

СТАТИСТИЧЕСКАЯ ОЦЕНКА ЭФФЕКТОВ ПОЛИТИЧЕСКИХ РЕШЕНИЙ И СОБЫТИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДА “РАЗНОСТЬ РАЗНОСТЕЙ”¹

Беленков В.Е., Конча В.Л., Седашов Е.А.

Национальный Исследовательский Университет «Высшая Школа Экономики»,
Москва, Россия

vbelenkov@hse.ru, vkoncha@hse.ru, esedashov@hse.ru

Аннотация. Статистический анализ эффектов политических решений и событий – одна из актуальных методологических проблем современной политической науки. В силу невозможности рандомизации исследователи часто прибегают к квази-экспериментальным исследовательским дизайнам. В данной работе мы анализируем возможности применения метода “разность разностей” в данном контексте.

Ключевые слова: метод разность разностей, каузальный анализ, эффекты вмешательства, естественные эксперименты.

Введение

Анализ влияния политических решений и событий на общественные процессы является одной из центральных проблем современной политической науки и социальных наук в целом. В то же время, методологическая дискуссия вокруг возможности получения надёжных статистических оценок этого влияния активно ведётся по сей день. Зачастую требования, которые тот или иной метод предъявляет к данным, невозможно проверить, и исследователям приходится полагаться на логику и здравый смысл, обосновывая применимость того или иного подхода к конкретной задаче.

Фокусом данной статьи является применение метода “разность разностей” для статистической оценки эффектов политических решений и событий. Данный метод является популярным способом получения выводов о наличии или отсутствии причинно-следственных связей между переменными. [1] использовали данный метод для получения квази-экспериментальных оценок эффекта повышения МРОТ на занятость, тем самым проложив дорогу для других исследований со схожим дизайном. Так, исследовательские дизайны, основанные на методе “разность разностей”, применялись для анализа эффектов массовой стрельбы в школах на результаты выборов [2], помощи в целях развития (developmental aid) на насилие со стороны повстанческих группировок [3], отключений Интернета на протестную мобилизацию [4], отмены губернаторских и мэрских выборов в России на экономические и социальные показатели [5, 6]. В основе данных исследовательских дизайнов лежит достаточно простой набор идей. Во-первых, предполагается существование двух временных периодов $t = 0$ и $t = 1$, в промежутке между которыми происходит некоторое событие. Это событие может быть внешним шоком (например, стихийным бедствием или упоминавшейся выше массовой стрельбой в школах) или же контролируемым воздействием (повышение МРОТ, отключение Интернета, отмена выборов губернаторов и мэров). Во-вторых, событие должно оказывать влияние только на одну из двух исследуемых групп. Например, [1] исследовали две группы ресторанов быстрого обслуживания, одна из которых находилась в штате Пенсильвания (в котором МРОТ не повышался между $t = 0$ и $t = 1$) и в штате Нью-Джерси (в котором МРОТ был повышен между $t = 0$ и $t = 1$). При этом крайне желательным является исключение «эффектов переливания» (spillover effects), когда воздействие на экспериментальную группу может косвенно влиять и на контрольную группу, тем самым нарушая предпосылку SUTVA (Stable Unit Treatment Value Assumption) [7]. В-третьих, должно соблюдаться предположение о параллельности трендов, в соответствии с которым тренд экспериментальной группы должен точно следовать тренду контрольной группы в отсутствие воздействия. Для исследовательского дизайна из [1] это предположение означает, что уровень занятости в Нью-Джерси, в отсутствие повышения МРОТ, должен был бы измениться на ту же величину между периодами $t = 0$ и $t = 1$, что и в штате Пенсильвания.²

Необходимо отметить, что отсутствие «эффектов переливания» и параллельность трендов сложно гарантировать даже в лучших исследовательских дизайнах. Так, в исследовании [1] предположение о параллельности трендов гарантировалось тем, что рестораны быстрого обслуживания, из которых

¹ Исследование выполнено за счет гранта Российского научного фонда № 20-18-00274, <https://rscf.ru/project/20-18-00274/>, Национальный исследовательский университет «Высшая школа экономики»

² В более строгом смысле математические ожидания изменения уровня занятости для штатов Пенсильвания и Нью-Джерси должны быть равны. При этом в качестве генеральной совокупности используются все рестораны быстрого обслуживания, лежащие на границе двух штатов.

