

Прикладная эконометрика, 2025, т. 80, с. 93–116.

Applied Econometrics, 2025, v. 80, pp. 93–116.

DOI: 10.22394/1993-7601-2025-80-93-116

EDN: SXRRPY

Е.С. Котырло¹

Некоторые подходы к оценке воздействия различной интенсивности

В регрессионной модели существует проблема интерпретации оценки регулирующего воздействия, ассоциирующегося с непрерывной величиной. Это связано с тем, что величина воздействия, как правило, тесно связана с наблюдаемыми и ненаблюдаемыми характеристиками объекта, на который это воздействие направлено. Статья представляет обзор некоторых подходов к оценке непрерывного воздействия на основе: 1) функции отклика на непрерывное воздействие; 2) оценки непрерывного воздействия на панельных данных; 3) использования методов машинного обучения. Методы поясняются эмпирическими примерами, иллюстрирующими зависимость показателей безработицы от индекса возможности работать удаленно в период COVID-19.

Ключевые слова: обобщенный метод соответствия; функция отклика на непрерывное воздействие; метод разности разностей; оценка политики; несмещенный алгоритм машинного обучения; условный средний эффект воздействия.

JEL classification: C18; C21; C23; C87.

1. Введение

Необходимость оценки воздействия, ассоциирующегося с непрерывной величиной, нередко возникает в экономических задачах. Можно привести несколько примеров.

1. В 2025 г. администрацией Президента США были введены новые тарифы на импорт, дифференцированные по странам. Можно ожидать, что и воздействие на их экономики будет разной силы.

2. Санкционные режимы, инициированные в отношении ряда стран, различаются по тому, какие санкции введены. Следовательно, экономический эффект также должен быть различен.

3. В период коронавирусных ограничений получил распространение режим удаленной работы. Многим это помогло сохранить заработки. Но воспользоваться «удаленкой» смогли только те, кто часть своей работы выполнял за компьютером. То, как заработки меняются в зависимости от величины индекса, измеряющего интенсивность использования нового режима работы для разных профессий (teleworkability), может рассматриваться как пример такого непрерывного воздействия.

4. Инвестиции в региональную экономику в зависимости от их размера могут иметь противоречивое воздействие на благосостояние жителей региона. Оно может быть как

¹ Котырло Елена Станиславовна — Национальный исследовательский университет «Высшая школа экономики», Москва; ekotyrllo@hse.ru.

положительным — благодаря созданию рабочих мест, так и отрицательным — вследствие возросшей нагрузки на окружающую среду.

Несмотря на то что многие реформы и регулирующие воздействия логично интерпретировать как непрерывные величины, эконометрические методы для их оценки развиты относительно меньше по сравнению с подходами к оценке эффекта воздействия, описываемого бинарной переменной. Но в последние годы это направление получило новое развитие. Статья представляет три подхода к оценке непрерывного воздействия: оценку функции отклика на непрерывное воздействие; оценку непрерывного воздействия на панельных данных; оценку с использованием методов машинного обучения.

Второй раздел знакомит читателя с понятием эффекта воздействия, проблемами его измерения и кратко представляет методы соответствия и разности разностей, которые получают развитие далее. Третий раздел знакомит читателя с функцией отклика на непрерывное воздействие и связанным с ней обобщенным методом соответствия. Четвертый раздел представляет оценку непрерывного воздействия на панельных данных на базе метода разности разностей, пятый демонстрирует применение методов машинного обучения к оценке непрерывного воздействия. Теоретические модели и подходы к вычислению показателей причинно-следственной связи и их доверительных интервалов сопровождаются эмпирическими примерами. Также даны ссылки на пакеты и отдельные коды для Stata и R, реализующие эти методы, которые помогут заинтересованному читателю расширить знания в данной области.

2. Измерение эффекта воздействия

Напомним, что средний эффект воздействия (average treatment effect, ATE) — это величина, показывающая изменение результирующей переменной вследствие некоторой программы, реформы (воздействия) и элиминирующая изменение результирующей переменной вследствие каких-то сторонних факторов. Таким образом, ATE демонстрирует причинно-следственную связь между воздействием D и результатом Y^2 . Если воздействие описывается бинарной переменной ($D_i = 1$ или $D_i = 0$), то эффект воздействия выражается как

$$ATE = E\{Y_i(1)\} - E\{Y_i(0)\}, \quad (1)$$

где $Y_i(1), Y_i(0)$ — исходы при наличии/отсутствии воздействия. Если объекты $i = 1, \dots, n$ в контрольной группе и группе воздействия были абсолютно идентичны, то достаточно сравнить средние значения Y по группам. Если объекты в группах различаются только в наблюдаемых характеристиках X , но нет ненаблюдаемых причин, почему тот или иной объект оказался в контрольной группе или группе воздействия (*условная независимость воздействия и результата*, conditional independence assumption, CIA),

$$Y(1), Y(0) \perp D \mid X, \quad (2)$$

² Здесь и далее предполагается, что результат воздействия на объект i не зависит от того, получили ли другие объекты воздействие или нет (independence assumption).

то АТЕ можно оценить, устранив влияние X . Для этого достаточно проконтролировать влияние переменных X , включив их в качестве регрессоров в регрессионную модель:

$$Y_i = x_i' \beta_0 + D_i x_i' \beta_1 + \varepsilon_i, \quad (3)$$

где x_i — вектор характеристик объекта i , β — соответствующий вектор коэффициентов, ε_i — ошибка регрессии. Тогда оцениваемый эффект $ATE = E\{x_i' \beta_1\}$.

Другой популярный способ оценки АТЕ при условии CIA — использовать *метод соответствия* (propensity score matching, PSM) (Rosenbaum, Rubin, 1983). Суть метода PSM состоит в том, чтобы сопоставить контрольную и группу воздействия не по вектору наблюдаемых характеристик X , а по одному показателю — вероятности воздействия как функции X :

$$\Pr(D_i = 1 | x_i) = p(x_i). \quad (4)$$

Индекс соответствия $p(x_i)$ (propensity score) позволяет поставить в соответствие объектам из группы воздействия подходящие по характеристикам объекты контрольной группы, и наоборот. Параметры модели (4) обычно оцениваются с использованием пробит- или логит-регрессии. Для применения метода необходимо соблюдение условия CIA и *условия перекрывающихся выборок* (overlap assumption):

$$0 < \Pr(D = 1 | X = x) < 1. \quad (5)$$

Последнее свойство требует, чтобы ни для каких объектов $i = 1, \dots, n$ вероятность воздействия $\Pr(D_i = 1 | x_i)$ не была равна 0 или 1. Те объекты, для которых не нашлись соответствующие в группе сравнения, отбрасываются³. Доля объектов, включенных в сравнение (common support, CS), должна быть достаточно велика, чтобы сделать достоверные выводы. Чтобы убедиться в достигнутом соответствии между сравниваемыми группами⁴, проводятся *тесты на сбалансированность* (balance test). Средний эффект воздействия АТЕ рассчитывается как взвешенное среднее разности между объектами, попавшими под воздействие, и объектами в контрольной группе (Abadie et al., 2004; Cerulli, 2015, ch. 2).

Если объекты различаются не только по наблюдаемым, но и по ненаблюдаемым характеристикам, оба вышеприведенных метода не подходят. Проблема может быть решена методом разности разностей (difference-in-differences, DD) — популярным подходом к оценке эффекта, когда ненаблюдаемые характеристики определяют, куда попадет объект: в группу воздействия или в контрольную группу. Для оценки требуется наблюдение за объектами в течение хотя бы двух периодов или наличие повторяющихся кросс-секционных данных. Основное для реализации метода условие параллельного тренда (parallel trend assumption, PTA) — это предположение о том, что если бы не воздействие, то различие $E\{Y_{it}(1) - Y_{it}(0) | D_{it} = 0\}$ оставалось бы неизменным. Иными словами, ненаблюдаемая компонента объекта i неизменна во времени $t = 1, \dots, T$. Однако наличие ненаблюдаемой компоненты не позволяет оценить АТЕ, см., например, (Котырло, 2024) или (Roth et al., 2023). Доступной оценкой эффекта

³ Здесь не обсуждается проблема сокращения выборки из-за потери части наблюдений, а также возможное систематическое удаление несоответствующих объектов в двух группах, которое приводит к смещению оценки эффекта (King, Nielsen, 2019).

⁴ Более подробно о подборе соответствующих объектов см. (Cerulli, 2015, ch. 2).

воздействия является средний эффект воздействия на группу воздействия (average treatment effect on the treated, ATT):

$$ATT = E\{Y_{it}(1) - Y_{it}(0) | D_{it} = 1\}. \quad (6)$$

На основе приведенных методов оценки воздействия далее описываются подходы к оценке непрерывного воздействия.

