

**В.В. УСТЮЖАНИН\***  
**АНАЛИЗ РЕДКИХ СОБЫТИЙ:  
ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ  
И АЛЬТЕРНАТИВЫ<sup>1</sup>**

*Аннотация.* Революции всегда вызывали интерес социологов и политологов, однако только в последнее десятилетие авторы начали активно использовать количественные методы для их анализа. В таких исследованиях зависимая переменная, революции или ее характеристики, часто представляет собой «редкие события». Некоторыми авторами было показано, что логистическая регрессия – самый используемый метод – дает смещенные результаты при анализе таких данных, однако единого стандарта для анализа так и не было выработано. Более того, часто авторы сталкиваются и с другими сопутствующими проблемами – панельными данными и малой выборкой. При этом надежность классической логистической регрессии при сочетании этих проблем еще не была проверена, хотя большая часть современных авторов сталкивается именно с их пересечением. В рамках настоящего исследования мы постарались оценить, как использование логистической регрессии при разных исследовательских дизайнах с редкими событиями – в случае кросс-секционных и панельных данных – влияет на смещенность и устойчивость результатов. Также мы предлагаем альтернативу – логистическую регрессию со штрафом. Мы нашли, что в случае с кросс-секционными данными сочетание малой выборки и редкого события порождает огромное смещение в классической логистической регрессии, и исследователь рискует не только не получить относительно точной оценки эффекта, но и сделать абсолют-

---

\* **Устюжанин Вадим Витальевич**, аспирант Института общественных наук, Российская академия народного хозяйства и государственной службы при Президенте РФ; младший научный сотрудник Центра изучения стабильности и рисков, Национальный исследовательский университет «Высшая школа экономики» (Москва, Россия), e-mail: [vvustiuzhanin@yandex.ru](mailto:vvustiuzhanin@yandex.ru)

<sup>1</sup> Исследование выполнено в рамках Программы фундаментальных исследований НИУ ВШЭ в 2026 г. при поддержке Российского научного фонда (проект № 24-18-00650).

ное ложное суждение о направлении связи. В свою очередь, логистическая регрессия со штрафом дает почти что несмещенные оценки при любой редкости события и даже при очень малой выборке. В случае с панельными данными классическая логистическая регрессия в ситуации малой выборки и редких событий не применима вовсе. На удивление, логистическая регрессия со штрафом решает все эти проблемы – панельные данные, малая выборка и редкость события. На наш взгляд, это самое важное наблюдение настоящей работы: логистическая регрессия со штрафом отлично справляется с панельными данными, тогда как в профессиональной литературе единственной опцией считается условная логистическая регрессия.

*Ключевые слова:* редкие события; логистическая регрессия; панельные данные; смещение; революции; метод максимального правдоподобия.

*Для цитирования:* Устюжанин В.В. Анализ редких событий: логистическая регрессия и альтернативы // Политическая наука. – 2026. – № 2. – С. 259–283. – DOI: <http://www.doi.org/10.31249/poln/2026.02.12>

## Введение

Революции всегда вызывали интерес социологов и политологов, однако только в последнее десятилетие авторы начали активно использовать количественные методы для их исследования. Во многом это стало возможным благодаря выходу кросс-национальной базы данных по революциям – NAVCO [Chenoweth, Lewis, 2013], – за счет которой авторы смогли подтвердить (или опровергнуть) давние гипотезы. Стоит выделить несколько вопросов, которыми занимаются современные исследователи революций<sup>1</sup> – анализ последствий [Biglaiser, Lee, McGauvran, 2024; Celestino, Gleditsch, 2013], анализ успеха [Chenoweth, Stephan, 2011; Dahlum, 2023], анализ выборка тактики [Ustyuzhanin, Korotayev, 2023] и, наконец, анализ причин начала революций [Beissinger, 2022; Chenoweth, Ulfelder, 2017; Korotayev et al., 2024], который привлек наибольшее внимание. При этом в настоящее время интерес к эмпирическому анализу революций только набирает обороты (см. рис. 1).

---

<sup>1</sup> Также читатель может найти систематический обзор последних исследований революций в [Пятое поколение..., 2025] или в [Korotayev et al., 2025].

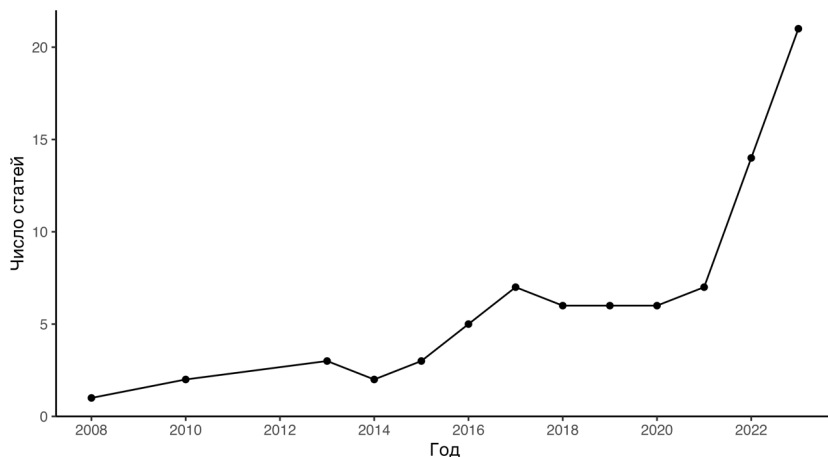


Рис. 1.  
**Число статей, исследующих революции  
количественными методами и с использованием  
кросс-национальных баз данных<sup>1</sup>**

В таких исследованиях зависимая переменная, революции или ее характеристики, часто представляет собой «редкие события» – «бинарные зависимые переменные, в которых от десятков до тысяч раз меньше единиц (наличие события), чем нулей (отсутствие события)» [King, Zeng, 2001, p. 138]. Некоторыми авторами было показано, что логистическая регрессия – самый частый метод для анализа – дает смещенные результаты при анализе редких событий [King, Zeng, 2001], однако единого стандарта для анализа так и не было выработано. Примечательным выглядит то, что в недавней фундаментальной работе М. Бейсинджер [Beissinger, 2022], столкнувшись с проблемой одновременно и панельных данных, и редких событий, при описании методологии часто ссылается на онлайн-форумы, а не на учебную или академическую литературу. Это связано, с одной стороны, с тем, что в учебниках по эконометрике не рассматривается проблема редких событий и тем более не рассматривается сочетание нескольких проблем сразу: и редких событий, и малой выборки, и панельных данных.

В статистической литературе довольно хорошо изучены все перечисленные проблемы по-отдельности: при малых выборках

<sup>1</sup> Данные взяты из [Korotayev et al., 2025].

[Nemes et al., 2009; Rainey, McCaskey, 2021], при панельных данных [Beck, 2018; Katz, 2001] и отчасти при редких событиях [King, Zeng, 2001]. Тем не менее надежность классической логистической регрессии при сочетании этих проблем еще не была проверена, хотя большая часть современных авторов, исследующих революции количественными методами, сталкивается именно с пересечением этих проблем. В такой ситуации, к сожалению, не выработано единого методологического стандарта, и многие исследователи прибегают к разным техникам, чья надежность не изучена.

