки и результата власти) для 6 надежной оказывается реконструкция только с помощью одного из двух методов, который выбирается на основе детального анализа. Именно эта реконструкция считается окончательной. Для остальных 16 значений оба метода дают надежные (и весьма близкие) реконструкции, так что в качестве окончательной реконструкции уместно использовать их среднее.

Разность и отношение официальных значений электоральных характеристик и их окончательных реконструкций показаны на рис. 7. Он позволяет подтвердить все выводы, сделанные при проведении политологического анализа в работе [10], распространив их на парламентские выборы 2021 г. Применение на них нового инструментария фальсификаций, апробированного ранее на плебисците 2020 г., позволило власти превзойти прежние максимумы как для разности между официальным и истинным ее результатом, так и для той доли в официальных значениях всех трех рассматриваемых электоральных характеристик, которая обусловлена фальсификациями.

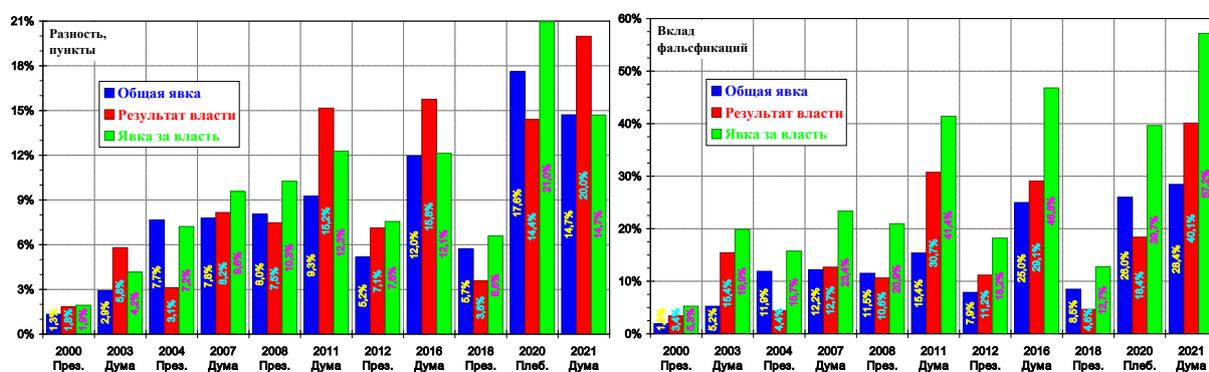


Рис. 7. Превышение официальными итогами выборов реконструированных и доля величины официальных итогов, приходящая на фальсификации

5. Сравнение методов и их синергия

Наличие двух методов реконструкции, основанных на независимых предположениях, имеет принципиальное значение как с точки зрения противопоставления результатов реконструкции, так и с точки зрения их сопоставления.

Немногочисленные расхождения результатов, каждое из которых получает политологическое объяснение, дают представление об области применимости каждого из методов, об их сильных и слабых сторонах.

Основным преимуществом интегрального метода является его простота как в части описания, так и в части компьютерной реализации. Этот метод не требователен к вычислительным ресурсам и может применяться даже на уровне электронных таблиц, т.е. без написания кода. А вот алгоритм, ищущий центр основного кластера диаграммы, уже предполагает программную реализацию и очень ресурсоемок. Без глубокой оптимизации он может потребовать до нескольких суток машинного времени на обработку данных выборов федерального уровня с $p \sim 10^5$.

Анализ распределения избирателей обладает большей устойчивостью по отношению к очень сильным фальсификациям, ориентированным на критически важные для власти показатели, такие как результат партии власти в 2011 и общая явка в 2020 гг. Отчасти это компенсируется тем, что анализ диаграммы рассеяния участков оказывается устойчив к экзотическим фальсификациям, таким как преодоление порога явки, осуществлявшееся в 2004 г.

Кроме того, диаграммный метод обладает и рядом преимуществ. Во-первых, он не использует данных о размере участка, что существенно, поскольку фальсификация итогов голосования осуществлялась на уровне участка, а вовсе не на уровне отдельных избирателей. Во-вторых, в основной кластер не попадают экстремальные участки не только с большими общей явкой или результатом власти, но и с малыми, т.е. метод исключает из рассмотрения и те участки, где из-за ошибок или фронды фальсификаторов лишние голоса были приписаны оппозиции. В-третьих, анализ диаграммы рассеяния потенциально расширяем на случай любого числа электоральных характеристик, т.е. реконструкция может быть уточнена за счет учета и результатов оппозиции.

Имеющая место в большинстве случаев близость результатов, которые дают оба метода реконструкции, свидетельствует о применимости исходных предположений каждого из них к российской политической системе. И если предположение об эллиптически симметричном унимодальном виде распределения участков по электоральным характеристикам в целом не вызывает возражений, то предположение о взаимной независимости общей явки и результата власти до сего момента оставалось спорным. Теперь не только его следствия получили независимое подтверждение, но и

само это предположение может быть косвенным образом проверено на основе анализа диаграмм рассеяния.