формировалась генеральная совокупность, находились на границе между двумя штатами и, по сути, представляли собой набор идентичных заведений со схожим персоналом. При этом отсутствие «эффекта переливания» сложно формально гарантировать, потому что, теоретически, можно представить ситуацию, когда работники ресторанов в Пенсильвании будут пробовать переходить на работу в Нью-Джерси, где более высокий уровень МРОТ. В случае, если формальное квази-экспериментальное разделение групп (например, в виде географической границы) на экспериментальную и контрольную отсутствует, методологические вызовы оказываются ещё более существенными.

1. Формальное описание метода

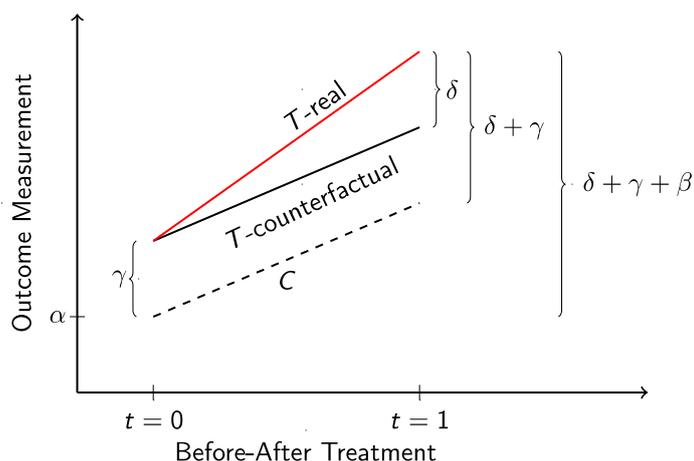


Рис. 1. Иллюстрация метода разность разностей

Даны два периода наблюдения, $t = 0$ и $t = 1$, а также две группы, которые мы обозначим как T (экспериментальная) и C (контрольная). Между указанными периодами происходит некоторое событие (внешний шок или сознательное воздействие), которое оказывает влияние на группу T и не оказывает влияния на группу C . Уравнение МНК-регрессии для генеральной совокупности можно записать следующим образом:

$$y_{kt} = \alpha + \beta * Period_t + \gamma * Group_k + \delta * Period_t * Group_k + \epsilon_{kt} \quad (1)$$

где y_{kt} – значение зависимой переменной для испытуемого k в период t , $Period_t = 0$ если $t=0$ и $Period_t = 1$, если $t=1$, $Group_k = 0$, если $k \in C$ (испытуемый k в контрольной группе) и 1 в противном случае, ϵ_{kt} – ошибка. Глядя на данное уравнение и Рисунок 1, несложно установить, что δ – оценка эффекта воздействия при соблюдении предположения о параллельности трендов (линия T -counterfactual на рисунке). Заметим, что все наблюдения, соответствующие T -counterfactual в момент $t = 1$, описываются уравнением: $\alpha + \beta + \gamma + \epsilon_{kt}$, в то время как наблюдения, соответствующие линии T в момент $t = 1$, описываются уравнением $\alpha + \beta + \gamma + \delta + \epsilon_{kt}$. Беря во внимание предположение о математическом ожидании ошибки $E[\epsilon_{kt}] = 0$, мы получаем δ – оценку каузального эффекта воздействия как разницы между T и T -counterfactual в момент времени $t=1$. Заметим, что, несмотря на существенные методологические прорывы последних лет³, описанный здесь исследовательский дизайн по-прежнему остаётся наиболее популярным способом оценки эффектов политических решений и событий, прежде всего в силу своей интуитивной простоты.

2. Проблемы метода “разность разностей” и способы их решения

Как указывалось выше, метод “разность разностей” опирается на две ключевых идеи: отсутствие «эффектов переливания» и параллельность трендов. В силу отсутствия возможности прямой проверки исследователи нередко полагаются на здравый смысл и логику для обоснования параллельности

³ Прежде всего, речь идёт о моделях “разность разностей” с вмешательствами одного типа, но осуществленными в разное время [8, 9]. Ключевым предположением здесь является *remain treated*, то есть объект наблюдения, получив воздействие, рассматривается в качестве члена экспериментальной группы в течение всего оставшегося анализируемого периода.