В рамках настоящего исследования мы постараемся оценить, как использование логистической регрессии при разных исследовательских дизайнах с редкими событиями – в случае кросс-секционных и панельных данных – влияет на смещенность и устойчивость результатов. Кроме того, мы рассмотрим несколько альтернатив классической логистической регрессии и проверим, справляются ли они лучше с представленными проблемами. Другими словами, есть ли в контексте перечисленных проблем у классической логистической регрессии альтернативы? А если есть, то стоит ли отказываться от привычных методов?

### **Классическая логистическая регрессия: МП-оценщик**

Логистическая регрессия – один из самых часто используемых методов для анализа бинарных данных и, конечно, революций. В этом случае зависимая переменная принимает всего два возможных исхода: 0 («неуспех» / «революции не было») или 1 («успех» / «революция случилась»). Чтобы определить ее формально, можно представить, что каждое наблюдение в данных представляет из себя простой эксперимент, в котором событие происходит с некоторой вероятностью. Тогда пусть зависимая переменная  $Y \sim Ber(p)$ , то есть представляет собой дискретную случайную величину с распределением Бернулли с вероятностью события  $0 < p < 1$  и всего с двумя возможными исходами. Мы же хотим оценить, от чего эта вероятность зависит. Предположим, что  $p$  – это некоторая функция от независимых переменных. В качестве такой функции, как правило, используют логистическую функцию:

$$p(X) = [1 + \exp(-X\beta)]^{-1} \quad (1)$$

где  $X$  – это матрица независимых переменных, а  $\beta$  – вектор оцениваемых параметров, которые связывают вероятность с набором ковариат, то есть позволяют оценить условную вероятность  $P(Y=1|X\beta)$ . При этом логистическую регрессию можно представить в следующем виде, напоминая обычную линейную регрессию, где зависимая переменная  $y^*$  – это натуральный логарифм отношения шансов:

$$y^* = \ln \left[ \frac{P(Y = 1|X\beta)}{1 - P(Y = 1|X\beta)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = X\beta \quad (2)$$

Для оценки параметров  $\beta$  используется метод максимального правдоподобия (МП). К сожалению, полученные с его помощью оценки – вектор  $\beta$  – только асимптотически несмещены, то есть разница между оцененными и «настоящими» параметрами близка к 0 только при достаточно большом числе наблюдений:  $E|\beta| = \beta$  при  $n \rightarrow \infty$ . Таким образом, при малых выборках логистическая регрессия дает смещенные оценки<sup>1</sup>.

В случае с редкими событиями смещение в оценках ведет себя тем же образом, что и в случае с малой выборкой. Так, когда число событий мало, логистическая регрессия дает смещенные оценки, даже если сама выборка довольно большая. При наложении обеих проблем друг на друга – малой выборки и редкого события – величина смещения может быть крайне большой. В статистической литературе довольно хорошо изучены проблемы логистической регрессии при малых выборках [Nemes et al., 2009; Rainey, McCaskey, 2021], а также уже были предприняты попытки к изучению смещения при редких событиях [King, Zeng, 2001]. Мы же покажем, как ведет себя логистическая регрессия в случае комбинации этих проблем.

Также стоит отдельно отметить, что большинство исследований революций проводятся на панельных данных, когда выборка состоит из кластеров – например, страны или регионы, для которых мы наблюдаем некоторые переменные в определенном временном промежутке. Тогда единицей наблюдения могут выступать, например, страна-года – самый распространённый тип

---

<sup>1</sup> Читатель может обратиться к приложению, где формально описываются свойства оценщиков – смещенность и состоятельность, – которые довольно часто встречаются в тексте работы.

данных при изучении революций. У исследователей есть все основания ожидать, что такая структура данных влияет как на зависимую, так и на независимые переменные, потому что в стране (которая представляет собой переменную) может содержаться неучтенная в исследовании информация, которая не изменяется со временем (например, наличие горной местности, которая способствует продолжительности партизанских движений). Однако стоит отметить, что введение фиксированных эффектов в логистическую регрессию само по себе создает смещение [Katz, 2001]. Обычной стратегией борьбы с этим является оценка модели по стратам (*conditional logistic model* [Chamberlain, 1980]) за счет условного максимального правдоподобия (далее – УМП оценщик), но такая модель не дает интерпретируемых оценок предельных эффектов, учитывающих фиксированные эффекты. Исходя из этого, многие авторы принимают смещение. Тем не менее оно небольшое при относительно большом числе групп (например, стран) и количества наблюдений (например, страна-лет) в них [Beck, 2018; Katz, 2001]. Во многом проблема фиксированных эффектов является отражением проблемы малой выборки, но в этом случае именно небольшое число наблюдений в каждой подвыборке (например, в стране) создает смещение.

В случае с анализом редких событий нередко возникает ситуация, когда в определенных странах за рассматриваемый период революций не было, то есть для них зависимая переменная всегда равна нулю. В таком случае введение фиксированного эффекта на страну будет равносильно исключению данных (фильтрации) по таким странам [Beck, 2020] и, следовательно, недоучетом важной информации: например, почему в одних странах за рассматриваемой промежуток вообще не было революций? Однако в рамках настоящего исследования нас интересует то, как смещение, вызванное введением фиксированного эффекта, сочетается с проблемой редких событий: к сожалению, исследований, посвященных этой проблеме, нет, но в рамках настоящей работы мы постараемся дать предварительные результаты.

### **Скорректированная логистическая регрессия: ШМП-оценщик**

Как было сказано, смещение, порождаемое логистической регрессией, довольно хорошо изучено в статистической литерату-

ре. Его можно асимптотически представить в следующем виде [Firth, 1993]:

$$b(\beta) = \frac{b_1(\beta)}{n} + \frac{b_2(\beta)}{n^2} + \dots \quad (3)$$

где  $b_1(\beta)$ ,  $b_2(\beta)$  и т.д. – это компоненты общего смещения  $b(\beta)$ , которое обратно зависит от числа наблюдений  $n$  и определяется как  $b(\beta) = E(\hat{\beta}) - \beta$ . Д. Фиртф [Firth, 1993] предложил корректировать оценки  $\hat{\beta}$  на первый член из (3), который вносит самый большой вклад в смещение. Это достигается за счет изменения функции МП, куда вводится штраф (далее этот метод будет называться максимальное правдоподобие со штрафом – ШМП – *penalized maximum likelihood*):

$$L(\beta^{\text{ШМП}}) = L(\beta^{\text{МП}}) |A|^{1/2} \quad (4)$$

$$\ell(\beta^{\text{ШМП}}) = \ell(\beta^{\text{МП}}) + \frac{1}{2} \ln |A| \quad (5)$$

где  $L(\cdot)$  – это функция правдоподобия, а  $\ell(\cdot)$  – это ее натуральный логарифм;  $\beta^{\text{ШМП}}$ ,  $\beta^{\text{МП}}$  – коэффициенты, оцененные через ШМП или через классическое МП соответственно,  $A$  – информационная матрица Фишера, а  $|A|$  – ее детерминант<sup>1</sup>. ШМП-оценщик обладает теми же свойствами, что и МП-оценщик: он состоятелен, но оценки получаются смещенными, хоть и в меньшей степени за счет штрафа.

Исследователи революций в последние несколько лет начали активно использовать этот подход [Beissinger, 2022; Korotayev et al., 2024; Коррупция и революционная дестабилизация..., 2023], который сокращает смещение в малых выборках и косвенно помогает справляться с проблемой редких событий. Отдельно отметим еще

---

<sup>1</sup> Получение оценок методом МП достигается за счет максимизации функции правдоподобия, откуда становится понятным название этого метода. С практической точки зрения это задача максимизации, которая зачастую решается числовым способом, то есть через итеративную процедуру. Д. Фиртф предложил корректировать не уже полученные за счет этой процедуры оценки, а саму функцию оценки, которая «штрафует» экстремальные оценки параметров, тем самым уменьшая смещение. Величина этого штрафа задается самой функцией правдоподобия (то есть не требует от исследователя сильных априорных предположений) и основана на информации Фишера.