3. Обобщенный метод соответствия и функция отклика на непрерывное воздействие

Теперь предположим, что воздействие непрерывно в некотором интервале $d_{\min} \leq d_i \leq d_{\max}$ или принимает достаточно много значений, поэтому логично рассматривать это воздействие как непрерывную величину. В оценке непрерывного воздействия в регрессионной модели возникает проблема интерпретации этого коэффициента. Прежний подход к его интерпретации, как изменение на одну единицу воздействия, явно не подходит, поскольку переменная может меняться в небольшом интервале, например (0,1). Даже если воздействие меняется на большем интервале, нет причин ожидать, что переход от воздействия d_k к $d_k + 1$ будет давать такой же эффект, что и переход от d_m к $d_m + 1$ для произвольных различных d_k и d_m . То есть линейная регрессионная модель явно не подходит для интерпретации непрерывного воздействия. Что еще более важно, это то, что величина воздействия, как правило, тесно связана с наблюдаемыми характеристиками X . Это также затрудняет интерпретацию эффекта воздействия. Например, помощь международного валютного фонда странам, попавшим в долговую зависимость, зависит от величины долга, ВВП страны, степени ее индустриализации и т.д. Также и сами возможности выйти из долговой ямы зависят от этих характеристик. Трудно понять, как предоставленная финансовая помощь повлияла бы на экономику, если характеристики страны были бы другими. Возникает задача — описать зависимость исхода Y от величины воздействия d , когда влияние X устранено как из результирующей переменной, так и из самого воздействия.

Первое решение было предложено нобелевским лауреатом 2021 г. Г. Имбенсом (Imbense, 2000). Концепция получила название *функции отклика на непрерывное воздействие* (dose-response function, DRF). В ее оценке используется *обобщенный метод соответствия* (generalised propensity score, GPS). Кратко опишем базовые шаги этого подхода.

Imbens (2000) в условиях CIA и перекрывающихся выборок предложил измерить DRF, элиминируя прямое влияние X заменой на обобщенный индекс соответствия. Модель GPS описывает вероятность воздействия в зависимости от величины воздействия d и набора переменных x :

$$r(d, x) \equiv \Pr(D = d | X = x). \quad (7)$$

Далее он предложил оценить зависимость исхода Y как функцию $\beta(d, r)$ только величины воздействия d и обобщенного индекса соответствия $r(d, X)$. Усреднение функции по $r(d, X)$ дает DRF — функцию отклика на непрерывное воздействие $\mu(d)$:

$$\mu(d) = E\{Y(d)\} = E\{\beta(d, r(d, X))\},$$

где $\beta(d, r) = E\{Y(d) | r(D, X) = r\} = E\{Y | D = d, R = r\}.$ (8)

Как видно из формулы, характеристики объекта X используются только при расчете вероятности воздействия, но не влияют напрямую на результирующую переменную Y , что решает проблему измерения причинно-следственной связи между воздействием и результатом. Функция $\mu(d)$ демонстрирует ожидаемый отклик в Y на величину воздействия $D = d$ в сравнении с отсутствием воздействия $d = 0$. Hirano, Imbens (2004) показывают, что при корректной спецификации модели для $r(d, X)$ и $\beta(d, r)$ оценка $\mu(d)$ состоятельна со скоростью сходимости $\sqrt{N}$, где N — объем выборки, и асимптотически нормальна. Для расчета стандартных ошибок и доверительных интервалов предлагается бутстрапирование.

3.1. Алгоритм оценки DRF

Оценка GPS на первом шаге

Как видно из определения DRF, для ее оценки необходимо на первом шаге оценить обобщенный индекс соответствия $r(d, x)$. Аналогично методу PSM, для достижения соответствия между сравниваемыми группами, помимо самой оценки $r(d, x)$, требуется определение интервала значений $r(d, x)$, где выполняется условие перекрывающихся выборок, а также проверка сбалансированности групп воздействия и контрольной группы.

Обобщенный индекс соответствия $r(d, x)$ может быть оценен в рамках *обобщенной линейной модели* (generalized linear model, GLM), позволяющей расширить предположения о распределении ошибок на широкий класс распределений⁵ и связывающей дисперсию расчетного значения с самим значением с помощью некоторой *функции связи* (link function) (Bia, Mattei, 2008):

$$r(d_i | x_i) \sim \psi\{g(\gamma, x_i), \theta\}, \quad (9)$$

где ψ — функция плотности выбранного распределения, g — функция связи, γ — вектор оцениваемых параметров, θ — коэффициент приведения.

В достижении соответствия сравниваемых объектов сложность состоит в том, что воздействие — это непрерывная величина. Следовательно, соответствие должно соблюдаться для каждого значения в отдельности, что невозможно. Flores et al. (2012) предлагают подход к расчету перекрывающейся выборки, который используется в большинстве алгоритмов для DRF. Он основан на подборе сравниваемых объектов по оцененной вероятности на некоторых интервалах равной длины $\{(d_1, d_2), \dots, (d_{K-1}, d_K)\}$, где $d_1 = d_{\min}$, $d_K = d_{\max}$ ⁶. Алгоритм расчета перекрывающейся выборки для интервала k (CS_k) состоит в том, что нижняя граница вероятности попадания в интервал равна максимуму из нижних границ вероятности для объектов самого интервала и всех объектов, не включенных в интервал. Наоборот, верхняя граница вероятности попадания в конкретный интервал равна минимуму из верхних границ вероятности для объектов самого интервала и всех объектов, не включенных в интервал:

$$CS_k = \left\{ i : \hat{r}_i^k \in \left[\max \left(\min_{j: Q_j = q_k} \hat{r}_j^k, \min_{j: Q_j \neq q_k} \hat{r}_j^k \right), \min \left(\max_{j: Q_j = q_k} \hat{r}_j^k, \max_{j: Q_j \neq q_k} \hat{r}_j^k \right) \right] \right\}. \quad (10)$$

⁵ В пакете *causaldrf* из R возможны варианты нормального, пуассоновского, отрицательного биномиального и гамма распределения.

⁶ В пакетах для R и Stata интервалы могут быть заменены квантилями распределения.

Затем по всем интервалам CS_k определяется интервал, общий для выборки, как их пересечение $CS = \bigcap_{k=1}^K CS_k$. В сравниваемые группы попадают те объекты, которые соответствуют этому условию, остальные отбрасываются. Для иллюстрации алгоритма определения перекрывающихся выборок рассмотрим небольшой пример, представленный в табл. 1. Пусть воздействие варьируется в пределах от 0 до 10. Предположим, что для 5 равномерных интервалов получены минимальное и максимальное значения $\hat{r}_{k,\min}, \hat{r}_{k,\max}$. Тогда перекрывающиеся выборки для каждого интервала CS_k дают итоговый интервал $CS = (0.2, 0.73)$.

Таблица 1. Пример расчета перекрывающейся выборки для непрерывного воздействия

Воздействие	[0, 2]	(2, 4]	(4, 6]	(6, 8]	(8, 10]
Интервал k	1	2	3	4	5
$\hat{r}_{k,\min}, \hat{r}_{k,\max}$	(0.12, 0.78)	(0.08, 0.81)	(0.18, 0.73)	(0.15, 0.84)	(0.2, 0.83)
CS_k	(0.2, 0.78)	(0.2, 0.81)	(0.2, 0.73)	(0.2, 0.83)	(0.2, 0.83)

Для тестирования сбалансированности может использоваться t -тест, предложенный Hirano, Imbens (2004). Суть теста состоит в том, чтобы убедиться, что средние по интервалам для каждой переменной в отдельности не различаются на статистически значимом уровне. Эта процедура требует довольно много вычислений. Гораздо проще для реализации тест отношения правдоподобия, предложенный Flores et al. (2012). Модель без ограничений описывает зависимость воздействия d_i от объясняющих переменных x_i , рассчитанного обобщенного индекса соответствия, его квадрата и куба: $\hat{r}_i, \hat{r}_i^2, \hat{r}_i^3$. Модель с ограничениями использует в качестве объясняющих переменных только $\hat{r}_i, \hat{r}_i^2, \hat{r}_i^3$. Незначимость различий в качестве моделей свидетельствует о достижении сбалансированности. Иными словами, индекс соответствия хорошо объясняет воздействие и без включения переменных x_i .

Модель для результирующей переменной

На следующем шаге строится модель для контрафактуального Y . Рассмотрим три популярных подхода. Hirano, Imbens (2004) предлагают использовать *полиномиальную модель* с включением степеней d и r , а также их произведения. Например, квадратичная модель выглядит как

$$E\{Y_i | d_i, r_i\} = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^2 + \alpha_3 \hat{r}_i + \alpha_4 \hat{r}_i^2 + \alpha_5 d_i \hat{r}_i, \quad (11)$$

где $\hat{r}_i$ рассчитан на первом шаге. Эта модель может быть оценена простой линейной регрессией. Затем функция отклика на воздействие DRF рассчитывается как среднее по i для каждого уровня d (Hirano, Imbens 2004):

$$\hat{\mu}(d) = \frac{1}{N} \sum_{i=1}^N [\hat{\alpha}_0 + \hat{\alpha}_1 d + \hat{\alpha}_2 d^2 + \hat{\alpha}_3 \hat{r}(d, x_i) + \hat{\alpha}_4 \hat{r}(d, x_i)^2 + \hat{\alpha}_5 d \cdot \hat{r}(d, x_i)^2]. \quad (12)$$

Недостатком такого приближения является жестко заданная функциональная форма DRF. Как правило, в граничных значениях d она дает существенные искажения (Bia et al., 2014).