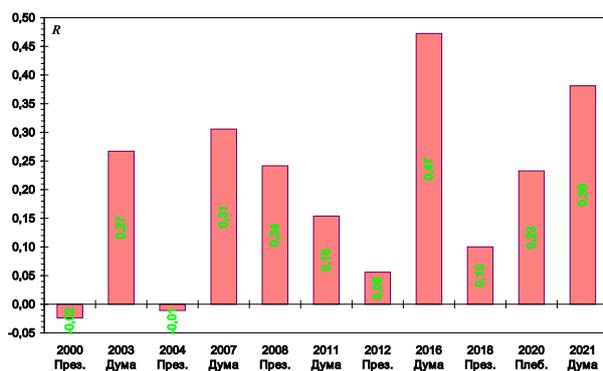


Рис. 8. Коэффициент корреляции между общей явкой и результатом власти

Взаимозависимость общей явки и результата власти характеризуется коэффициентом корреляции $R = C / (D_i \cdot D_r)^{1/2}$, вычисляемым для точек основного кластера. Как можно видеть из рис. 8, в целом корреляция не очень велика (среднее значение R^2 составляет чуть более 6%). Однако она, безусловно, положительна, что допускает две трактовки. С одной стороны, некоторое увеличение результата власти при росте общей явки может быть реальным феноменом, обусловленным зависимостью между полити-

ческими активностью и симпатиями избирателей. Однако в этом случае непонятно, почему эта гипотетическая зависимость совершенно по-разному проявляет себя для разных выборов. С другой стороны, корреляция может объясняться попаданием в основной кластер некоторого количества точек, соответствующих фальсифицированным участкам, где общая явка и результат власти завышаются одновременно. В этом случае коэффициент корреляции будет тем больше, чем сильнее фальсифицированы выборы. И в самом деле, вклад фальсификаций в официальные итоги выборов (см. рис. 7 справа) примерно наполовину (47% для общей явки, 52% для результата власти и 56% для явки за власть) детерминируется величиной R . Таким образом, предпочтение следует отдать второй трактовке природы корреляции.

Представляется, что возможность что-то узнать о базовых свойствах политической системы, анализируя результаты реконструкции фальсифицированных данных о ней, является важнейшим результатом выполненного анализа. По значимости он превосходит реконструкцию как таковую. Ведь одно дело просто выбраться из болота, а совсем другое – вытащить себя оттуда за волосы.

Работа выполнена при поддержке РФФИ (проект 20-511-00003).

Литература

1. Kobak D., Shpilkin S., Pshenichnikov M.S. Statistical anomalies in 2011–2012 Russian elections revealed by 2D correlation analysis. [arXiv:1205.0741](https://arxiv.org/abs/1205.0741)
2. Kobak D., Shpilkin S., Pshenichnikov M.S. Integer percentages as electoral falsification fingerprints // [Ann. Appl. Stat. 2016, 10\(1\), 54-73](https://doi.org/10.1080/10673895.2016.1191733).

3. *Kobak D., Shpilkin S., Pshenichnikov M.S.* Statistical fingerprints of electoral fraud? // [Significance. 2016, 13\(4\), 20-23.](#)
4. *Kobak D., Shpilkin S., Pshenichnikov M.S.* Putin's peaks: Russian election data revisited // [Significance. 2018, 8\(3\), 8-9.](#)
5. *Kobak D., Shpilkin S., Pshenichnikov M.S.* Suspect peaks in Russia's "referendum" results // [Significance. 2020, 17\(5\), 8-9.](#)
6. *Подлазов А.В.* Исследование статистических методов выявления вымышленных результатов выборов: Часть 1. Круглые числа // [Препринт ИПМ им. М.В. Келдыша РАН. 2019, №147.](#)
7. *Подлазов А.В.* Исследование статистических методов выявления вымышленных результатов выборов. Часть 2. Стустки явки и результатов // [Препринт ИПМ им. М.В. Келдыша РАН. 2020, №20.](#)
8. *Подлазов А.В.* Формальные методы выявления масштабных электро-ральных фальсификаций // [Проектирование будущего. Проблемы цифровой реальности. 2019. Вып.2, 120-137.](#)
9. *Подлазов А.В.* Формальное выявление вымышленных результатов выборов // [Проектирование будущего. Проблемы цифровой реальности. 2020. Вып.3, 176-190.](#)
10. *Подлазов А.В.* Реконструкция фальсифицированных результатов выборов с помощью интегрального метода Шпилькина // [Проектирование будущего. Проблемы цифровой реальности. 2021. Вып.4, 193-208.](#)
11. *Шпилькин С.* Статистическое исследование результатов российских выборов 2007–2009 гг. // [Троицкий вариант – наука. 2009, 21\(40\), 2-4.](#)
12. *Шпилькин С.* Математика выборов – 2011 // [Троицкий вариант – наука. 2011, 25\(94\), 2-4.](#)
13. *Шпилькин С.* Выборы 2018 года: Фактор X и «пила Чурова» // [Троицкий вариант – наука. 2018, 8\(252\), 8-10.](#)
14. *Шпилькин С.* Поправки на 27 миллионов // [Троицкий вариант – наука. 2020, 14\(308\), 4-5](#)
15. *Hubert M., Debruyne M.* Minimum covariance determinant // [WIREs Comput Stat. 2010, 2\(1\), 36-43.](#)