одно преимущество этого метода – решение проблемы полной сепарации. Например, когда для бинарного предиктора зависимая переменная всегда равна единице, оценки коэффициентов и их вариаций стремятся к бесконечности и становятся неустойчивыми, потому что метод МП, работающий через итеративные алгоритмы, не может найти единственный максимум. Следовательно, это не позволяет исследователю сделать вывод о связи между переменными, а также рассчитать интересующие величины на основе модели. Проблема полной сепарации может возникнуть в случае с малым числом наблюдений и когда зависимая и/или независимая переменная является редким событием [Mansournia et al., 2018]. Оценщик ШМП за счет штрафа позволяет дать конечную оценку коэффициента и его вариации [Kosmidis, Firth, 2021].

Проблема введения фиксированного эффекта в логистическую регрессию тоже косвенно связана с проблемой малой выборки: смещение велико в случае, когда число кластеров мало и величина наблюдений в них тоже мала. Соответственно, мы предполагаем, что ШМП также способен снизить смещение по сравнению с МП в случае модели с фиксированными эффектами. В настоящей работе мы сравним ШМП с МП в ситуации с фиксированными эффектами в контексте исследования с редкими событиями.

### Точная логистическая регрессия

Другим важным методом является «точная» логистическая регрессия (*exact logistic regression*), которая решает проблему как малой выборки, так и редкости события. Однако для ее вычислений требуются значительные вычислительные мощности и нацелена она прежде всего на тип данных с небольшой и дискретной размерностью. По сути, если данные в исследовании могут быть представлены как таблица сопряженности, то точная логистическая регрессия является оптимальным выбором: она даст конечные и более точные (но при этом не несмещенные) оценки коэффициентов и их вариации [Mehta, Patel, 1995]. В случае же большого числа переменных (и особенно – фиксированных эффектов), которые к тому же могут быть непрерывными, точная логистическая регрессия неприменима, что и объясняет ее малую популярность у исследователей.

### Симуляционное исследование

Для оценки смещения, порождаемого разными оценщиками в контексте редких событий, мы используем метод Монте-Карло. Мы создаем набор данных, приближенный к используемому исследователями в области изучения революций, а далее проводим анализ каждым из рассмотренных методов. Так как мы сами порождаем эти данные, то точно знаем истинные значения коэффициентов, что позволяет оценить размер смещения. Чтобы оценки смещения получились устойчивыми и отражали реальность, мы проводим 5000 симуляций для каждой гипотетической ситуации, а дальше считаем его среднее: при таком большом числе идентичных симуляций полученные результаты, по закону больших чисел, будут довольно точно отражать действительность.

Мы проводим эксперименты в нескольких сценариях. Первый – это кросс-секционные данные, когда наблюдения не имеют важных пространственных отличий. Такая ситуация с некоторыми оговорками соответствует, например, исследованию вопроса «почему одни революции стали успешными, а другие – нет?». В таком случае единица наблюдения – это отдельный революционный кейс. Мы предполагаем следующий процесс порождения данных:

$$y_i^* = \beta_0 + \beta_1 x_i + \beta_2 z_i$$

$$(i = 1, \dots, n)$$
(6)

$$\beta_1 = \beta_2 = 1, \beta_0 = g[p]$$

$$Z \sim N(0,1), X = D + Z, D \sim N(0,1)$$

где нотация повторяет (1), главная независимая переменная –  $x_i$ , чей коэффициент,  $\beta_1$ , для удобства равен 1. Контрольной переменной выступает  $z_i$ , которая оказывает эффект и на зависимую переменную  $y_i^*$ , и на независимую  $x_i$ . Константа моделируется для каждой заданной вероятности  $p$ <sup>1</sup>. Соответственно, при заданном процессе порождения

---

<sup>1</sup> Вероятность  $p$  представляет собой математическое ожидание зависимой переменной. Как видно из (1), эта функция нелинейна и предполагает зависимость между переменными  $X$  и  $Z$ , что задано процессом порождения данных. Для того чтобы выразить  $\beta_0$  из математического ожидания от (1), мы нашли приближительное значение этой функции через ряд Тейлора второго порядка (первый

данных бинарная зависимая переменная,  $Y = (y_1, \dots, y_n)$ , моделируется как набор случайных независимых переменных с распределением Бернулли, где  $p_i(X_i) = [1 + \exp(-X_i\beta)]^{-1}$ .

Второй сценарий – это панельные данные, когда наблюдения имеют важные временные и пространственные характеристики. В случае с исследованием революций для ответа на вопрос «почему революции начинаются в определенных странах в определенное время?», как правило, используют наблюдения вида страна-год. Мы предполагаем следующий процесс порождения данных, где панельные данные для простоты предполагаются сбалансированными<sup>1</sup>:

$$y_{ij}^* = \alpha_j + \beta_1 x_{ij} + \beta_2 z_{ij}$$

$$(i = 1, \dots, n; j = 1, \dots, 20; \forall j n_j = n / 20) \quad (7)$$

$$\beta_1 = \beta_2 = 1, \alpha_j = N(g[p], 0.5)$$

$$Z \sim N(0, 1), X = D + Z + (g[p] - \alpha_j), D \sim N(0, 1)$$

где нотация повторяет предыдущий сценарий: главная независимая переменная –  $x_i$ , а ее коэффициент  $\beta_1$  для удобства равен 1. Контрольной переменной выступает  $z_i$ . В модель включаются дамми-переменные  $\alpha_j$  (пространственный фиксированный эффект), каждая из которых представляет собой реализацию случайной нормальной величины со средним в общей константе  $g[p]$ , которая зависит от заданной общей вероятности  $p^2$ . Более того,  $\alpha_j$  влияет на независимую переменную  $x_{ij}$ , что также делает дамми-переменные важной контрольной пе-

и второй моменты (1) легко выводятся из модели порождения данных) и получили кубическое уравнение, которое решили численным способом. В итоге  $\beta_0 \approx \ln[q/(1-q)]$ , где  $q$  – это логистическая функция из (1) и для каждой заданной  $p$  мы находим, чему должно равняться  $q$ . За счет относительно небольшой общей вариации переменных  $X$  и  $Z$ , аппроксимация получилась достаточно точной и ее отклонение от целевой  $p$  в среднем не превышает 0.01.

<sup>1</sup> То есть для каждого кластера (страны) число наблюдений (лет, за которые мы наблюдаем переменные) одинаково.

<sup>2</sup> Таким образом, в среднем в симуляциях нам удастся приблизиться к заданной вероятности (т.е. редкости события). Метод для оценки  $g[p]$  такой же, как и при первом сценарии.

ременной. Наконец, итоговая бинарная зависимая переменная получается тем же образом, что и в первом сценарии.

