Приложение № 28

Описание используемого подхода к количественной оценке влияния реализации мер государственной поддержки по привлечению инвестиций в приоритетные направления проектов технологического суверенитета и структурной адаптации экономики Российской Федерации на экономическую и инвестиционную деятельность предприятий

Количественная оценка влияния реализации мер государственной поддержки по привлечению инвестиций в приоритетные направления проектов технологического суверенитета и структурной адаптации экономики Российской Федерации на экономическую и инвестиционную деятельность предприятий проводилась с применением метода эконометрического анализа, а именно разность разностей. Искомые причинные параметры были определены с помощью модели Рубина как функционал распределения «полных» данных, то есть данных, содержащих все возможные исходы (англ. possible outcomes, counterfactuals). При таком подходе задача оценки причинных параметров подобна задаче об устранении пропусков в данных. Распределение, которое порождает данные, можно представить следующим образом: $Y = (1 - D) \times Y^{d=0} + D \times Y^{d=1}$, где Y — случайная величина для показателя, влияние на который требуется оценить (для краткости будем называть такой показатель выбранным; в роли Y выступали среднесписочная численность работников, выручка и изменение внеоборотных материальных активов); D — индикаторная функция воздействия, влияние которого требуется оценить (далее также участия в отобранных программах, заключения договора займа); $Y^{d=0}$ — случайная величина для выбранного показателя в том случае, когда воздействия не было (то есть если бы значение d случайной величины D было 0); $Y^{d=1}$ — случайная величина для выбранного показателя в том случае, когда воздействия было (то есть если бы значение d случайной величины D было 1).

Фундаментальная проблема причинного вывода в том, что из множества возможных исходов наблюдается самое большее один. Тем не менее при определенных условия возможна идентификация причинного параметра. Под идентификацией понимается подбор такого функционала распределения наблюденных («неполных») данных, который при данных условиях идентификации был бы равен искомому причинному параметру – функционалу распределения «полных» данных.

Основной искомый параметр – это средний причинный эффект среди тех, кого подвергли воздействию (англ. average treatment effect on the treated, далее — ATT): $\mathbb{E}(Y^{d=1} - Y^{d=0} | D = 1)$. Такой параметр удобен тем, что позволяет оценить и суммарный эффект среди предприятий – участников оцениваемых программ. Для этого достаточно умножить ATT на фактическое число предприятий, средний эффект среди которых оценивался, n_1. Очевидно, что и статистические выводы о таком производном параметре будут возможны, ведь квадратичная ошибка оценки такого суммарного параметра будет больше квадратичной ошибки соответствующей оценки ATT в n_1 раз (при этом n_1 можно считать данным параметром, а не случайной величиной). Аналогично можно определить процентное приращение выбранного показателя, вызванное участием в оцениваемых программах, или долю вызванного участием в оцениваемых программах прибавлении в суммарном (по предприятиям – участникам оцениваемых программ) значении выбранного показателя. Пусть $\tilde{Y}$ — суммарное значение выбранного показателя среди участников оцениваемых программ, $\bar{Y}$ — среднее значение. Тогда процентное приращение — $\frac{ATT \times n_1}{\tilde{Y} - (ATT \times n_1)}$, или $\frac{ATT}{\bar{Y} - ATT}$; доля фактического значения, которая объясняется участием в программах — $\frac{ATT \times n_1}{\tilde{Y}}$, или $\frac{ATT}{\bar{Y}}$. Статистические выводы о таких параметрах возможны благодаря дельта-методу: если известна квадратичная ошибка $\mathrm{se}(\hat{\theta})$ асимптотически нормального параметра θ, то квадратичная ошибка производного параметра $f(\theta)$ равна $|f'^{(\theta)}|\mathrm{se}(\hat{\theta})$.

Для идентификации ATT использовались предположения метода разности разностей:

1) $\mathbb{E}(Y^{d=0}|D = 1, X) - \mathbb{E}(Y^{d=0}|D = 0, X) = \mathbb{E}(C|D = 1, X) - \mathbb{E}(C|D = 0, X)$, где C — значение выбранного показателя в год, предшествующий заключению договора займа; X — контрольные случайные величины из года, предшествующего заключению договора займа (англ. pretreatment variables);

2) $\mathbb{P}(D = 1) > 0$ и с вероятностью 1 $\mathbb{P}(D = 1|X) < 1$.