Более сглаженное приближение дает *полупараметрическая оценка с использованием сплайнов*⁷. Модель, основанная на регрессионных сплайн-функциях со штрафами (additive penalized spline estimator), обеспечивает более гладкую оценку контрафактуального Y и меньше искажает его значения для граничных значений d (Bia et al., 2014). Пример такой функции с линейными членами:

$$\begin{aligned} E\{Y_i | d_i, \hat{r}_i\} = & \alpha_0 + \alpha_1 d_i + \alpha_2 \hat{r}_i + \alpha_3 t_i \hat{r}_i + \sum_{k=1}^{K^d} u_k^d (d_i - k_k^d)_+ + \sum_{k'=1}^{K^r} u_{k'}^r (\hat{r}_i - k_{k'}^r)_+ + \\ & + \sum_{k=1}^{K^d} v_k^d \hat{r}_i (d_i - k_k^d)_+ + \sum_{k'=1}^{K^r} v_{k'}^r d_i (\hat{r}_i - k_{k'}^r)_+ + \sum_{k=1}^{K^d} \sum_{k'=1}^{K^r} v_{kk'}^{dr} (d_i - k_k^d)_+ (\hat{r}_i - k_{k'}^r)_+, \end{aligned} \quad (13)$$

где $k_1^d < \dots < k_{K^d}^d$ и $k_1^r < \dots < k_{K^r}^r$ — узлы решетки, в которых берутся значения d и r для расчета. Нижний индекс «+» означает, что $(z)_+ = \max(0; z)$, т.е. в расчете используются только положительные значения. $\alpha_0, \dots, \alpha_3, u_k^d, u_{k'}^r, v_k^d, v_{k'}^r, v_{kk'}^{dr}, k=1, \dots, K^d, k'=1, \dots, K^r$ — оцениваемые параметры регрессии. Таким образом, модель имеет коэффициенты, как постоянные для всех наблюдений, так и зависящие от того, к какому узлу наблюдение ближе. Такая регрессия может быть оценена как многоуровневая модель со случайными коэффициентами. Для применения этой модели размер данных должен быть достаточно велик (Bia et al., 2014).

Flores et al. (2012) предлагают *непараметрическую оценку* $\mu(d)$, основанную на сглаживании с использованием *ядерной функции*⁸ и использующую веса важности (importance weights, IW). В этом методе расчета локальной регрессии используются не $\hat{r}_i$, а $\hat{r}_i^d$ — оценки GPS для уровня d :

$$\hat{\mu}(d) = \frac{\sum_{i=1}^N \tilde{K}_{h,X}(d_i - d) \cdot y_i}{\sum_{i=1}^N \tilde{K}_{h,X}(d_i - d)}. \quad (14)$$

Здесь $\tilde{K}_{h,X}(d_i - d) = K_h(d_i - d) / \hat{r}_i^d$, K_h — ядро со стандартными свойствами, h — ширина полосы, $h \rightarrow 0$, $Nh \rightarrow \infty$, $N \rightarrow \infty$.

Перечисленные методы позволяют получить представление о подходах к оценке DRF, но отнюдь не исчерпывают их многообразие. Для более подробного знакомства см. (Galagate, 2016; Kennedy et al., 2017; Zhao et al., 2020). В частности, Kennedy et al. (2017) предлагают метод оценки с двойной устойчивостью, который допускает некорректную спецификацию одной из моделей: GPS или DRF.

Представляет интерес не только оценка DRF как эффекта в сравнении с нулевой дозой, но и изменение результирующей переменной при некотором изменении воздействия (DRF derivative). Эта величина зависит от уровня воздействия d_j . Zhao et al. (2020) предлагают расчет

⁷ Сплайн — это кусочно-полиномиальная функция. В каждом интервале, задаваемого узлами, она описывается полиномом выбранной степени. Смежные полиномы для определенного узла возвращают одинаковое значение функции, что обеспечивает ее непрерывность. Более подробно в (Анатольев, 2009).

⁸ Ознакомиться с непараметрическими регрессиями можно в (Анатольев, 2009).

такого изменения для воздействия d_j относительно предыдущего и последующего как усредненное отношение изменений в результирующей переменной к изменению в воздействии:

$$\frac{1}{2} \left[\frac{\hat{\mu}(d_{j+1}) - \hat{\mu}(d_j)}{d_{j+1} - d_j} + \frac{\hat{\mu}(d_j) - \hat{\mu}(d_{j-1})}{d_j - d_{j-1}} \right], \quad (15)$$

тогда как в (Flores et al., 2012) используется только первое слагаемое.

Расчет стандартных ошибок

Расчет стандартных ошибок и доверительных интервалов базируется либо на *бутстраповании* оценок, либо на *функции влияния* (influence function). Первый подход предполагает многократное повторение процедуры оценивания на случайных подвыборках из имеющейся выборки. Точечные (pointwise) оценки $\hat{\mu}(d)$ для каждого d , таким образом, получаются многократно. Это позволяет оценить дисперсию каждого значения и рассчитать соответствующие доверительные интервалы (Cameron, Trivedi, 2005, ch. 11). Второй подход нередко применяется для расчета ковариационной матрицы ошибок для непараметрических оценок (Ichimura, Newey, 2022; Wasserman, 2006, ch. 2).

3.2. Ограничения методов оценки DRF

Zhao et al. (2020) методом Монте-Карло сравнивают подходы по качеству подгонки для некоторой известной аналитической функции непрерывного воздействия. Результаты демонстрируют чувствительность модели Hirano, Imbens (2004) к спецификации DRF. Метод, предложенный Flores et al. (2012), демонстрирует нестабильность относительно выбора ширины полосы. Полупараметрическая оценка с использованием сплайн-функций со штрафами лучше описывает эффект непрерывного воздействия. Однако, по мнению Zhao et al. (2020), все эти методы могут давать довольно плохую оценку DRF на реальных данных. Одна из проблем связана с тем, что вероятность воздействия зависит от уровня воздействия. Поскольку определенный уровень воздействия получают не все объекты, а объекты с некоторыми специфическими характеристиками, оценка DRF представляет собой совокупность оценок для разных подвыборок. Другая проблема связана с корректностью спецификации модели для вероятности получения воздействия GPS и предполагаемого распределения ошибок этой модели.

3.3. Реализация в статистических пакетах

Оценка непрерывного воздействия на основе DRF с расчетом GPS на первом шаге реализована в Stata пакетами *drf* (Bia et al., 2014), *doseresponse* (Mattei, Bia, 2009), *glmdose* (Guardabascio, Ventura, 2014), а в R можно воспользоваться пакетом *causaldrf* (Galagate, Schafer, 2022).

В частности, *drf*, *doseresponse* и *glmdose* оценивают GPS; определяют область CS согласно Flores et al. (2012), тестируют сбалансированность групп и оценивают DRF и производную DRF тремя перечисленными выше методами. Пакеты *drf* и *glmdose*, в отличие

от *doseresponse*, допускают более широкий класс распределений для переменной воздействия. Результаты иллюстрируются графически, рассчитываются доверительные интервалы. Пакет *causaldrf* для R более гибок в выборе шагов. Можно отдельно от оценки DRF разобратся с GPS и перекрывающимися выборками, постепенно улучшая качество подгонки. А затем перейти к оценке DRF. Удобство работы с пакетом *causaldrf* в том, что оценки, полученные разными методами, хранятся отдельно. Можно усовершенствовать оценки постепенно. Тогда как в методах для Stata все процедуры выполняются одной командой, и если на каком-то шаге не достигнута требуемая сходимость, то приходится повторять всю процедуру сначала, на что может уйти довольно много времени. К сожалению, в пакете *causaldrf* нет опций для расчета стандартных ошибок и доверительных интервалов. Также пакет не рассчитывает изменение DRF по уровню дозы.

Пример 1. Оценка роста численности безработных как функции индекса возможности удаленной работы.

В период ограничений, связанных с COVID-19, в России, как и во многих других странах, были усилены меры помощи безработным. Помимо продления срока выплаты пособий по безработице и существенного увеличения его размера, процедура подачи заявлений была перенесена на онлайн-платформу. Для регистрации в статусе безработного в этот период не требовалось подтверждения занятости и потери работы, что значительно упростило процесс регистрации и добавило дополнительные стимулы для получения статуса безработного. В то же время онлайн-регистрация стала барьером в получении статуса для тех, кто не имел навыков работы в Интернете. Как показывают исследования, новые меры в отношении безработных привлекли тех, кто был занят неофициально или не работал ранее (Гимпельсон, 2022). Рост уровня безработицы практически не был связан со снижением уровня занятости (см. там же). Kotyrló (2024) исследует, в какой мере навыки работы в Интернете, которые можно измерить индексом возможности работать удаленно (*teleworkability index*, TW), были связаны с ростом численности безработных. Результаты показывают, что численность безработных в этот период больше выросла в тех профессиональных группах, которые характеризовались высоким индексом TW.