Оба сценария для наглядности можно представить в виде ациклических направленных графов (НАГ или *DAG*)<sup>1</sup>, которые отражают каузальную модель каждого из сценариев – кросс-секционных (рис. 2а) и панельных (рис. 2б) данных – и повторяют нотацию из этих сценариев. Отметим, что оба сценария приближены к исследовательским дизайнам, к которым часто прибегают ученые. Так, переменная *D* представляет набор ненаблюдаемых факторов (т.е. тех, которые нельзя или трудно операционализировать, но которые играют важную роль в теоретической модели – например, уровень депривации населения или его ожидания относительно уровня жизни). Переменная *Z* – это наблюдаемые важные факторы, которые исследователь включает в модель в качестве контрольных переменных (например, численность населения страны или уровень ее экономического развития). Переменные *X* и *Y* представляют собой независимую и зависимую переменные соответственно. В свою очередь, *FE* – это набор дамми-переменных, которые вбирают в себя всю неизменяющуюся во времени вариацию, влияющую на *X* и на *Y* (в случае реальных исследований это могут быть страны, регионы, округа и т.д., в которых заложена информация, например, о культуре и исторически сложившихся институтах). Конечно, оба графа на рисунке 2 не отражают реальность в полной мере (*Y*, например, может влиять на сторонние переменные или даже на *X*, а у *Z* может быть множество сторонних переменных-причин), однако представляют важный инструмент для оценки состоятельности каузальной модели, оцениваемой исследователем. В обоих представленных сценариях эффект *X* на *Y* может быть оценен – это теоретически возможно, потому мы учли всю критически необходимую информацию, которая могла бы исказить оценку эффекта *X* на *Y* (т.е.  $\beta_1$ ), и дело остается только за технической оценкой этого эффекта.

---

<sup>1</sup> Подробнее про НАГи и их использование для определения дизайна исследования см.: [Cinelli, Forney, Pearl, 2024].

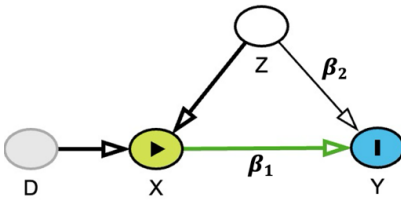


Рис. 2 а

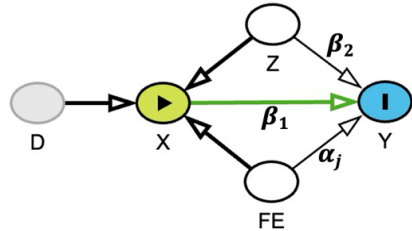


Рис. 2 б

Рис. 2.

**Каузальная модель порождения данных  
для симуляционного исследования:  
кросс-секционный (а) и панельный (б) сценарии**

Для каждого из сценариев мы также варьируем редкость события и размер выборки, чтобы учесть максимальное количество гипотетических ситуаций, в которых может оказаться исследователь. Редкость событий мы меняем от очень редкого до почти редкого, рассматривая следующие вероятности  $p$ : 0,01, 0,05, 0,1, 0,2, то есть на одну единицу в выборке около 100, 20, 10 и 5 нулей соответственно. В свою очередь, размер выборки  $n$  изменяется от маленькой до большой: 200, 300, ..., 1000 наблюдений (с шагом в 100). Таким образом, мы предполагаем, что величина смещения будет максимальной в случае с малой выборкой и очень редким событием, когда  $p = 0,01$ ,  $n = 200$  и число единиц (революций) приблизительно равно 2. В свою очередь, мы ожидаем, что наименьшее смещение будет в случае с  $p = 0,2$ ,  $n = 1000$  и числом единиц (революций) приблизительно равным 200. Чтобы комплексно рассмотреть сочетание двух проблем – малой выборки и редкости события – мы проводим анализ по всем возможным комбинациям  $p$  и  $n$ .

Все расчеты делались на языке программирования R с использованием пакета *stats*<sup>1</sup> для МП оценщика, *brglm2*<sup>2</sup> для ШМП оценщика и пакета *survival*<sup>1</sup> для УМП оценщика (функция *clogit*).

<sup>1</sup> Это предустановленный в R пакет. Реализация для панельных данных: `model = glm(data = data, y ~ x + z + FE, family = binomial(link = "logit"))`.

<sup>2</sup> Kosmidis, I. (2024). *brglm2*: bias reduction in generalized linear models. R package version 0.9.2. Реализация для панельных данных: `model = glm(data = data, y ~ x + z + FE, family = binomial(link = "logit"), method = "brglmFit", type = "MPL_Jeffreys")`.

## Симуляции: результаты

В настоящем разделе даны результаты, сравнивающие два описанных метода оценки логистической регрессии – МП и ШМП. Сначала мы описываем результаты первого сценария – кросс-секционных данных – при разных размерах выборки (от 200 до 1000) и разной средней вероятности события: от 0,01 (крайне редкое событие) до 0,2 (умеренно редкое событие).

На рисунке 3 представлены точечные диаграммы между величиной смещения в константе (ось  $y$ ) и целевом коэффициенте при независимой переменной (ось  $x$ ) при  $n = 200, 300, 400, 500$  и при  $p = 0,01$  (то есть для всех диаграмм событие очень редкое) для каждой симуляции  $i$ . Таким образом, смещение минимально около точки 0 при пересечении осей: тогда отклонения и в константе, и в коэффициенте незначительны. Видно, что с увеличением размера выборки смещение все больше концентрируется вокруг 0, однако ШМП-оценка всегда превосходит МП-оценку как по величине смещения, так и по его вариации, что отражено эллипсами, отражающими совместную плотность распределения смещений. Таким образом, ШМП-оценщик доминирует над МП-оценщиком, который дает сравнительно большое смещение в константе<sup>2</sup> и сильно недооценивает вероятность события, что дает ложные выводы при прогнозировании или любой работе с вероятностями на основе оцененной через МП модели.

На рисунке 4 представлено сравнение МП- и ШМП-оценщиков в смещении в целевом коэффициенте (ось  $x$ ) при независимой переменной  $X$  при разных  $p$  (разные графики) и  $n$  (ось  $y$ ). На удивление, ШМП-оценщик не только доминирует над МП, но и оказывается в среднем несмещенным даже при малой выборке и крайне редком событии, то есть при самом негативном сценарии. В свою очередь, МП оценщик всегда дает смещенные оценки: даже при достаточно большом  $n$  и при относительно небольшой редкости события. При малой выборке и крайне редком событии, коэффициент МП смещен в 4 раза (!) относительно реального коэффициента, а ШМП – на уровне погрешности. Таким образом, исследователь, использующий МП, не только не получил бы удовлетворительной

---

<sup>1</sup> Therneau, T. (2024). survival: survival analysis. R package version 3.7.0. Реализация для панельных данных: `model = clogit(y ~ x + z + strata(FE), method = "exact")`.

<sup>2</sup> Отметим, что настоящая константа всегда отрицательна.

оценки величины коэффициента, но и мог бы сделать ложное суждение о направлении связи.