Второе условие дает руководство к тому, кого следует отбирать в контрольную группу: там должны быть лица с ненулевой условной вероятностью участия. Именно поэтому в контрольную группу отбирались российские предприятия (то есть только коммерческие, согласно коду КОПФ/ОКОПФ юридические лица; только такие лица заключали договоры займа) с классом вида деятельности (основной согласно ЕГРЮЛ), который бывал хотя бы у одного лица, заключившего договор займа (в определенный год).

Первое условие (условная параллельность трендов) можно представить иначе, если данные панельные (также англ. repeated outcomes). Речь о том, что известны значения выбранного показателя для того же предприятия и до, и после воздействия. Альтернатива – повторные поперечные срезы, при которых наблюдения до воздействия и после воздействия – это случайные выборки из подходящей совокупности предприятий, так что одно и то же предприятие может наблюдаться лишь в один период (до или после воздействия). Если данные панельные, условная параллельность трендов соответствует более стандартной для литературы о причинных выводах (англ. causal inference) условной заменимости (англ. conditional exchangeability; также в англ. selection upon covariates, conditional ignorability, unconfoundedness). Речь о том, что при известных значениях контрольных величин ожидание возможного исхода не зависит от того, какое значение приняла индикаторная переменная D: $\forall d\ \mathbb{E}(Y^d|L, D=1) = \mathbb{E}(Y^d|L, D=0)$, где L — подходящие контрольные величины. Идея метода разности разностей в том, чтобы ослабить требования к подходящему набору контрольных величин. Однако отмечается, что при панельных данных условная параллельности трендов эквивалентна условной заменимости *разности* выбранного показателя после воздействия и до воздействия. Пусть ΔY — такая разность, то есть $\Delta Y^d = Y^d - C$. Тогда условную параллельность трендов по $Y^{d=0}$ можно представить как условную заменимость по этой разности при $d = 0$: $\mathbb{E}(\Delta Y^{d=0}|X, D=1) = \mathbb{E}(\Delta Y^{d=0}|X, D=0)$. Заметим, что разность разностей требует условной заменимости ΔY^d лишь при $d = 0$. Если такое условие выполняется и для $d = 1$, то можно идентифицировать не только ATT $((Y^{d=1} - Y^{d=0}|D=1))$, но и средний эффект среди тех предприятий, которые не заключили договор займа, если бы эти предприятия его все же заключили: $(Y^{d=1} - Y^{d=0}|D=0)$. На сходство условной заменимости и условной параллельности трендов явно указывали по меньшей мере Gavrilova, Langørgen и Zoutman (2023)[1], но сходство очевидно и при сравнении эффективной функции влияния для оценок ATT при двух альтернативных условиях идентификации.

При выполнении условий идентификации метода разности разностей и при наличии панельных данных ATT идентифицирует формула: $\mathbb{E}\big(\mu_{1,\Delta}(X) - \mu_{0,\Delta}(X)|D=1\big)$, где $\mu_{d,\Delta}$ — условное ожидание (регрессия) $\Delta Y = Y - C$ при известных контрольных величинах X и $D = d$, то есть $\mu_{\mathrm{d},\Delta} = \mathbb{E}(\Delta Y|D=d, X)$ [2]. Этот результат в точности совпал бы с формулой, идентифицирующей ATT при условной заменимости по ΔY^d [3].

Сходство метода разности разностей с условной заменимостью полезно потому, что позволяет обосновать подбор контрольных величин X. Pearl свел проверку условной

[1] Gavrilova E., Langørgen A., Zoutman F. Dynamic Causal Forests, with an Application to Payroll Tax Incidence in Norway // SSRN Journal. 2023. DOI 10.2139/ssrn.4500857.

[2] Heckman J. J., Ichimura H., Todd P. E. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme // The Review of Economic Studies. 1997. T. 64, № 4. C. 605–654 ; Sant'Anna P. H. C., Zhao J. Doubly robust difference-in-differences estimators // Journal of Econometrics. 2020. T. 219, № 1. C. 101–122 ; Roth J. и др. What's trending in difference-in-differences? A synthesis of the recent econometrics literature // Journal of Econometrics. 2023. T. 235, № 2. C. 2218–2244.

[3] См., например, приложение A.13 в Laan M. J. van der, Rose S. Targeted learning: causal inference for observational and experimental data. New York: Springer, 2011. 626 c.