На основе тех же данных, что и в (Kotyrló, 2024), проанализируем связь роста численности безработных (месяц 2020 г. к соответствующему месяцу 2019 г.) в зависимости от классификации профессии по индексу TW в период апрель—август 2020 г. Здесь используется индекс, предложенный Sostero et al. (2020). Месячные данные Федеральной службы занятости населения (ФСЗН) (июль 2019 — декабрь 2020), агрегированы по профессии и региону обращения⁹. Период наблюдений — с января 2019 г. по декабрь 2020 г. По безработным доступны пол, возраст, образование, профессия на последнем месте работы, регион места обращения, зарплата, индекс интенсивности социального взаимодействия в профессиональной группе (F2F) (Sostero et al., 2020), месяц и год регистрации в качестве безработного, а также месяц и год, когда этот статус утрачивается. В качестве результирующей переменной рассмотрим изменение численности безработных (месяц 2020 г. к соответствующему месяцу 2019 г.) в логарифмах в период с апреля по август 2020 г. Индекс TW принимает значения от 0 до 1 и рассматривается как непрерывное воздействие. Наблюдения со значениями TW 0 или 1 не включены в рассмотрение согласно алгоритму расчета DRF.

⁹ Доступ предоставлен Инфраструктурой научно-исследовательских данных (ИНИД, <https://data.rcsi.science/>).

Предполагается, что индекс TW как непрерывная величина является функцией наблюдаемых характеристик безработного — пола, возраста, уровня образования и региона обращения. Поэтому простое включение индекса TW в регрессию не позволяет интерпретировать коэффициент при этой переменной как эффект воздействия. Для оценки использовался пакет *R causaldrf*. На основе характеристик объекта оценивается GPS и формируются группы сравнения. В таблице 2 приводится описательная статистика для данных после процедуры мэтчинга.

Таблица 2. Описательная статистика по профессиональным группам

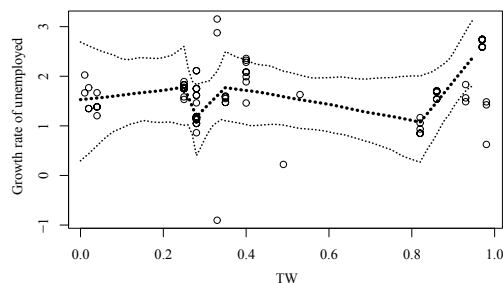
	Рост численности безработных (%)	TW	Возраст (лет)	Доля женщин (%)	F2F	Зарплата (руб.)	Образование (лет)
Min	–110	0.01	20	0	0.31	20000	9
Q1	119	0.28	40	0	0.4	51792	11
Median	159	0.35	40	0	0.5	62575	12
Mean	163	0.523	39.9	23	0.529	64152	12.01
Q3	208	0.86	40	41	0.66	77661	12.79
Max	330	0.98	60	100	0.8	146667	16

Результат представлен на рис. 1. Метод, основанный на сплайн-функциях со штрафами (Bia, Mattei, 2008), дает кривую, демонстрирующую наиболее высокий рост численности безработных в профессиях, тесно связанных с использованием Интернета. Среднее значение эффекта воздействия $ATE_{Bia, Mattei} = 1.534$. Параметрическая оценка DRF с использованием регрессии с линейными и квадратичными членами (Hirano, Imbens, 2004) не показывает зависимости между ростом численности безработных и индексом TW, $ATE_{Hirano, Imbens} = 1.527$. Оценка DRF с использованием ядерной функции (Flores et al., 2012) с $ATE_{Flores et al.} = 1.639$ демонстрирует неожиданные «всплески» в значениях TW, которые наиболее часто встречаются в выборке, а также меньшую определенность результатов для граничных значений TW. Этот метод демонстрирует несколько больший рост численности безработных при более высоких значениях TW. В целом результаты сходны с полученными в (Kotyrlo, 2024, Table II, Figure 3), но при этом различаются не только методы оценки, но и группы сравнения. В отличие от (Kotyrlo, 2024), в приведенном примере не включались профессиональные группы с граничными значениями индекса TW. Кроме того, различны периоды анализа, более широкий (февраль 2019 — декабрь 2020) был в (Kotyrlo, 2024).

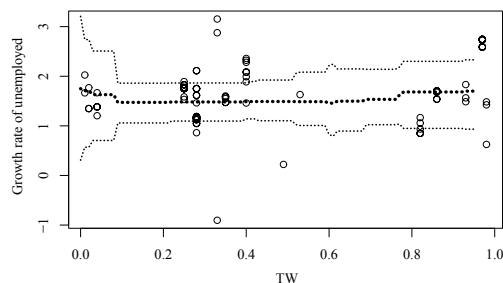
4. Оценка непрерывного воздействия на панельных данных

Панельные данные дают возможность элиминировать неслучайность отбора в группы воздействия и контрольную, если факторы, его определяющие, не меняются во времени. Метод оценки воздействия *разности разностей* основан на этом свойстве, обзор по применению DD на русском языке см. в (Котырло, 2024).

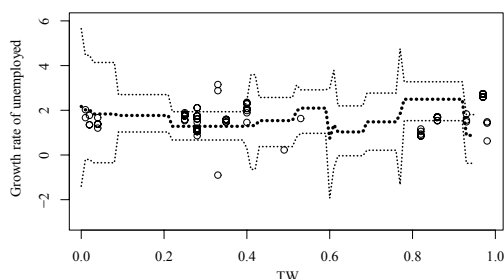
Callaway et al. (2024b) предлагают адаптировать дизайн разности разностей для оценки непрерывного воздействия. Воздействие для различных групп может начинаться в разные



а) Метод оценки DRF-GPS, предложенный Bia, Mattei (2008)



б) Метод оценки DRF-GPS, предложенный Hirano, Imbens (2004)



в) Метод оценки DRF-GPS, предложенный Flores et al. (2012)

Рис. 1. Оценка численности безработных в зависимости от индекса TW
в период апрель—август 2020 г.
(95% доверительные интервалы получены бутстрапом)

моменты времени и длиться несколько периодов. Сгруппируем объекты по моменту начала воздействия $g \in \{2, \dots, T, \infty\}$, где $g = \infty$ для объектов контрольной группы. Таким образом, g_i — это время начала воздействия для объекта i . Обозначим как $d_i \in [0, d_H]$ дозу, которую объект i получил в начале воздействия. Метод использует стандартные DD условия.

- (1) Данные $\{Y_{i,1}, \dots, Y_{i,T}, d_i, g_i\}$, $i = 1, \dots, n$ независимы и одинаково распределены.
- (2) Воздействие непрерывно, и существует группа объектов, которые не испытывали воздействие (контрольная группа, для которой доза равна 0)¹⁰. Прирост $\Delta Y_i = Y_i - Y_{i-1}$ — непрерывно дифференцируем по d .
- (3) Для любого объекта i из группы воздействия g_i интенсивность d_i постоянна во времени, начиная с момента g_i (staggered treatment constant in time).
- (4) Для всех объектов начало воздействия неожиданно¹¹.
- (5) Объекты до начала воздействия и не попавшие под воздействие следуют параллельными трендами.

Обозначим потенциальное значение выходной переменной для отдельного объекта i в момент времени t через $Y_{i,t}(g, d)$ для объектов из группы воздействия g с интенсивностью

¹⁰ Callaway et al. (2024b) развивают основные идеи с этим условием, а затем расширяют их на случай, когда контрольной группы с нулевой дозой в данных нет.

¹¹ Можно ослабить это условие и считать, что период ожидания один и тот же. В этом случае фактическое начало воздействия следует сместить на начало ожидания.

воздействия d , и $Y_{i,t}(0)$ — для объектов из контрольной группы. Определим $ATT(g, t, d)$ (group-time-dose-specific average treatment effect) как средний эффект воздействия для группы воздействия в момент t в группе g с дозой d в сравнении с нулевой дозой в этот же момент времени:

$$ATT(g, t, d) = E\{Y_t(g, d) - Y_t(0) | D = d\}. \quad (16)$$

$ATT(g, t, d)$ можно усреднить (с некоторыми весами) по времени или группе, или по обоим компонентам. Это позволяет оценить как эффект воздействия в динамике, так и групповые эффекты, и получить общее для всех объектов и моментов времени значение АТТ.

Другим показателем причинно-следственной связи между непрерывным воздействием и результатом является средний каузальный отклик

$$ACR(g, t, d) = \left. \frac{\partial E[Y_t(g, \tilde{d})]}{\partial \tilde{d}} \right|_{\tilde{d}=d}. \quad (17)$$

Этот показатель демонстрирует предельный эффект для уровня воздействия d для всех объектов из группы g в момент времени t . Он позволяет понять, какой уровень воздействия существенно меняет результат (Callaway et al., 2024b).