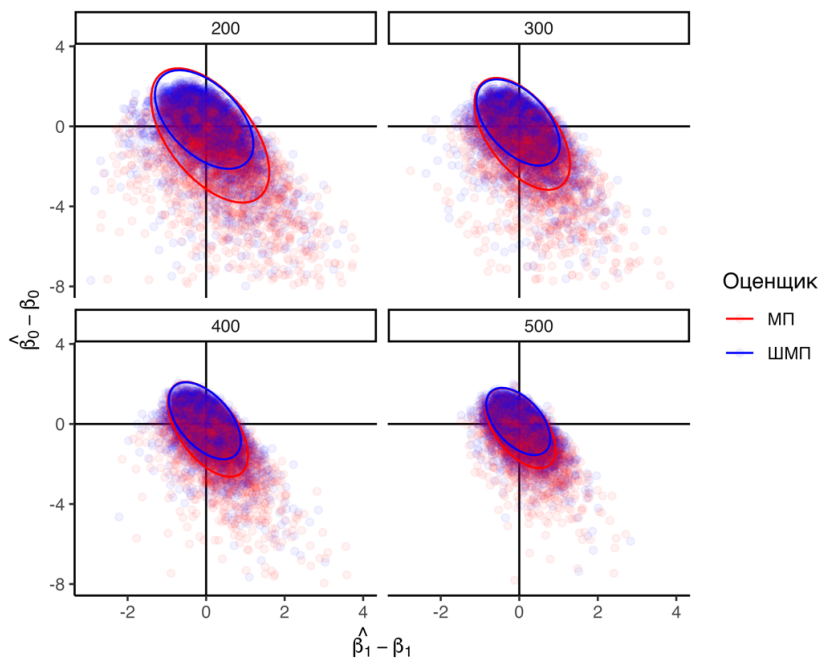


Рис. 3.

**Смещение в коэффициентах, порождаемое разными оценщиками при разном размере выборки и при  $p = 0.01$ : кросс-секционные данные<sup>1</sup>**

<sup>1</sup> Эллипсы отражают плотность двумерного распределения между переменными, аппроксимируя его как t-распределение.

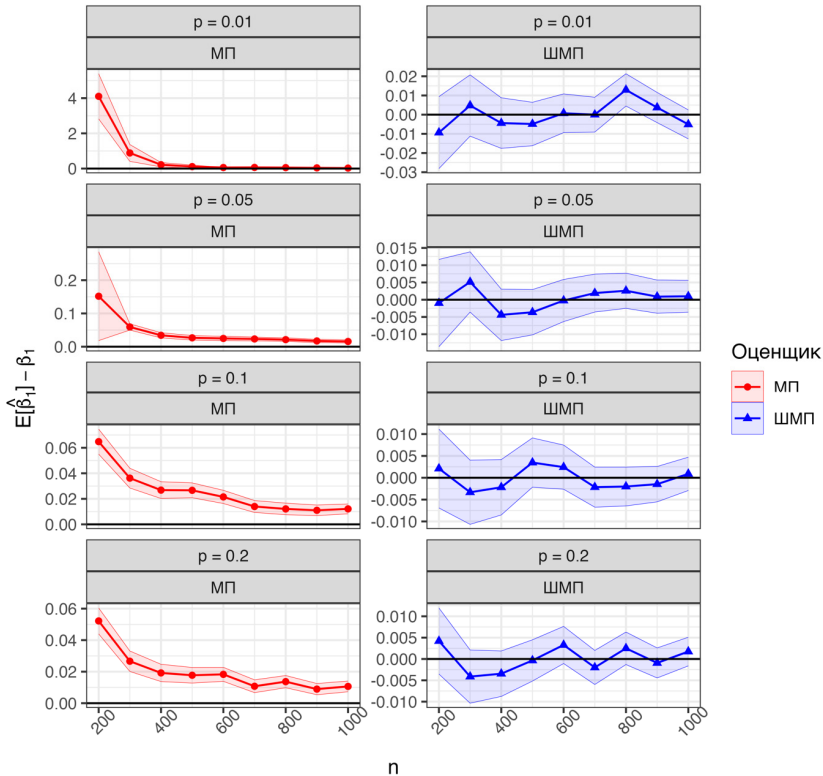


Рис. 4.

**Смещение в целевом коэффициенте, порождаемое разными оценщиками при разном размере выборки и при разных  $p$ : кросс-секционные данные<sup>1</sup>**

В таблицах 1 и 2 представлены результаты по величине смещения в абсолютных величинах в целевом коэффициенте и в константе соответственно. Так, при крайне редком событии ( $p = 0,01$ ) МП завышает величину эффекта на 4,1, что потенциально может поменять знак эффекта. При остальных  $n$  знак уже не меняется при любой редкости события и вывод о направлении связи получится верным, однако величина смещения будет довольно большой. В свою очередь, ШМП-оценщик, наоборот, склонен не-

<sup>1</sup> Ось  $y$  имеет разную размерность на графиках; залитая область представляет 95% ДИ.

сколько занижать величину целевого эффекта, однако общее смещение не превышает одной сотой при любых сценариях, что позволяет сделать не только корректный вывод о связи между переменными, но и использовать эту модель для прогнозирования.

Таблица 1

**Смещение в целевом коэффициенте, порождаемое разными оценщиками при разном размере выборки и при разных  $p$ :  
кросс-секционные данные<sup>1</sup>**

p / n	$b(\beta_1^{МП})$				$b(\beta_1^{ШМП})$			
	200	300	400	500	200	300	400	500
<b>0.01</b>	4,103	0,886	0,219	0,112	-0,009	0,005	-0,004	-0,005
<b>0.05</b>	0,152	0,059	0,034	0,027	-0,001	0,005	-0,004	-0,004
<b>0.1</b>	0,065	0,036	0,027	0,027	0,002	-0,003	-0,002	0,003
<b>0.2</b>	0,052	0,027	0,019	0,018	0,004	-0,004	-0,003	0

В таблице 2 видно, что величина смещения в константе ведет себя противоположным образом относительно целевого коэффициента. МП-оценщик дает крайне смещенные оценки, особенно в случае с малой выборкой и очень редким событием. В свою очередь, ШМП-оценщик всегда дает приблизительно несмещенные оценки. Отметим, что хотя константа редко анализируется в академических исследованиях, она вносит огромный вклад при расчетах вероятностей событий на основе моделей, что делает ее очень важным параметром, если исследователи интересуются, например, прогнозированием.

Таблица 2

**Смещение в константе, порождаемое разными оценщиками при разном размере выборки и при разных  $p$ :  
кросс-секционные данные<sup>2</sup>**

p / n	$b(\beta_0^{МП})$				$b(\beta_0^{ШМП})$			
	200	300	400	500	200	300	400	500
<b>0.01</b>	-24,545	-4,823	-1,465	-0,546	0,022	-0,052	-0,021	-0,002
<b>0.05</b>	-0,54	-0,213	-0,144	-0,106	-0,028	-0,014	-0,001	0,006
<b>0.1</b>	-0,19	-0,113	-0,083	-0,083	-0,003	0,006	0,004	-0,014
<b>0.2</b>	-0,115	-0,067	-0,05	-0,042	-0,006	0,003	0,001	-0,001

<sup>1</sup> МП – максимальное правдоподобие, ШМП – максимальное правдоподобие со штрафом; настоящий коэффициент равен 1.

<sup>2</sup> МП – максимальное правдоподобие, ШМП – максимальное правдоподобие со штрафом; настоящая константа для разных  $p$  разная, но всегда отрицательная.

Таким образом, ШМП-оценщик всегда доминирует над МП-оценщиком в случае с кросс-секционными данными и редкими событиями. Если размер выборки довольно большой (от 400), а событие не крайне редко ( $p$  от 0,05), логистическая регрессия, оценённая через МП, даёт удовлетворительные результаты, позволяющие определить направление связи, однако не величину эффекта, что не позволяет рассчитывать, например, отношение шансов (odds ratio) или предельные эффекты / условные предсказания. В свою очередь, ШМП-оценщик всегда даёт оценки с минимальным смещением, что позволяет её использовать при почти любых  $n$  и редкости события, а также рассчитывать все интересные исследователя величины, которые не ограничиваются только целевым коэффициентом.