заменимости к d-разделению в причинном ациклическом орграфе, кодирующем качественные представления об изучаемой системе. Вершина в таком графе – случайная величина, а стрелка от одной вершины к другой указывает на причинность. Один из результатов состоит в том, что достаточно включить в множество контрольных величин показатели, представляющие вероятностные причины воздействия, эффект которого оценивается. За этим стоит в точности такая логика, которая оправдывает причинные выводы по результатам рандомизированного испытания: когда исследователь распределяет подопытных в опытную и контрольную группу с помощью броска честной монеты, он отсекает естественные причины, которые бы определяли, кого подвергнут оцениваемому воздействию, а кого нет[4].

Такими вероятностными причинами для участия в программах / заключения договора займа являются:

- вид деятельности (в паспортах оцениваемых программ особо выделяются судоремонтные предприятия, станкостроение, производство роботов, оборонная промышленность);
- наличие категории МСП (банковская ставка ниже (программа привлекательнее), если предприятие получает гарантию корпорации МСП);
- показатели финансового состояния (например, банковская ставка ниже (программа привлекательнее), если предприятие получает гарантию ВЭБ.РФ. Однако ВЭБ.РФ, согласно постановлению Правительства № 92 от 05.02.2020, при принятии решения учитывает коэффициент текущей ликвидности предприятия, рентабельность продаж, норму чистой прибыли за прошедший период);
- получение льгот:
    - благодаря специальному налоговому режиму;
    - участию в преференциальных режимах, таких как ОЭЗ или центр «Сколково»;
- возраст предприятия (нужна определенная зрелость для осуществления крупных проектов);
- регион предприятия (влияет, к примеру, на осведомленность предприятий о мерах поддержки, отражает различия в величине региональных налоговых и иных льгот);
- размер предприятия (участие в программах предполагало определенные вложения со стороны участников).

[4] Pearl J., Glymour M., Jewell N. P. Causal Inference in Statistics: A Primer. John Wiley & Sons, 2016. 160 с. ; Hernán M. A., Robins J. M. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, 2020. 310 с.

Для указанных вероятностных причин участия в программах были подобраны соответствующие контрольные величины (в год, предшествующий рассматриваемому году заключения договора займа):

- класс основного вида деятельности согласно ЕГРЮЛ;
- регион согласно ОГРН;
- возраст предприятия с учетом истории его реорганизаций;
- признак наличия данной категории МСП согласно единому реестру МСП (с учетом того, что в том же году предприятие могло переходить из одной категории в другую);
- текущая ликвидность предприятия (в типовой отчетности отношение показателя строки БФО 1200 к сумме показателей в строках 1510, 1520, 1540 и 1550; в упрощенной БФО отношение суммы из строк 1210, 1230 и 1250 к сумме из строк 1510, 1520 и 1550);
- рентабельность продаж (в типовой отчетности отношение строк БФО 2200 и 2110, в упрощенной – отношение суммы строк 2110 и 2120 к значению в строке 2110);
- норма чистой прибыли за прошедший период (отношение суммы в строке 2400 БФО к сумме в строке 2110);
- признак применения предприятием того или иного налогового спецрежима (УСН, ЕСХН, ЕНВД, СРП, АУСН);
- признак участия в промышленно-производственной, или технико-внедренческой, или портовой, или туристско-рекреационной (с разделением на зоны в кластере и не в кластере) ОЭЗ, дальневосточных ТОР, ТОР в монопрофильном образовании, ТОР в ЗАТО Росатома, калининградской или Магаданской ОЭЗ, центре «Сколково»;
- выбранный показатель в год, предшествующий рассматриваемому году заключение договора займа.

Представленная ранее формула, идентифицирующая при допущениях метода разности разностей АТТ, требует оценки мешающих параметров: $\mu_{d,\Delta}$. Альтернативный подход – идентификация ATT через показатель склонности $\pi(X)$ — условное ожидание D при известных контрольных величинах X[5]. Однако ни $\mu_{d,\Delta}$, ни π исследователям не известны. Традиционный параметрических подход, сводящийся к оценке параметров

[5] Abadie A. Semiparametric Difference-in-Differences Estimators // The Review of Economic Studies. 2005. Т. 72, № 1. С. 1–19.