Основная идея подхода к оценке непрерывного воздействия на панельных данных состоит в том, чтобы отдельно исследовать: 1) эффект воздействия в динамике, когда оно ограничено некоторым интервалом значений $d_L \leq d \leq d_H$; 2) эффект воздействия как DRF в некотором временном интервале.

Эти подходы предполагают два разных метода измерения.

Подход 1. Эффект воздействия в динамике

Когда воздействие принимает значения в некотором интервале $d_L \leq d \leq d_H$, Callaway et al. (2024b, c) предлагают сопоставить ему бинарную переменную: $D_i = \begin{cases} 1, & \text{если } d_L \leq d_i \leq d_H, \\ 0, & \text{если } d_i = 0. \end{cases}$

В этом подходе игнорируется изменение интенсивности внутри каждой групп. Для группы с началом воздействия g к моменту времени t средний эффект воздействия для получивших воздействие можно записать как

$$ATT(g, t) = E\{Y_t - Y_{g-1} | D = 1\} - E\{Y_t - Y_{g-1} | U_{\max\{t, g-1\}} = 1, D = 1 \cup D = 0\}. \quad (18)$$

Здесь $U = I\{g > t \mid D = 0\}$. Как видно из формулы, этот эффект для группы g не для некоторого момента времени, а для периода от начала воздействия (event-study analysis). Обобщим (18) по всем группам воздействия g к эффекту на некоторый период от начала воздействия $ATT(e)$, где $e = t - g$. Здесь необходимо условие, что объект был под воздействием в периоды $g + e \in [2, T]$:

$$ATT(e) = E\{ATT(g, g + e) | g + e \in [2, T], D = 1\}. \quad (19)$$

Этот подход позволяет применить DD и те методы, которые для этого разработаны (Callaway, Sant'Anna, 2021). Callaway et al. (2024b) демонстрируют, что полученная оценка не будет страдать от использования отрицательных весов¹² в расчете эффекта.

Алгоритм оценки непрерывного воздействия в динамике

1. На первом шаге регрессия¹³ (Callaway, Sant'Anna, 2021) Y на X по всем периодам оценивается на объектах контрольной группы. Контрольная группа формируется из объектов, не попавших под воздействие (never treated, NT), либо *еще* не попавших под воздействие (not-yet-treated, NYT). Затем по предсказанным значениям $\hat{Y}$ рассчитывается $m_{g,t}(X)$ как средний прирост $\hat{Y}$ на момент t в сравнении с моментом $g-1$, периодом, предшествующим воздействию:

$$m_{g,t}(X) = E\{\hat{Y}_t - \hat{Y}_{g-1}\}. \quad (20)$$

2. Для тех, кто попал под воздействие в группе g на момент времени t , рассчитывается первая разность по фактическим значениям:

$$\Delta Y_{g,t} = Y_t - Y_{t-g-1}. \quad (21)$$

3. Средний эффект воздействия для тех, кто попал под воздействие в группе g на момент времени t , рассчитывается как разность разностей:

$$ATT_{g,t} = E\{\Delta Y_{g,t}/n_g - m_{g,t}(X)\}, \quad (22)$$

где n_g — число объектов в группе g .

Достоинство метода состоит в упрощении модели. Он позволяет установить причинно-следственную связь, несмотря на сложность данных, имеющих панельную структуру, когда воздействие различной интенсивности для разных объектов начинается в разные периоды.

Очевиден недостаток метода. Он оценивает не непрерывное воздействие, а воздействие с интенсивностью в некотором интервале, выбор которого зависит от предпочтений исследователя и не всегда обоснован. Если внутри этого интервала есть какие-то необнаруженные резкие изменения в эффекте воздействия, то среднее по группе не позволит его обнаружить. Чтобы избежать эту проблему, можно проанализировать распределение дозы и графики зависимости выходной переменной от дозы. Но графический анализ не просто реализовать, когда воздействие начинается в разные моменты времени.

¹² Проблема оценки эффекта воздействия методом DD для воздействия, которое начинается для разных групп в разные моменты времени, посвящены работы (Borusyak et al., 2024; De Chaisemartin, D'Haultfoeuille, 2020; Goodman-Bacon, 2021). В них демонстрируется, что при стандартном методе оценки на панельных данных с фиксированными эффектами (TWFE) оцениваемый параметр будет представлять собой взвешенное среднее по группам. При этом веса могут быть как положительными, так и отрицательными, что непредсказуемо влияет на величину оцениваемого параметра. Следовательно, оценку параметра методом TWFE не следует рассматривать как показатель причинно-следственной связи между воздействием и результатом.

¹³ В (Callaway, Sant'Anna, 2021), помимо регрессии, на первом шаге предлагается также использовать метод оценки на основе обратных вероятностных весов и метод двойной устойчивости.

Подход 2. Изменение эффекта воздействия как отклика на интенсивность воздействия

В этом подходе опускается временная компонента. Вместо эффектов воздействия на момент времени, Callaway et al. (2024b) предлагают построить DRF, усредненную на некотором отрезке $[e_1, e_2]$ от начала воздействия. Например, отрезок можно ассоциировать с краткосрочным воздействием, взяв периоды $0 \leq e \leq 2$, или с долгосрочным, рассматривая эффект во временном отрезке $3 \leq e \leq 6$:

$$ATT_{e_1, e_2}(d) = \frac{1}{e_2 - e_1 + 1} \sum_{e=e_1}^{e_2} E\{ATT(g, g+e, d) | g+e \in [2, T], D=d\}. \quad (23)$$

Callaway et al. (2024b) показывают, что такое упрощение сводит оценку непрерывного воздействия к модели разности разностей с двумя периодами («до» и «после»). Для идентификации эффекта воздействия необходимо *более строгое предположение о параллельных трендах* (strong parallel trends assumption, SPTA). Это есть предположение о том, что среднее изменение результирующей переменной, если бы все испытали воздействие d , равно тренду, который соответствует группе, фактически испытавшей это воздействие. Иными словами, SPTA — это предположение параллельного тренда для любого воздействия относительно другого произвольного воздействия:

$$E\{Y_{after}(d) - Y_{before}(0)\} = E\{Y_{after}(d) - Y_{before}(0) | D=d\}. \quad (24)$$

Callaway et al. (2024b) предлагают более простую процедуру оценки DRF без расчета GPS, основанную на разложении переменной воздействия с использованием *базисных сплайнов* (B-splines).

Алгоритм оценки отклика на интенсивность воздействия в панельных данных

1. Для каждого объекта i , впервые получившего воздействие в момент g , рассчитывается первая разность для всех периодов от начала воздействия $e \in [e_1, e_2]$ к периоду, предшествующему воздействию, обозначенному здесь как (-1) :

$$\Delta Y_{i,g,e}(d) = Y_{i,g,e}(d) - Y_{i,-1,g}(d). \quad (25)$$

2. Контрольная группа формируется из объектов, не попавших под воздействие, для которых есть наблюдения как в момент $g-1$, предшествующий воздействию, так и в момент e . Для этих объектов рассчитывается разность

$$\Delta Y_{i,g,e}(0) = Y_{i,g,e}(0) - Y_{i,-1,g}(0), \quad (26)$$

которая затем усредняется по i , что дает $\Delta Y_{g,e}(0)$.

3. Рассчитывается среднее значение разности разностей для объекта i в интервале $[e_1, e_2]$:

$$ddY_{i,g}(d) = \frac{1}{e_2 - e_1} \sum_{e=e_1}^{e_2} [\Delta Y_{i,g,e}(d) - \Delta Y_{i,g,e}(0)]. \quad (27)$$

4. С помощью базисных сплайнов осуществляется декомпозиция непрерывного воздействия d на $\varphi_{j,K}(d)$, где K — степень полинома, а j — номер узла. Разложение зависит от выбранного числа узлов и степени полинома.

5. DD-DRF оценивается как предсказанные значения в OLS регрессии $ddY_i(d)$ на разложение $\varphi_{j,K}(d)$:

$$ddY_i = \sum_{j=1}^J \varphi_{j,K}(d) \beta_{j,K} + \varepsilon_i. \quad (28)$$

Callaway et al. (2024c) рассчитывают стандартные ошибки с использованием функции влияния. В расчете используются остатки, полученные OLS регрессией.

Алгоритм расчета DD-DRF пока не реализован как отдельный пакет. Callaway et al. (2024c) разместили код к статье в репозитории Mendeley (Callaway et al., 2024a), и он может быть легко адаптирован к конкретной задаче.

Пример 2. Влияние возможности работать удаленно на численность безработных в период COVID-19.

Рассмотрим пример, реализующий оба подхода. На основе тех же данных, что и в примере 1, проанализируем связь между числом зарегистрированных безработных и индексом возможности работать удаленно (TW) в соответствии с классификацией профессий по этому индексу. Для всех групп момент воздействия — апрель 2020 г. (период $g = 16$).