Рассмотрев поведение МП- и ШМП-оценщиков при первом сценарии с кросс-секционными данными, мы переходим ко второму сценарию с панельными данными. Гипотетические ситуации остаются теми же, а число групп (фиксированных эффектов) равно 20.

На рисунке 5 представлены точечные диаграммы между величиной смещения в константах (ось  $y$ , дано среднее отклонение во всех 20 константах) и целевом коэффициенте при независимой переменной (ось  $x$ ) при  $n = 200, 300, 400, 1000$  и при  $p = 0.01$  (то есть для всех диаграмм событие очень редкое). Таким образом, смещение минимально около точки 0 при пересечении осей: тогда отклонения и в константах, и в коэффициенте незначительны. Видно, что с увеличением размера выборки смещение стремится к 0, однако МП-оценка даже при достаточно большом  $n = 1000$  оказывается крайне смещённой и обладающей очень большой вариацией. В свою очередь, ШМП-оценка оказывается наименее смещённой, хоть и отклоняется в положительную сторону по константам при  $n = 200$ .

На рисунке 6 представлено сравнение МП- и ШМП-оценщиков в смещении в целевом коэффициенте (ось  $x$ ) при независимой переменной при разных  $p$  (разные графики) и  $n$  (ось  $y$ ). Также мы даём результаты по УМП-оценщику, который, как было сказано выше, является общепризнанной и рекомендуемой заменой МП в случае логистической регрессии с фиксированными эффектами. На удивление, ШМП-оценщик не только даёт почти несмещённые результаты по оценке целевого эффекта, но и превосходит УМП при любых сценариях, справляясь сразу со всеми проблемами: и малой выборкой, и панельными данными, и редкими событиями, тогда как УМП корректирует только одну часть смещения классического МП – панельные данные. В свою очередь, МП-оценщик не

только всегда дает смещенные результаты, но и при малых  $n$  (до 500) и небольшом  $p$  (до 0,01) может вообще не давать конечную оценку эффекта, в то время как остальные оценщики производят не только более точные оценки (особенно ШМП), но и конечные, что позволяет сделать вывод на основе данных. Таким образом, исследователь, использующий МП для панельных данных при довольно редком событии и относительно малой выборке, может вообще не получить хоть сколько-нибудь адекватных оценок эффекта. Более того, в отличие от сценария с кросс-секционными данными, МП-оценка даже при большой выборке ( $n = 1000$ ) остается крайне смещенной относительно УМП и ШМП оценщиков. При этом последний при  $p = 0,2$  дает в среднем несмещенные оценки при любом размере выборки.

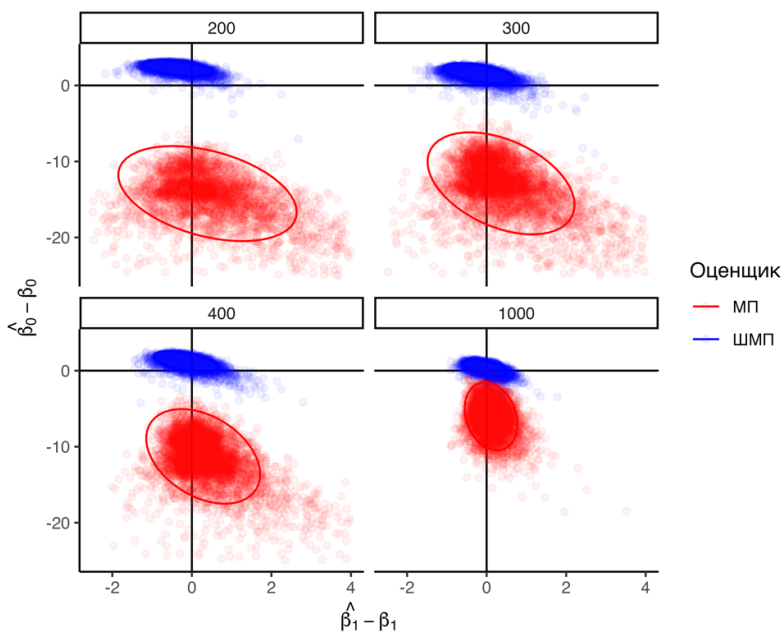


Рис. 5.

**Смещение в коэффициентах, порождаемое разными оценщиками при разном размере выборки и при  $p = 0,01$ : панельные данные<sup>1</sup>**

<sup>1</sup> Эллипсы отражают плотность двумерного распределения между переменными, аппроксимируя его как t-распределение.

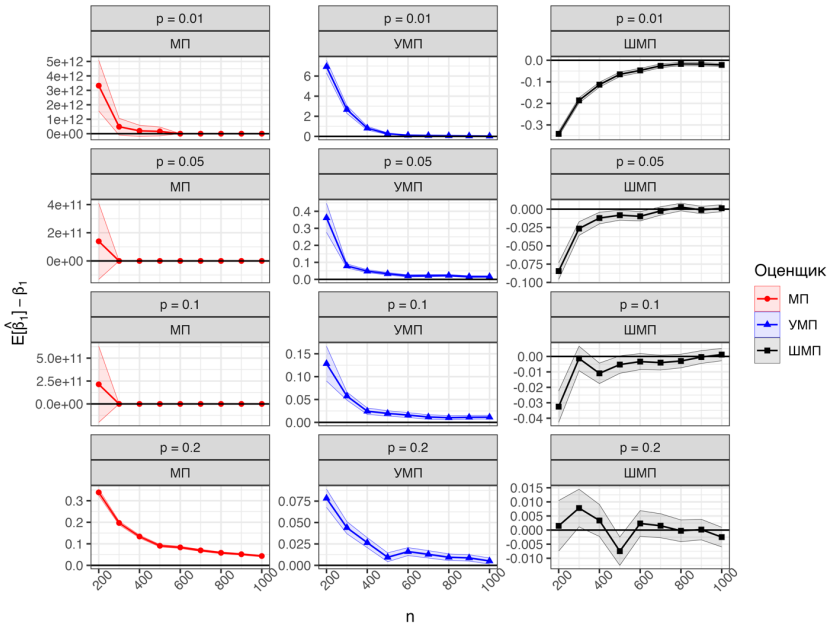


Рис. 6.

**Смещение в целевом коэффициенте, порождаемое разными оценщиками при разном размере выборки и при разных  $p$ : панельные данные<sup>1</sup>**

В таблицах 3 и 4 представлены результаты по величине смещения в абсолютных величинах в целевом коэффициенте и в константах (которых 20 по числу фиксированных эффектов) соответственно. Так, при крайне редком событии ( $p = 0,01$ ) МП при  $n$  до 500 не дает хоть сколько-нибудь адекватных оценок в принципе, в то время как ШМП и УМП позволяют получить хотя бы смещенные оценки. Отметим, что ШМП всегда оказывается менее смещенным, чем УМП и его смещение не превосходит 0,34 в самом худшем случае, что позволяет сделать корректный вывод о направлении связи. Таким образом, логистической регрессией, оцененной через классическое МП, к сожалению, при ситуации панельных данных, редких событий и не очень большого размера выборки (до 500) пользоваться нельзя: величина смещения слиш-

<sup>1</sup> Ось  $y$  имеет разную размерность на графиках; заштрихованная область представляет 95% ДИ.