линейной регрессии для $\mu_{d,\Delta}$ и логистической регрессии для $\pi(X)$, не имеет научного обоснования и наверняка ведет к смещенным оценкам. Вместо того, чтобы оценивать якобы заранее известные с точностью до конечного числа параметры функции, разумнее решать задачу восстановления зависимости по эмпирическим данным (по выражению В. Н. Вапника). Такую задачу решают методы статистического обучения, в последнее время часто именуемые методами машинного обучения. Тем не менее простая подстановка предсказаний результата обучения модели машинного обучения ведет к смещенным оценкам[6]. Фундаментально причина в том, что методы статистического обучения подбирают оценку так, чтобы уменьшить риск для мешающего параметра ($\mu_{d,\Delta}(X)$ или $\pi(X)$), а не для интересующего параметра (ATT), который от мешающего зависит.

Ключ к применению гибких методов статистического обучения для оценки причинных параметров – это функции влияния. В общем случае нельзя требовать от оценки, чтобы она была несмещенная, но для широкого класса параметров, среди которых и интересующий нас ATT, существуют асимптотически линейные оценки, то есть такие оценки параметра $T(F)$, что $T(\hat{F}_n) - T(F) = \frac{1}{n}\sum_{i=1}^{n} \hat{\varphi}_{\hat{F}}(x_i) + o_p\left(\frac{1}{\sqrt{n}}\right)$, где φ — функция влияния, функция с нулевым ожиданием и конечной дисперсией; F — истинная функция распределения; $\hat{F}_n$ — ее эмпирический аналог; n — число наблюдений, участвовавших в оценивании. Так как нормированная (умноженная на $\sqrt{n}$) разность оценки и истинного значения параметра при все большем n ведет себя как сумма одинаково и независимо распределенных величин с нулевым ожиданием, асимптотически линейные оценки распределены асимптотически нормально, возможно рассчитывать асимптотически корректную квадратичную ошибку оценки, строить доверительные интервалы и проверять статистические гипотезы с помощью критерия Вальда. Дисперсия оценки $T(\hat{F}_n)$ оценивается как $\frac{1}{n}\sum_{i=1}^{n} \hat{\varphi}_i^2$, где $\hat{\varphi}_i$ — подстановочная оценка функции влияния для наблюдения i.

Оценка, у которой есть функция влияния, является асимптотически линейной. И наоборот. Выведение функций влияния и построение с их помощью асимптотически линейных оценок бывает нетривиальным, однако для ATT при допущениях как об условной заменимости, так и об условной параллельности трендов, такая функция влияния известна. Более того, известна эффективная функция влияния, то есть функция влияния с наименьшей дисперсией. В таком случае и построенная с помощью эффективной функции влияния оценка оказывается эффективной, то есть имеющей наименьшую возможную дисперсию.

[6] Наглядный пример есть в Chernozhukov V. и др. Double/debiased machine learning for treatment and structural parameters // The Econometrics Journal. 2018. Т. 21, № 1. С. C1–C68.

Эффективную функцию влияния для ATT при допущениях метода разности разностей представили Sant'Anna и Zhao (2020). Если упростить приведенное в оригинальной статье (формула 2.11) выражение, получим: $\varphi = \frac{1}{\mathbb{E}(D)}\left(\frac{D-\pi}{1-\pi}\left(\Delta Y - \mu_{0,\Delta}\right) - D\tau\right)$, где для краткости записи опущены нижние индексы – указатели номера наблюдения, а также сокращены выражения вроде $\pi(X)$, τ — значение ATT [7].

Нормированную разность наивной подстановочной оценки и истинного значения параметра $\sqrt{n}\left(T(\hat{F}_n) - T(F)\right)$ можно разложить (разложение фон Мизеса, аналог разложения Тейлора для распределений) на сумму из трех членов[8].

Первый член в таком разложении отражает смещение из-за подстановки вместо мешающих параметров идентифицирующей параметр формулы предсказаний статистической модели. Есть несколько способов его устранения. Самый простой – добавить к наивной оценке выборочное среднее значений оценки функции влияния. Это так называемая одношаговая оценка. Более современный метод – выразить параметр как решение основанного на эффективной функции влияния решающего уравнения. Именно он был использован. У функции влияния нулевое ожидание, а потому оценивающее уравнение задается равенством $\mathbb{P}_n\left(\frac{1}{\mathbb{E}(D)}\left(\frac{D-\pi}{1-\pi}\left(\Delta Y - \mu_{0,\Delta}\right) - D\tau\right)\right) = 0$, где $\mathbb{P}_n$ — указатель выборочного среднего. Из решения уравнения следует формула для искомого τ: $\hat{\tau} = \frac{\mathbb{P}_n\left(\frac{D-\hat{\pi}}{\hat{\mathbb{E}}(D)(1-\hat{\pi})}\left(\Delta Y - \hat{\mu}_{0,\Delta}\right)\right)}{\mathbb{P}_n\left(\frac{D}{\hat{\mathbb{E}}(D)}\right)}$, где надстрочные символы указывают на оценку с помощью моделей статистического обучения.