При первом подходе для анализа динамики эффекта воздействия индекс TW преобразуется в бинарную переменную на основе его медианного значения:

$$D_i^{high} = \begin{cases} 1, & \text{если } TW_i > median(TW), \\ 0, & \text{если } TW_i = 0, \end{cases} \quad \text{и} \quad D_i^{low} = \begin{cases} 1, & \text{если } 0 < TW_i \leq median(TW), \\ 0, & \text{если } TW_i = 0. \end{cases}$$

Поскольку данный подход реализует DD, можно применить развитие метода для счетных моделей¹⁴. Чтобы не усложнять изложение, воспользуемся линейной моделью для логарифма численности безработных. Проанализируем динамику отклика в численности безработных на высокие и низкие значения индекса TW в сравнении с $TW = 0$. Результаты оценки представлены на рис. 2. Как видно из графиков, численность безработных существенно выше в период коронавирусных ограничений, если профессия характеризовалась высоким индексом возможности удаленной работы (рис. 2а). Для групп с индексом $TW = 0$ и ниже медианного различия не значимы (рис. 2б). Также можно оценить изменение эффекта во времени. Судя по графику на рис. 2а, рост численности продолжался первые четыре месяца (до июля 2020 г. включительно), а затем численность безработных стабилизировалась. По сравнению с мартом 2020 г. (базовый период), превышение численности безработных с высоким индексом TW, или средний эффект воздействия на группу воздействия, составляло около 50 п.п.¹⁵. Хотя результаты показывают больший рост численности безработных в профессиональных группах, где режим удаленной работы мог получить широкое распространение, оценки получились ниже, чем в примере 1. Различие состоит в том, что в DD-DRF были включены граничные значения индекса и контрольная группа — группа с нулевым индексом TW.

¹⁴ Wooldridge (2023) указывает на то, что условие ПТА для счетных данных более логично сформулировать не как разность средних для групп сравнения, а как отношение средних, что можно сделать, используя экспоненциальное условное среднее.

¹⁵ Расчеты велись для логарифма численности безработных, для интерпретации полученное значение должно быть преобразовано по формуле $[\exp(\beta) - 1] \cdot 100\%$.

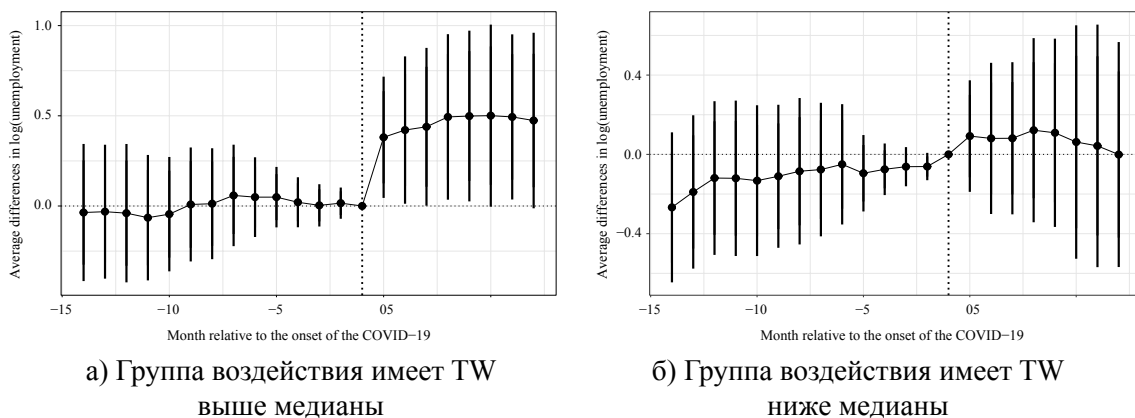


Рис. 2. Метод DD-DRF. Прирост численности безработных в сравнении с группой $TW = 0$ в динамике (отображены оценки с 95% доверительными интервалами)

Для иллюстрации второго подхода сгруппируем данные в краткосрочный эффект апрель — август 2020 г. и более поздний, сентябрь — декабрь 2020 г. Рисунок 3 показывает эффект возможности работать удаленно на прирост численности в краткосрочном и в более позднем периодах в сравнении с приростом численности безработных в профессиях с $TW = 0$. Как видно, такая классификация подтверждает роль высокого индекса TW в изменении численности безработных в период COVID-19.

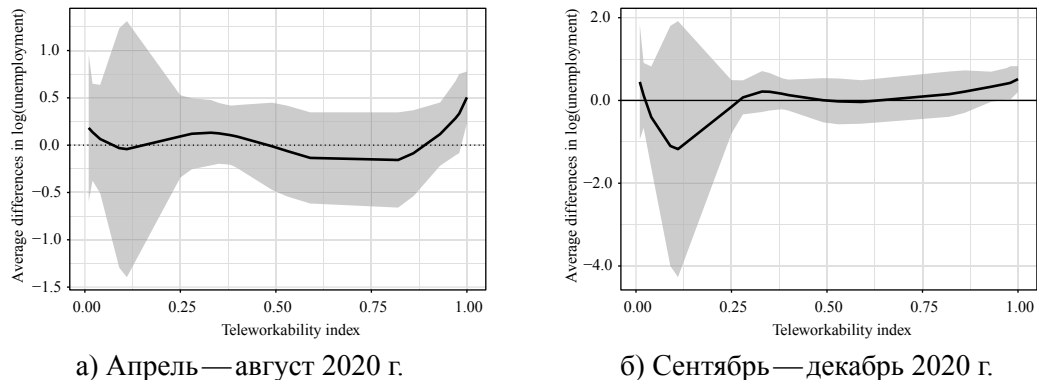


Рис. 3. Метод DD-DRF. Прирост численности безработных в сравнении с группой $TW = 0$ как отклик на интенсивность воздействия (заштрихованная зона — 95%-ный доверительный интервал)

5. Оценка непрерывного воздействия с использованием машинного обучения

Semenova, Chernozhukov (2021) предлагают оценивать непрерывное воздействие как вариант оценки условного среднего эффекта воздействия (conditional average treatment effect, CATE), который может применяться для более широкого класса задач. Подход использует двойное машинное обучение (double machine learning или debiased machine learning, DML).

Сама идея DML предложена ранее (Belloni et al., 2016) и приобрела в последние годы большую популярность.

В оценке непрерывного воздействия с использованием DML из зависимости между переменными интереса X и результирующей Y устраняется шумовая компонента $\eta(Z)$, функция от большого числа регрессоров Z . В отличие от вышеприведенных методов, помимо фактора переменной интенсивности этого воздействия, некоторой переменной X , в оценке непрерывного воздействия с использованием CATE вводится дополнительная бинарная переменная воздействия D . Задача остается прежней — найти эффект воздействия X на Y , элиминируя зависимость обеих переменных от Z :

$$g(x) = E\{Y(1) - Y(0) | X = x\},$$

где $Y(1)$ и $Y(0)$ — потенциальные значения Y для группы воздействия и контрольной группы при некотором значении X . Условия, накладываемые на воздействие, аналогичны сформулированным в разделе 2 условиям CIA и перекрывающихся выборок (4).

Алгоритм оценки DRF с использованием DML

Алгоритм состоит из двух шагов. На первом шаге методом ML¹⁶ (принципиальных различий нет каким — lasso, random forest, neural networks¹⁷) оцениваются зависимости D на Z и две модели для результирующей переменной при условии воздействия и без него, $Y(1)$ и $Y(0)$ на Z . Обозначим соответствующие функции как вероятности воздействия

$$\pi(Z) = \Pr(D = 1 | Z = z), \quad (29)$$

а для $Y(d)$

$$\vartheta(d, Z) = E\{Y | Z = z, D = d\}. \quad (30)$$

Затем рассчитывается ортогональный сигнал $Y(\eta)$ — преобразованная результирующая переменная, которая «очищена» от Z :

$$Y(\eta) = \vartheta(1, Z) - \vartheta(0, Z) + \frac{D[Y - \vartheta(1, Z)]}{\pi(Z)} - \frac{(1 - D)[Y - \vartheta(0, Z)]}{1 - \pi(Z)}. \quad (31)$$

Аналогично (Callaway et al., 2024c), на втором шаге оценки DRF получаются как предсказанные значения в OLS регрессии $Y(\eta)$ на B-сплайны переменной интереса X . Для оценки стандартных ошибок обсуждается (гауссовский) бутстрап и функция влияния.

Достоинство метода DML — это возможность установления причинно-следственной связи в условиях, когда переменных слишком много. Semenova, Chernozhukov (2021) демонстрируют, что полученные оценки DRF состоятельны и асимптотически нормальны. Однако на выборке оценка непрерывного воздействия может быть смещена. Непрозрачность ML алгоритмов затрудняет анализ причин смещения оценки. Наиболее частые причины — это пропущенные переменные, или включение в набор Z переменных, зависящих как от результирующей

¹⁶ С базовыми принципами методов машинного обучения можно ознакомиться, например, в (Shalev-Shwartz, Ben-David, 2014).

¹⁷ Semenova, Chernozhukov (2021) не акцентируют внимание на различиях этих методов. Но отметим, что lasso регрессия реализует линейное приближение, тогда как random forest и neural networks не накладывают ограничений на форму взаимосвязи.