ком большая. В свою очередь, УМП-оценщик помогает решить проблему панельных данных (накладывая при этом огромное ограничение в виде невозможности получить оценки фиксированных эффектов), однако не справляется с малой выборкой и редкими событиями: при очень редком событии потенциально может поменяться знак целевого коэффициента. На удивление, ШМП-оценщик решает все эти проблемы, оставаясь сильно смещенным только в самых негативных гипотетических ситуациях, где остальные методы либо не дают интерпретируемых оценок, либо потенциально искажают оценки вплоть до смены знака. Повторимся, что на наш взгляд, это самое важное наблюдение настоящей работы: ШМП-оценщик отлично справляется с панельными данными, тогда как в профессиональной литературе единственной опцией оценки логистической регрессии с фиксированными эффектами признается только УМП-оценщик.

Таблица 3

**Смещение в целевом коэффициенте, порождаемое разными оценщиками при разном размере выборки и при разных  $p$ : панельные данные<sup>1</sup>**

p / n	$b(\beta_1^{МП})$				$b(\beta_1^{ШМП})$				$b(\beta_1^{УМП})$			
	200	300	400	500	200	300	400	500	200	300	400	500
0,01	$3,32 \times 10^{12}$	$4,8 \times 10^{11}$	$1,99 \times 10^{11}$	$1,57 \times 10^{11}$	-0,341	-0,186	-0,113	-0,065	6,935	2,671	0,816	0,266
0,05	$1,39 \times 10^{11}$	0,307	0,206	0,156	-0,084	-0,027	-0,012	-0,008	0,362	0,079	0,048	0,033
0,1	$2,14 \times 10^{11}$	0,242	0,153	0,119	-0,033	-0,001	-0,011	-0,005	0,128	0,058	0,024	0,02
0,2	0,339	0,197	0,134	0,091	0,001	0,008	0,003	-0,007	0,078	0,044	0,026	0,009

В таблице 4 видно, что величина смещения в константе ведет себя схожим образом, что и ситуации с кросс-секционными данными (см. табл. 2). МП-оценщик дает крайне смещенные оценки, а в случае с малой выборкой и очень редким событием не дает адекватных оценок в принципе. В свою очередь, ШМП-оценщик всегда дает конечные оценки, которые при этом остаются заметно смещенными при очень редких событиях ( $p = 0.01$ ) и очень малой выборке ( $n$  до 400). Отметим, что хотя константа редко анализируется в академических исследованиях, она вносит огромный вклад при расчетах вероятностей событий на основе моделей, что делает

<sup>1</sup> МП – максимальное правдоподобие, ШМП – скорректированное максимальное правдоподобие, УМП – условное максимальное правдоподобие; настоящий коэффициент равен 1.

ее очень важным параметром, если исследователи интересуются, например, прогнозированием.

Таблица 4

**Среднее смещение в константах, порождаемое разными оценщиками при разном размере выборки и при разных  $p$ : панельные данные<sup>1</sup>**

p / n	$\bar{b}(\alpha^{МП})$				$\bar{b}(\alpha^{ШМП})$			
	200	300	400	500	200	300	400	500
<b>0,01</b>	-1,34 x 10 <sup>13</sup>	-2,3 x 10 <sup>12</sup>	-7 x 10 <sup>11</sup>	-6,8 x 10 <sup>11</sup>	2,205	1,469	1,083	0,814
<b>0,05</b>	-5,3 x 10 <sup>11</sup>	-6,556	-4,765	-3,72	0,657	0,345	0,225	0,14
<b>0,1</b>	-4,4 x 10 <sup>11</sup>	-3,921	-2,614	-1,868	0,317	0,148	0,086	0,053
<b>0,2</b>	3,625	-1,935	-1,158	-0,764	0,131	0,042	0,023	0,02

Таким образом, в случае с кросс-секционными данными сочетание малой выборки и редкого события порождает огромное смещение в МП оценщике логистической регрессии, и исследователь рискует не только не получить относительно точной оценки эффекта, но и сделать абсолютное ложное суждение о направлении связи. В свою очередь, ШМП оценщик дает наименее смещенные оценки при любой редкости события и даже при очень малой выборке, что позволяет сделать исследователю выводы о величине эффекта и рассчитать интересующие величины на основе оцененной модели.

### Заключение

В рамках настоящего исследования мы постарались оценить, как использование логистической регрессии при разных исследовательских дизайнах с редкими событиями – в случае кросс-секционных и панельных данных – влияет на смещенность и устойчивость результатов. Мы нашли, что в случае с кросс-секционными данными сочетание малой выборки и редкого события порождает огромное смещение в МП-оценщике логистической регрессии и исследователь рискует не только не получить относительно точной оценки эффекта, но и сделать абсолютно ложное

<sup>1</sup> МП – максимальное правдоподобие, ШМП – скорректированное максимальное правдоподобие; настоящие 20 констант для разных  $p$  разные, но всегда отрицательные; среднее смещение в 20 константах рассчитывалось как арифметическое среднее.

суждение о направлении связи. В свою очередь, ШМП-оценки смещены в наименьшей степени, и такой оценщик дает почти что несмещенные оценки при любой редкости события и даже при очень малой выборке, что позволяет сделать исследователю выводы о величине эффекта и рассчитать интересующие величины на основе оцененной модели.

В случае с панельными данными МП-оценщик логистической регрессии в ситуации относительно малой выборки (и даже при  $n = 1000$ ) и редких событий не применим вовсе. На удивление, ШМП-оценщик решает все эти проблемы – фиксированные эффекты, малая выборка и редкость события, – оставаясь заметно смещенным только в самых негативных гипотетических ситуациях ( $n$  до 300 и  $p = 0.01$ ), где остальные методы либо не дают интерпретируемых оценок, либо могут изменить направление связи. На наш взгляд, это самое важное наблюдение настоящей работы: ШМП-оценщик отлично справляется с панельными данными, тогда как в профессиональной литературе единственной опцией оценки логистической регрессии с фиксированными эффектами признается только УМП-оценщик.

**V.V. Ustyuzhanin\***

### **Rare events analysis: logistic regression and alternatives**

*Abstract.* The study of revolutions has long been a topic of interest to sociologists and political scientists. However, only in recent years authors have begun to employ quantitative methods with greater regularity in their analysis of these events. In such studies, the dependent variable, namely revolutions or their characteristics, frequently represent a “rare event”. The most commonly used method, logistic regression, has been demonstrated by some authors to yield biased results when analysing such data. However, no unified standard for analysis has been established. Furthermore, authors frequently encounter additional challenges, including the analysis of panel data and the use of small sample sizes. The reliability of classical logistic regression in addressing these issues has yet to be evaluated, despite the fact that the majority of contemporary authors frequently encounter these challenges. The present study aimed to assess the bias of results produced by classical logistic regression when used in different research designs with rare events, specifically cross-sectional and panel data. Furthermore, we put forward an alternative approach, namely penalized logistic regression. In the case of cross-sectional data, the combination of a small sample size and a rare event leads to a significant bias in classical logistic regression.

---

\* **Ustyuzhanin Vadim**, RANEPА University; HSE University (Moscow, Russia), e-mail: vvustiuzhanin@yandex.ru

This result shows that researchers face the challenge of obtaining not only an inaccurate estimate of the effect but also making a false judgement about the direction of the relationship. In contrast, penalized logistic regression produces almost unbiased estimates regardless of the rarity of the event or the size of the sample. In the context of panel data, the application of classical logistic regression is not viable in scenarios characterized by a limited sample size and infrequent events. In contrast, our findings suggest that penalized logistic regression is a viable alternative for analyzing panel data, whereas conditional logistic regression is currently the only option recommended in the professional literature.