При такой корректировке в разложении $\sqrt{n}\left(T(\hat{F}_n) - T(F)\right)$ остается член, от которого требуется быстрая сходимость, то есть он должен быть $o_p(1)$. Для этого отдельно рассматриваются два оставшихся члена в разложении. Сходимость первого гарантируется, если при оценивании применяется разделение данных (в англ. литературе используется cross-fitting). Идея в том, чтобы предсказания для подстановки в формулу $\hat{\tau}$ для каждого наблюдения давали модели, которые при обучении это наблюдение не использовали. Последний член требует достаточно быстрой сходимости оценок мешающих параметров, что достигается благодаря применению гибких методов машинного обучения. Таким гибким методом послужил обобщенный случайный лес. Подробное его описание дано в

[7] Sant'Anna P. H. C., Zhao J. Doubly robust difference-in-differences estimators // Journal of Econometrics. 2020. Т. 219, № 1. С. 101–122 ; Sant'Anna P. H. C., Zhao J. B. Doubly Robust Difference-in-Differences Estimators: Supplementary Appendix. 2020.

[8] Hines O. и др. Demystifying Statistical Learning Based on Efficient Influence Functions // The American Statistician. Taylor & Francis, 2022. Т. 76, № 3. С. 292–304.

работе Athey, Tibshirani и Wager (2019) [9]. Главное (по меньшей мере для стоявшей задачи оценки мешающих параметров) отличие от обычного случайного леса в том, что создатели модели называют «честностью»: структура деревьев определяется на одной части данных, а предсказания делаются на другой. Это свойство было особенно полезно, так как фактически решало проблему с разделением данных. Для оценивания мешающих параметров использовалась библиотека grf [10] для языковой среды R. При этом калибровался полный набор параметров для каждой модели[11].

С помощью обобщенного случайного леса получались оценки мешающих параметров $\hat{\pi}$ и $\hat{\mu}_{0,\Delta}$. И $\mu_{0,\Delta}$, и π — это регрессии, условные математические ожидания типа $\mathbb{E}(Y_i|X_i = x)$, а потому алгоритм обобщенного случайного леса имел такой вид:

1. Инициировать нулевой вектор $\vec{\alpha}$ длиной, равной числу наблюдений (строк в таблице данных).
2. B (2000) раз (число деревьев):
   а) отобрать случайно часть исходных наблюдений;
   б) разбить отобранные наблюдения на 2 части;
   в) с помощью первой части построить дерево: последовательно разбить пространство признаков на прямоугольные области по значению той или иной контрольной величины так, чтобы при каждом делении максимальным было различие значений оценки искомого параметра в двух образованных разбиением областях (мера величины различий — $\frac{n_1 n_2}{(n_1+n_2)^2}\left(\hat{Y}_1 - \hat{Y}_2\right)^2$, где n_1 и n_2 — число наблюдений в узлах, на которые разбивается родительский узел, $\hat{Y}_1, \hat{Y}_2$ — предсказания регрессии в первом и втором узлах, на которые разбивается родительский узел); при каждом разбиении участвуют только несколько случайно отобранных характеристик предприятий;
   г) с помощью второй части и построенного дерева для данной точки x найти список наблюдений-«соседей» $\mathrm{L_b}(\mathrm{x})$ второй части наблюдений (которые не использовались для построения дерева), которые попадают в тот же конечный узел («листок»), что и точка x;
3. Рассчитать искомый параметр $\mathbb{E}(Y_i|X_i = x)$ как $\sum_{i=1}^{n} \frac{\alpha_i}{B} Y_i$.

[9] Athey S., Tibshirani J., Wager S. Generalized random forests // The Annals of Statistics. Institute of Mathematical Statistics, 2019. Т. 47, № 2. С. 1148–1178.
[10] https://grf-labs.github.io/grf/
[11] https://grf-labs.github.io/grf/reference/regression_forest.html, функция применялась с tune.parameters = "all".