трендов (как в случае [1]) либо же используют информацию, присутствующую в собранных данных, для конструирования массива для анализа, в рамках которого соблюдаются предположения метода. Тем не менее, нередко возникают ситуации, когда в данных отсутствуют переменные, позволяющие сконструировать подобный массив. [4] предложили оригинальный выход из подобной ситуации – генерировать выборки для контрольной группы случайным образом с целью получения распределения возможных эффектов вмешательства. Этот подход не может рассматриваться как универсальный, т.к. требует наличия значительного дисбаланса в количестве наблюдений между экспериментальной и контрольной группами. Тем не менее, в некоторых случаях это действительно единственный возможный способ реализации подхода “разность разностей”. Но насколько мы можем быть уверены, что данный подход даёт надёжные и устойчивые результаты? Используя сгенерированные данные, мы попытались ответить на этот вопрос. Целью нашего анализа является сравнение результатов двух разных подходов. В рамках первого подхода мы использовали матчинг по критерию propensity score [10], а в рамках второго подхода мы случайным образом выбирали наблюдения из контрольной группы с целью получения распределения возможных эффектов вмешательства (как в [4]). Ниже мы приводим подробное описание алгоритма и результаты.

Начальной точкой является генерация 10000 экспериментальных единиц, которые будут использоваться в сравнительном анализе.⁴ В качестве первого шага мы создаём три предиктора, от которых будет зависеть получение той или иной экспериментальной единицы воздействия:

$$X_1 \sim Normal(2, 3)$$

$$X_2 \sim Poisson(16)$$

$$X_3 \sim Poisson(4)$$

Само воздействие генерируется по следующему алгоритму:

$$prediction = 0.5 + 0.6 * X_1 - 0.25 * X_2 + 0.1 * X_3 + Normal(0,1)$$

$$treatment \sim Bern\left(\frac{1}{1 + e^{-prediction}}\right)$$

То есть воздействие (treatment) создаётся, используя генератор случайных чисел по распределению Бернулли. Вероятности p для распределения Бернулли, в свою очередь, берутся на основе заданного уравнения, значения которого вставляются в сигмоиду. Мы выбрали такой подход к генерации данных, чтобы максимально приблизить ситуацию к «плохому» для аналитика сценарию, когда распределение наблюдений в экспериментальную и контрольную группу не является случайным или квази-случайным, а зависит от некоторых характеристик данных наблюдений. Далее, мы рассматривали два кейса. В рамках первого кейса ожидаемый эффект воздействия⁵ δ зависит от переменной X_3 :

$$y \sim Poisson\left(e^{-0.05 + 0.2 * Period + 0.4 * Group + 1_{[X_3 > 5]} * 0.3 * Period * Group + 1_{[X_3 \leq 5]} * 0.1 * Period * Group}\right)$$

где 1 – индикаторная функция, фиксирующая соответствие X_3 заданному критерию.

В рамках второго кейса ожидаемый эффект воздействия δ зависит от переменных X_3 и X_2 :

$$y \sim Poisson\left(e^{-0.05 + 0.2 * Period + 0.4 * Group + T}\right)$$

$$T = 1_{[X_3 > 5, X_2 > 16]} * 0.4 * Period * Group + 1_{[X_3 > 5, X_2 \leq 16]} * 0.2 * Period * Group + 1_{[X_3 \leq 5, X_2 > 16]} * 0.3 * Period * Group + 1_{[X_3 \leq 5, X_2 \leq 16]} * 0.1 * Period * Group$$

Мы использовали данные два кейса для того, чтобы понять, зависит ли разница результатов в двух исследуемых подходах (матчинг и симуляции со случайной компоновкой контрольной группы) от поведения самого эффекта воздействия, который не однороден и зависит от значений генерирующих воздействие предикторов. Также заметим, что зависимая переменная генерируется, используя распределение Пуассона. Это соответствует большому количеству зависимых переменных, встречающихся в эмпирической литературе (см. [3, 4, 11]).

⁴ Общее количество наблюдений – 20000, потому что для каждой экспериментальной единицы доступно два временных периода.

⁵ Строго говоря, термин “эффект воздействия” здесь не совсем корректен, потому что для генерации данных используется распределение Пуассона. Тем не менее, эффект воздействия пропорционален δ , поэтому данную терминологию можно считать оправданной.

Описав алгоритм генерации данных, мы можем переходить к более подробному описанию двух исследуемых подходов. В рамках первого подхода мы используем матчинг по критерию propensity score [10] в качестве предварительной обработки данных. В данном случае мы предполагаем, что у аналитика есть доступ к переменным, от которых, предположительно, может зависеть получение экспериментальными единицами воздействия. Применительно к нашему алгоритму, это три созданных нами предиктора. После того, как для каждого наблюдения в экспериментальной группе найден наиболее подходящий матч из контрольной, мы оценили МНК-регрессию в соответствии с уравнением 1. В рамках второго подхода мы считали, что у аналитика нет доступа к предикторам, от которых зависело получение воздействия, и единственно возможный способ получения статистической оценки δ – симуляции. Алгоритм симуляций достаточно прост: экспериментальная группа является постоянной от симуляции к симуляции, в то время как контрольная группа формируется случайным образом из всего пула доступных контрольных наблюдений, при этом размер контрольной группы должен быть равен размеру экспериментальной. По итогам 1000 симуляций мы вычисляем средний эффект как оценку эффекта воздействия.