переменной Y , так и от воздействия D (collider bias). DML особенно чувствителен ко второй причине смещения, поскольку эндогенные переменные часто являются сильными предикторами, а потому могут существенно влиять на оценку взаимосвязи между результатом и воздействием (Hünemund et al., 2023).

Авторы учебника Chernozhukov et al. (2024) сопровождают теоретический материал по оценке CATE кодами в Python и R (DML for CATE, 2025). Коды содержат подходы к оценке стандартных ошибок и точечных и общих доверительных интервалов¹⁸. Пример реализации кода можно найти в практических заданиях к учебнику.

Пример 3. Прирост численности зарегистрированных безработных как эффект возможности работать удаленно.

Воспользуемся теми же данными, что и в примере 1. Для реализации алгоритма сформируем бинарную переменную, ассоциирующуюся с воздействием. Например, это может быть округление индекса до 1 или 0 с использованием медианы в качестве порогового значения. Чтобы избежать полной предопределенности в расчете вероятности воздействия, бинарную переменную сформируем на основе частоты использования Интернет для работы в каждой профессиональной группе по данным РМЭЗ за 2019 г. Здесь, как в примере 2, контрольная группа состоит из профессий с нулевым индексом TW, а в группу воздействия включены профессии с индексом 1. Как и в примере 1, объясняющими переменными служат возраст, пол, категориальная переменная для региона и месяца, зарплата и индекс интенсивности социального взаимодействия в профессиональной группе. Непрерывные переменные — возраст, зарплата и F2F представлены разложением с помощью локального полинома для создания большего числа регрессоров. Аналогичный подход предложен в коде (DML for CATE, 2025).

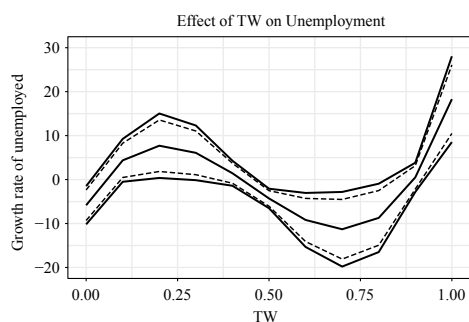


Рис. 4. Метод DML-DRF. Рост численности безработных (точечные и общие 95% доверительные интервалы отмечены соответственно пунктирной и сплошной линиями)

На рисунке 4 показан темп роста численности безработных (за апрель — август 2020), рассчитанный как разность логарифма численности на месяц 2020 г. и аналогичного показателя в соответствующем месяце 2019 г. В отличие от алгоритма для DRF, $TW = 0$ и $TW = 1$ не исключены из набора данных. На первом шаге используется lasso регрессия. Представленный результат демонстрирует значимый прирост безработных начиная с апреля 2020 г. по август 2020 г., в сравнении с аналогичным периодом 2019 г., в профессиях с индексом TW выше 0.9, что соответствует предыдущим результатам. Однако 95%-ные доверительные интервалы оказываются значительно шире в сравнении с методами, проиллюстрированными выше. Результаты похожи на оценки DRF, полученные для фиксированного

¹⁸ Chernozhukov et al. (2018) обсуждают проблему расчета доверительных интервалов при использовании ML и предлагают подход к точечной оценке доверительных интервалов, основанный на квантилях распределения оценки в каждой точке. Подход к оценке общего доверительного интервала изложен в (Semenova, Chernozhukov, 2021).

временного окна методом DD-DRF (рис. 3). Можно заключить, что оценки DRF-GPS (рис. 1) слишком оптимистичны. Однако разброс полученных значений функции DML-DRF от -11.31 до 18.29 со средним -0.088 (означающее разброс в приросте численности безработных от 100%-ного снижения до экспоненциального роста) вызывает сомнения в состоятельности ее оценки этим методом.

5. Заключение

Существует проблема интерпретации оценки непрерывного воздействия в регрессионной модели вследствие зависимости уровня воздействия и результата от характеристик объекта. Статья представляет некоторые подходы к оценке непрерывного воздействия. Методы DRF и DML-DRF основываются на условной независимости воздействия и результата. Метод DD-DRF позволяет ослабить это условие до неслучайного отбора в группы сравнения, если ненаблюдаемые факторы, определяющие этот отбор, остаются постоянными во времени. Независимо от выбора метода оценки, исследователю важно теоретически обосновать свойства исследуемой зависимости и убедиться в устойчивости полученных результатов. Каждый из методов имеет свои сильные и слабые стороны. Выбор методов зависит от данных, с которыми имеет дело исследователь. Перечисленные методы оценки непрерывного воздействия разработаны относительно недавно, потому исследование свойств новых методов — большое поле для поиска.

Благодарность. Данная статья является результатом исследовательского проекта, реализованного в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ). Автор благодарит анонимного рецензента за ценные замечания, которые помогли сделать изложение более ясным.

Список литературы

- Анатольев С. А. (2009). Непараметрическая регрессия. *Квантиль*, 7, 37–52. <http://quantile.ru/07/07-SA.pdf>.
- Гимпельсон В. Е. (2022). Зарплата и потоки на российском рынке труда в условиях коронакризиса. *Вопросы экономики*, 2, 69–94. DOI: 10.32609/0042-8736-2022-2-69-94.
- Котырло Е. С. (2024). Простой и сложный метод разности разностей. *Прикладная эконометрика*, 73, 102–118. DOI: 10.22394/1993-7601-2024-73-119-142.
- Abadie A., Drukker D., Herr J. L., Imbens G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, 4 (3), 290–311. DOI: 10.1177/1536867X0400400307.
- Belloni A., Chernozhukov V., Wei Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34, 606–619. DOI: 10.1080/07350015.2016.1166116.
- Bia M., Flores C., Flores-Lagunes A., Mattei A. (2014). A Stata package for the application of semiparametric estimators of dose-response functions. *The Stata Journal*, 14 (3), 580–604. DOI: 10.1177/1536867X1401400307.
- Bia M., Mattei A. (2008). A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *The Stata Journal*, 8 (3), 354–373. DOI: 10.1177/1536867X1401400307.

- Borusyak K., Jaravel X., Spiess J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91 (6), 3253–3285. DOI: 10.1093/restud/rdae007.
- Callaway B., Goodman-Bacon A., Sant'Anna P. H. (2024a). *Data and code for: Event-studies with a continuous treatment*. Nashville, TN: American Economic Association. DOI: 10.3886/E201785V1.
- Callaway B., Goodman-Bacon A., Sant'Anna P. H. (2024b). Difference-in-differences with a continuous treatment. *Working Paper 32117*. National Bureau of Economic Research.
- Callaway B., Goodman-Bacon A., Sant'Anna P. H. (2024c). Event studies with a continuous treatment. *AEA Papers and Proceedings*, 114, 601–605. DOI: 10.1257/pandp.20241047.
- Callaway B., Sant'Anna P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225 (2), 200–230.
- Cameron A. C., Trivedi P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Cerulli G. (2015). Econometric evaluation of socio-economic programs. *Advanced Studies in Theoretical and Applied Econometrics Series*, 49. Springer.
- Chernozhukov V., Demirer M., Duflo E., Fernandez-Val I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. *Working Paper 24678*. National Bureau of Economic Research.
- Chernozhukov V., Hansen C., Kallus N., Spindler M., Syrgkanis V. (2024). Applied causal inference powered by ML and AI. <https://causalml-book.org/>.
- De Chaisemartin C., D'Haultfoeuille X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110 (9), 2964–2996. DOI: 10.1257/aer.20181169.
- DML for CATE (2025). <https://colab.research.google.com/github/CausalAIBook/MetricsMLNotebooks/blob/main/T/dml-for-conditional-average-treatment-effect.irnb>.
- Flores C. A., Flores-Lagunes A., Gonzalez A., Neumann T. C. (2012). Estimating the effects of length of exposure to instruction in a training program: The case of job corps. *Review of Economics and Statistics*, 94 (1), 153–171. DOI: 10.1162/REST_a_00177.
- Galagate D. (2016). Causal inference with a continuous treatment and outcome: Alternative estimators for parametric dose-response functions with applications. *Ph D. Dissertation*. University of Maryland, College Park. <https://www.proquest.com/openview/888e50198b2563222f8e2030e5a8c9e7/1?pq-origsite=gscholar&cbl=18750>.
- Galagate D., Schafer J. L. (2022). Estimating average dose response functions using the R package *causaldrf*. https://cran.r-project.org/web/packages/causaldrf/vignettes/Using_causaldrf.pdf.
- Goodman-Bacon A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225, 254–277. DOI: 10.1016/j.jeconom.2021.03.014.
- Guardabascio B., Ventura M. (2014). Estimating the dose–response function through a generalized linear model approach. *The Stata Journal*, 14 (1), 141–158. DOI: 10.1177/1536867X1401400110.
- Hirano K., Imbens G. W. (2004) The propensity score with continuous treatments. In: *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, 73–84. New York: Wiley.
- Hünermund P., Louw B., Caspi I. (2023). Double machine learning and automated confounder selection: A cautionary tale. *Journal of Causal Inference*, 11 (1), 20220078.