От обычного случайного леса представленный выше алгоритм фактически отличается дополнительным разделением данных, с тем чтобы усреднение не происходило по наблюдениям, которые использовались для построения дерева (следствие пункта 2б).

При этом с помощью перекрестной проверки (англ. cross-validation) калибровались следующие настроечные параметры обобщенного случайного леса:

- параметры, которые управляют построением деревьев:
  - доля наблюдений, которая используется при построении каждого дерева; между 0 и 1;
  - верхняя граница числа наблюдений в конечном узле дерева; при калибровке значение выбрилось по формуле $2^{U(0,\,1)\times\left(\frac{\ln n}{\ln 2}-4\right)}$, где $U(0,\,1)$ — случайная величина с равномерным распределением на интервале от 0 до 1, а n — число наблюдений;
  - число независимых переменных, которые случайно отбираются перед каждым разбиением внутреннего узла; не более числа независимых величин в матрице данных;
- параметры, которые управляют «честностью» алгоритма:
  - доля наблюдений, которые отбираются для построения дерева (уже после того, как была отобрана часть наблюдений sample.fraction); от 0,5 до 0,8;
  - следуют ли конечные узлы («листья») с нулевым числом не участвовавших в построении дерева наблюдений из отложенных на шаге 2б алгоритма выше данных; возможные значения – да или нет;
- параметры, которые управляют делением узлов при построении дерева:
  - верхняя граница доли наблюдений, в любом из двух узлов, на которые разбивается «родительский» узел при построении дерева; между 0 и 0,25;
  - штраф за деление «родительского» узла на узлы с очень разным числом наблюдений; по умолчанию нуль; от 0 (по умолчанию); штраф применяется к оценке качества разбиения «родительского» узла для выбора предпочтительного разбиения.

Использовалось 100 наборов настроечных параметров, выбранных случайно из областей их значений. Для оценивания строилось 2000 деревьев.

Параметр ATT оценивался отдельно для каждой пары года заключения договора займа g и года наблюдения/оценки эффекта t $(t \geq g)$. Такой подход дает более ясные оценки, чем попытки свести разнообразие величины эффекта (различия между группами предприятий, в разное время присоединившихся к оцениваемым программам, изменение эффекта с течением времени и т. п.) ко всего одному параметру[12].

Сосредоточение на парах (g, t) означает, что оценивался параметр $ATT(g,t) = \mathbb{E}\left(Y_t^{d_g=1} - Y_t^{d_g=0} | D_g = 1\right)$, где $Y_t^{d_g=1}$ — возможное значение выбранного показателя (ССЧР, выручки или «нетто-инвестиций») в год $t \geq g$, если (бы) предприятие присоединилось к оцениваемым программам в год g, $Y_t^{d_g=0}$ — возможное значение выбранного показателя (ССЧР, выручки или «нетто-инвестиций») в год $t \geq g$, если (бы) предприятие не присоединилось к оцениваемым программам в год g; D_g — индикаторная переменная, принимающая значение 1, если предприятие впервые присоединилось к оцениваемым программам в год g. Такая постановка задачи сводит оценивание к стандартной схеме метода разности разностей со всего двумя временными отрезками (до g и в год t) и всего двумя группами (те, кто присоединился к оцениваемым программам в год g, и те, кто вовсе никогда не присоединялся к оцениваемым программам.

Стандартные ошибки оценивались для оценки искомого параметра – причинного эффекта, не для оценок мешающих параметров – регрессий. В оценке стандартной ошибки регрессий, оцененных с помощью обобщенного леса, не было необходимости. Формула оценки искомого параметра: $\frac{\mathbb{P}_n\left(\frac{D-\hat{\pi}}{\hat{\mathbb{E}}(D)(1-\hat{\pi})}\left(\Delta\hat{Y}-\hat{\mu}_{0,\Delta}\right)\right)}{\mathbb{P}_n\left(\frac{D}{\hat{\mathbb{E}}(D)}\right)}$. В нее подставляются предсказания $\hat{\mu}_{0,\Delta}$ и $\hat{\pi}$, полученные благодаря двум моделям обобщенного леса. Стандартные ошибки $\hat{\mu}_{0,\Delta}$ и $\hat{\pi}$ не использовались, только сами предсказания $\hat{\mu}_{0,\Delta}(X_i)$ и $\hat{\pi}(X_i)$.