*Keywords:* rare events; logistic regression; panel data; bias; revolutions; maximum likelihood estimator.

*For citation:* Ustyuzhanin V.V. Rare events analysis: logistic regression and alternatives. *Political science (RU)*. 2026, N 2, P. 259–283. DOI: <http://www.doi.org/10.31249/poln/2026.02.12>

## References

- Beck N. Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues. *arXiv preprint arXiv:1809.06505*, 2018.
- Beck N. Estimating grouped data models with a binary-dependent variable and fixed effects via a logit versus a linear probability model: the impact of dropped units. *Political analysis*. 2020, Vol. 28, N 1, P. 139–145.
- Beissinger M. *The revolutionary city: urbanization and the global transformation of rebellion*. Princeton: Princeton university press, 2022, 566 p.
- Biglaiser G., Lee H., McGauvran R.J. Domestic political unrest and sovereign bond ratings in the developing world. *Journal of conflict resolution*. 2024, Vol. 68, N 7–8, P. 1580–1607.
- Celestino M.R., Gleditsch K.S. Fresh carnations or all thorn, no rose? Nonviolent campaigns and transitions in autocracies. *Journal of peace research*. 2013, Vol. 50, N 3, P. 385–400.
- Chamberlain G. Analysis of covariance with qualitative data. *The review of economic studies*. 1980, Vol. 47, N 1, P. 225–238.
- Chenoweth E., Lewis O.A. Unpacking nonviolent campaigns: Introducing the NAVCO 2.0 dataset. *Journal of peace research*. 2013, Vol. 50, N 3, P. 415–423.
- Chenoweth E., Stephan M.J. *Why civil resistance works: The strategic logic of nonviolent conflict*. New York: Columbia university press, 2011, 320 p.
- Chenoweth E., Ulfelder J. Can structural conditions explain the onset of nonviolent uprisings? *Journal of conflict resolution*. 2017, Vol. 61, N 2, P. 298–324.
- Cinelli C., Forney A., Pearl J. A crash course in good and bad controls. *Sociological methods & research*. 2024, Vol. 53, N 3, P. 1071–1104.
- Dahlum S. Joining forces: Social coalitions and democratic revolutions. *Journal of peace research*. 2023, Vol. 60, N 1, P. 42–57.
- Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993, Vol. 80, N 1, P. 27–38.
- Katz E. Bias in conditional and unconditional fixed effects logit estimation. *Political analysis*. 2001, Vol. 9, N 4, P. 379–384.

- King G., Zeng L. Logistic regression in rare events data. *Political analysis*. 2001, Vol. 9, N 2, P. 137–163.
- Korotayev A., Grinin L., Ustyuzhanin V., Fine E. The fifth generation of revolutionary studies. Systematic review. *Logos*. 2025, Vol. 35, N 1, P. 191–316. (In Russ.)
- Korotayev A., Ustyuzhanin V., Grinin L., Fain E. The fifth generation of revolution studies. Part II: A systematic review of substantive findings (Revolution causes, forms, and waves). *Critical sociology*. 2025, Vol. 51, N 4, P. 1–48 (In press).
- Korotayev A., Zhdanov A., Grinin L., Ustyuzhanin V. Revolution and democracy in the twenty-first century. *Cross-cultural research*. 2024, Vol. 59, N 2, P. 180–215.
- Kosmidis I., Firth D. Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika*. 2021, Vol. 108, N 1, P. 71–82.
- Mansournia M.A., Geroldinger A., Greenland S., Heinze G. Separation in logistic regression: causes, consequences, and control. *American journal of epidemiology*. 2018, Vol. 187, N 4, P. 864–870.
- Nemes S., Jonasson J.M., Genell A., Steineck G. Bias in odds ratios by logistic regression modelling and sample size. *BMC medical research methodology*. 2009, Vol. 9, N 56, P. 1–5.
- Mehta C.R., Patel N.R. Exact logistic regression: theory and examples. *Statistics in medicine*. 1995, Vol. 14, N 19, P. 2143–2160.
- Rainey C., McCaskey K. Estimating logit models with small samples. *Political science research and methods*. 2021, Vol. 9, N 3, P. 549–564.
- Ustyuzhanin V., Korotayev A. Education and revolutions: why do revolutionary uprisings take violent or nonviolent forms? *Cross-cultural research*. 2023, Vol. 57, N 4, P. 352–390.
- Ustyuzhanin V., Kostin M., Grinin L., Korotayev A. Corruption and revolutionary destabilization: the experience of quantitative analysis. *Journal of sociology and social anthropology*. 2023, Vol. 26, N 3, P. 53–99. (In Russ.)

## Литература на русском языке

- Коротаев А.В., Гринин Л.Е., Устюжанин В.В., Файн Е.Д. Пятое поколение исследований революции. Систематический обзор // Логос. – 2025. – Т. 35, № 1. – С. 191–316.
- Устюжанин В., Костин М., Гринин Л., Коротаев А. Коррупция и революционная дестабилизация: опыт количественного анализа // Журнал Социологии и Социальной Антропологии – 2023. – Т. 26, № 3. – С. 53–99.

## ПРИЛОЖЕНИЕ

### Смещенность и состоятельность оценок

Любой количественный анализ концентрируется вокруг оценок реальных параметров и способов их получения. Это может быть среднее, параметр распределения (например, вероятность  $p$  в

распределении Бернулли), регрессионный коэффициент и так далее. Конечно, оценки могут быть «хорошими» и «плохими». Для того, чтобы провести границу между ними, как правило, рассматриваются три ключевых свойства полученных оценок: смещенность, состоятельность и эффективность. В настоящей работе мы оперируем только первыми двумя свойствами.

Несмещенность оценок предполагает, что оценка параметра в среднем равна реальному значению этого параметра. Определим это более строго: оценка  $\theta$  параметра  $\theta \in \Theta$  называется несмещенной, если  $E|\theta| = \theta$ , где  $\Theta$  – это множество возможных значений параметра  $\theta$  (которое может быть открытым интервалом). Таким образом, при любом размере выборки математическое ожидание оценки равно реальному параметру.

Состоятельность оценок предполагает, что с увеличением размера выборки оценка параметра стремится к реальному значению параметра. Определим это более строго: оценка  $\theta$  параметра  $\theta \in \Theta$  называется состоятельной, если  $\theta \rightarrow \theta$  при  $n \rightarrow \infty$ . Другими словами, оценка  $\theta$  должна стремиться к  $\theta$  по вероятности, чтобы быть состоятельной. Практически это означает, что при  $n \rightarrow \infty$ : (1) смещение  $b(\theta) = E|\theta| - \theta \rightarrow 0$  и (2) вариация оценки  $Var(\theta) \rightarrow 0$ . При этом первое положение о стремлении смещения,  $b(\theta)$ , к нулю также называется асимптотической несмещенностью (потому что оценка становится несмещенной только в пределе, то есть при бесконечной выборке). Стоит заметить, что из несмещенности оценки не следует ее состоятельность и наоборот.

Отметим, что главный фокус работы – оценка смещения коэффициентов логистической регрессии. Как было сказано в основном тексте работы, мы получаем оценки параметров в классической логистической регрессии за счет метода максимального правдоподобия. Этот оценщик дает состоятельные, но смещенные оценки.