- Ichimura H., Newey W. K. (2022). The influence function of semiparametric estimators. *Quantitative Economics*, 13 (1), 29–61. DOI: 10.3982/QE826.
- Imbens G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87 (3), 706–710.
- Kennedy E. H., Ma Z., McHugh M. D., Small D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79 (4), 1229–1245. DOI: 10.1111/rssb.12212.
- King G., Nielsen R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27 (4), 435–454.
- Kotyrlo E. (2024). The impact of anti-COVID-19 restrictions and transitory unemployment insurance policies on unemployment in Russia. *Journal of Economic Studies*, 51 (1), 111–125. DOI: 10.1108/JES-12-2022-0615.
- Mattei A., Bia M. (2009). DoseResponse: Stata module to estimate dose-response function through adjustment for the generalized propensity score. *Statistical Software Components S457096*. Boston College Department of Economics.
- Rosenbaum P. R., Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Roth J., Sant’Anna P. H., Bilinski A., Poe J. (2023). What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235 (2), 2218–2244. DOI: 10.1016/j.jeconom.2023.03.008.
- Semenova V., Chernozhukov V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24 (2), 264–289. DOI: 10.1093/ectj/utaa027.
- Shalev-Shwartz S., Ben-David S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Sostero M., Milasi S., Hurley J., Fernandez-Macías E., Bisello M. (2020). Teleworkability and the COVID-19 crisis: A new digital divide? *JRC Working Papers Series on Labour, Education and Technology*, 2020/05.
- Wasserman L. (2006) *All of nonparametric statistics*. New York: Springer.
- Wooldridge J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26 (3), C31–C66.
- Zhao S., van Dyk D. A., Imai K. (2020). Propensity score-based methods for causal inference in observational studies with non-binary treatments. *Statistical Methods in Medical Research*, 29 (3), 709–727. DOI: 10.1177/0962280219888745.

Поступила в редакцию 07.06.2025;
принята в печать 21.08.2025

Kotyrlo E. S. Evaluation of continuous treatment. *Applied Econometrics*, 2025, v. 80, pp. 93–116. DOI: 10.22394/1993-7601-2025-80-93-116.

Elena Kotyrlo

HSE University, Moscow, Russian Federation;
ekotyrlo@hse.ru

Evaluation of continuous treatment

Continuous treatment effect evaluation is complicated by confounders. The dose is commonly related to observed characteristics of the treated unit. This paper provides an overview of approaches for evaluating continuous treatment effects. They are 1) the dose-response function; 2) application to panel data based on difference-in-differences; and 3) conditional treatment effect evaluation with debiased machine learning. Empirical examples of the teleworkability effect on unemployment indicators during the COVID-19 period illustrate the approaches.

Keywords: the generalized propensity score; dose-response function; difference-in-differences; policy evaluation; debiased machine learning; conditional average treatment effect.

JEL classification: C18; C21; C23; C87.

References

- Anatolyev S.A. (2009). Nonparametric regression. *Quantile*, 7, 37–52 (in Russian). <http://quantile.ru/07/07-SA.pdf>.
- Gimpelson V.E. Wages and labor market flows in times of the corona crisis. *Voprosy Ekonomiki*, 2, 69–94 (in Russian). DOI: 10.32609/0042-8736-2022-2-69-94.
- Kotyrlo E. S. (2024). Simple and complex difference-in-differences approach. *Applied Econometrics*, 73, 119–142 (in Russian). DOI: 10.22394/1993-7601-2024-73-119-142.
- Abadie A., Drukker D., Herr J. L., Imbens G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, 4 (3), 290–311. DOI: 10.1177/1536867X0400400307.
- Belloni A., Chernozhukov V., Wei Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34, 606–619. DOI: 10.1080/07350015.2016.1166116.
- Bia M., Flores C., Flores-Lagunes A., Mattei A. (2014). A Stata package for the application of semiparametric estimators of dose-response functions. *The Stata Journal*, 14 (3), 580–604. DOI: 10.1177/1536867X1401400307.
- Bia M., Mattei A. (2008). A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *The Stata Journal*, 8 (3), 354–373. DOI: 10.1177/1536867X1401400307.
- Borusyak K., Jaravel X., Spiess J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91 (6), 3253–3285. DOI: 10.1093/restud/rdae007.
- Callaway B., Goodman-Bacon A., Sant’Anna P. H. (2024a). *Data and code for: Event-studies with a continuous treatment*. Nashville, TN: American Economic Association. DOI: 10.3886/E201785V1.

- Callaway B., Goodman-Bacon A., Sant'Anna P. H. (2024b). Difference-in-differences with a continuous treatment. *Working Paper 32117*. National Bureau of Economic Research.
- Callaway B., Goodman-Bacon A., Sant'Anna P. H. (2024c). Event studies with a continuous treatment. *AEA Papers and Proceedings*, 114, 601–605. DOI: 10.1257/pandp.20241047.
- Callaway B., Sant'Anna P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225 (2), 200–230.
- Cameron A. C., Trivedi P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Cerulli G. (2015). Econometric evaluation of socio-economic programs. *Advanced Studies in Theoretical and Applied Econometrics Series*, 49. Springer.
- Chernozhukov V., Demirer M., Duflo E., Fernandez-Val I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. *Working Paper 24678*. National Bureau of Economic Research.
- Chernozhukov V., Hansen C., Kallus N., Spindler M., Syrgkanis V. (2024). Applied causal inference powered by ML and AI. <https://causalml-book.org/>.
- De Chaisemartin C., D'Haultfoeuille X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110 (9), 2964–2996. DOI: 10.1257/aer.20181169.
- DML for CATE (2025). <https://colab.research.google.com/github/CausalAIBook/MetricsMLNotebooks/blob/main/T/dml-for-conditional-average-treatment-effect.irnb>.
- Flores C. A., Flores-Lagunes A., Gonzalez A., Neumann T. C. (2012). Estimating the effects of length of exposure to instruction in a training program: The case of job corps. *Review of Economics and Statistics*, 94 (1), 153–171. DOI: 10.1162/REST_a_00177.
- Galagate D. (2016). Causal inference with a continuous treatment and outcome: Alternative estimators for parametric dose-response functions with applications. *Ph D. Dissertation*. University of Maryland, College Park. <https://www.proquest.com/openview/888e50198b2563222f8e2030e5a8c9e7/1?pq-origsite=gscholar&cbl=18750>.
- Galagate D., Schafer J. L. (2022). Estimating average dose response functions using the R package *causaldrf*. https://cran.r-project.org/web/packages/causaldrf/vignettes/Using_causaldrf.pdf.
- Goodman-Bacon A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225, 254–277. DOI: 10.1016/j.jeconom.2021.03.014.
- Guardabascio B., Ventura M. (2014). Estimating the dose–response function through a generalized linear model approach. *The Stata Journal*, 14 (1), 141–158. DOI: 10.1177/1536867X1401400110.
- Hirano K., Imbens G. W. (2004) The propensity score with continuous treatments. In: *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, 73–84. New York: Wiley.
- Hünermund P., Louw B., Caspi I. (2023). Double machine learning and automated confounder selection: A cautionary tale. *Journal of Causal Inference*, 11 (1), 20220078.
- Ichimura H., Newey W. K. (2022). The influence function of semiparametric estimators. *Quantitative Economics*, 13 (1), 29–61. DOI: 10.3982/QE826.
- Imbens G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87 (3), 706–710.

- Kennedy E. H., Ma Z., McHugh M. D., Small D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79 (4), 1229–1245. DOI: 10.1111/rssb.12212.
- King G., Nielsen R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27 (4), 435–454.
- Kotyrlo E. (2024). The impact of anti-COVID-19 restrictions and transitory unemployment insurance policies on unemployment in Russia. *Journal of Economic Studies*, 51 (1), 111–125. DOI: 10.1108/JES-12-2022-0615.
- Mattei A., Bia M. (2009). DoseResponse: Stata module to estimate dose-response function through adjustment for the generalized propensity score. *Statistical Software Components S457096*. Boston College Department of Economics.
- Rosenbaum P. R., Rubin D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Roth J., Sant’Anna P. H., Bilinski A., Poe J. (2023). What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235 (2), 2218–2244. DOI: 10.1016/j.jeconom.2023.03.008.
- Semenova V., Chernozhukov V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24 (2), 264–289. DOI: 10.1093/ectj/utaa027.
- Shalev-Shwartz S., Ben-David S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Sostero M., Milasi S., Hurley J., Fernandez-Macías E., Bisello M. (2020). Teleworkability and the COVID-19 crisis: A new digital divide? *JRC Working Papers Series on Labour, Education and Technology*, 2020/05.
- Wasserman L. (2006) *All of nonparametric statistics*. New York: Springer.
- Wooldridge J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26 (3), C31–C66.
- Zhao S., van Dyk D. A., Imai K. (2020). Propensity score-based methods for causal inference in observational studies with non-binary treatments. *Statistical Methods in Medical Research*, 29 (3), 709–727. DOI: 10.1177/0962280219888745.

Received 07.06.2025; accepted 21.08.2025.