Результаты применения двух подходов к сгенерированным данным представлены в таблице 1. Мы видим, что оценки воздействия в двух подходах различаются незначительно. То, что различия фактически одинаковы для двух альтернативных алгоритмов генерации зависимой переменной позволяет сделать вывод о том, что полученные результаты являются устойчивыми. В целом, можно сделать вывод, что подход, основанный на статистических симуляциях, действительно позволяет получить надежные выводы о существовании эффекта воздействия и его направленности.

Таблица 1. Зависимость эффекта воздействия от анализируемых аналитических подходов и алгоритмов генерации зависимой переменной

Алгоритм генерации зависимой переменной/Подход	Разность разностей с матчингом	1000 симулированных контрольных групп
Эффект вмешательства зависит от X_3	0.4	0.38
Эффект вмешательства зависит от X_2 и X_3	0.52	0.55

3. Заключение

В данной статье был рассмотрен один из популярных методов получения статистических оценок эффектов политических событий и решений – метод “разность разностей”. Мы описали общую схему метода, а также его базовые предпосылки. В связке с предпосылками были также проанализированы возможные проблемы метода.

Основным вкладом статьи в текущие методологические дебаты вокруг метода “разность разностей” является сравнение двух подходов к получению оценок эффекта воздействия. В рамках первого подхода предполагается предобработка данных с использованием матчинга, после которой на получившемся массиве проводится МНК-регрессия по уравнению 1. Второй подход предполагает генерацию контрольной группы путем случайного отбора из пула доступных контрольных наблюдений, повторение процесса 1000 раз и расчёт эффекта вмешательства как среднего эффекта во всех симуляциях. Преимуществом второго подхода является отсутствие необходимости сбора данных о предикторах, генерирующих получение экспериментальными единицами воздействия. Наше исследование показывает, что два подхода дают очень близкие оценки. Получение надёжных оценок эффектов политических событий и решений, таким образом, возможно даже без доступа к информации о предикторах, влияющих на данные события и решения. Более фундаментальный вывод работы состоит в том, что анализ эффектов тех или иных политических событий или решений может рассматриваться вне привязки к анализу причин данных событий и решений.

Литература

1. Card D., Krueger A.B. Minimal Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania // American Economic Review. – 1994. – Vol. 84, N. 4. – P. 772-793.
2. Garcia-Montoya L., Arjona A., Lacombe M. Violence and Voting in the United States: How School Shootings Affect Elections // American Political Science Review. – 2022. – Vol. 116, N. 4. – P. 807-826.
3. Weintraub M. Do All Good Things Go Together? Development Assistance and Insurgent Violence in Civil War // Journal of Politics. – 2016. – Vol. 78, N. 4. – P. 989-1002.
4. Sedashov E., Belenkov V., Koncha V. The Effects of Internet Shutdowns on Protest Activity in India: an Empirical Investigation // Proceedings of International Conference on Management of Large-Scale System Development, 2023.

5. *Beazer Q. H.* Political Centralization and Economic Performance: Evidence from Russia // *Journal of Politics*. – 2015. – Vol. 77, N. 1. – P. 128-145.
6. *Beazer Q.H., Reuter O.J.* Do Authoritarian Elections Help the Poor? Evidence from Russian Cities // *Journal of Politics*. – 2021. – Vol. 84, N. 1. – P. 437-454.
7. *Imbens G.W., Rubin D.B.* Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. – Cambridge University Press, 2015. – P. 1 – 625.
8. *Goodman-Bacon, A.* Difference-in-differences with Variation in Treatment Timing // *Journal of Econometrics*. – 2021. – Vol. 225, N. 2. – P. 254-277.
9. *Callaway B., Sant'Anna P.H.C.* Difference-in-differences with Multiple Time Periods // *Journal of Econometrics*. – 2021. – Vol. 225, N. 2. – P. 200-230.
10. *Rubin, D.* Matched Sampling for Causal Effects. – Cambridge University Press, 2012. – P. 1 – 489.
11. *Rydzak J.* Of Blackouts and Bandits: The Strategy and Structure of Disconnected Protest in India // SSRN. 2019.