Расчет квадратичной ошибки искомого параметра описывается следующим образом: Дисперсия оценки $T(\hat{F}_n)$ оценивается как $\frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_i^2$, где $\hat{\varphi}_i$ — подстановочная оценка функции влияния для наблюдения i. Приведенная функция влияния искомого параметра — $\varphi = \frac{1}{\mathbb{E}(D)}\left(\frac{D-\pi}{1-\pi}\left(\Delta Y - \mu_{0,\Delta}\right) - D\tau\right)$. Соответственно, ее подстановочная оценка — $\frac{1}{\hat{\mathbb{E}}(D)}\left(\frac{D-\hat{\pi}}{1-\hat{\pi}}\left(\Delta Y - \hat{\mu}_{0,\Delta}\right) - D\hat{\tau}\right)$, где символ $\hat{}$ указывает на подстановку оценки, в частности предсказаний обобщенного леса $\hat{\pi}(X_i)$ и $\hat{\mu}_{0,\Delta}(X_i)$, где X_i — значения контрольных величин у i-го наблюдения.

Точечная оценка $\hat{\tau}_{g,t}$ для $ATT(g,t)$ рассчитывалась по формуле:

---

[12] Roth J. и др. What's trending in difference-in-differences? A synthesis of the recent econometrics literature // Journal of Econometrics. 2023. Т. 235, № 2. С. 2218–2244.

$$\hat{\tau}_{g,t} = \frac{\mathbb{P}_{n_{g,t}}\left(\frac{D_{i,g} - \hat{\pi}_{g,t}(X_{i,g-1})}{\widehat{\mathbb{E}}_i(D_g)(1 - \widehat{\pi_{g,t}})}\ \left(\Delta Y_{i,g-1,t} - \hat{\mu}_{0,\Delta,g-1,t}(X_{i,g-1})\right)\right)}{\mathbb{P}_{n_{g,t}}\left(\frac{D_{i,g}}{\widehat{\mathbb{E}}_i(D_g)}\right)},$$

где $\mathbb{P}_{n_{g,t}}$ — указатель выборочного среднего по наблюдениям, отобранным для оценивания эффекта для пары (*g, t*);

$D_{i,g}$ — значение двоичной переменной у предприятия i, равное 1, если предприятие i стало участвовать в оцениваемых программах в год g, или равно 0, если предприятие никогда не участвовало в оцениваемых программах;

$\widehat{\mathbb{E}}_i(D_g)$ — подстановочная оценка средней частоты начала участия в оцениваемых программах в год g, полученная с учетом требования разделения данных (в оценке не участвует предприятие i для которого подставляется предсказание): $\widehat{\mathbb{E}}_i(D_g) = \frac{1}{n_{g,t}-1}\sum_{k:k\neq i}^{n_{g,t}} D_{k,g}$, где $n_{g,t}$ — число наблюдений, отобранных для оценивания эффекта для пары (*g, t*);

$\hat{\pi}_{g,t}$ — полученная с помощью обобщенного леса оценка условной вероятности начала участия в оцениваемых программах в год g при оценивании эффекта в год t (выбор t влияет на выборку предприятий, а потому и на получаемую оценку $\hat{\pi}_{g,t}$;

$\Delta Y_{i,g-1,t}$ — фактическое наблюдаемое у предприятия i приращение выбранного для измерения эффекта показателя: $Y_{i,t} - Y_{i,g-1}$, где $Y_{i,t}$ — фактическое наблюдаемое у предприятия i значение выбранного для измерения эффекта показателя в год t, $Y_{i,g-1}$ — фактическое наблюдаемое у предприятия i значение выбранного для измерения эффекта показателя в год $g-1$;

$\hat{\mu}_{0,\Delta,g-1,t}$ — полученная с помощью обобщенного леса оценка условного ожидания (регрессии) приращения выбранного для измерения эффекта показателя (в год t по сравнению с годом $g-1$) при фиксированном $D = 0$, то есть это оценка $\mathbb{E}(\Delta Y_{g-1,t} | X_{g-1}, D_g = 0) = \mathbb{E}(Y_t - Y_{g-1} | X_{g-1}, D_g = 0)$;

$X_{i,g-1}$ — значение у предприятия i контрольных величин в году $g-1$, предшествующему рассматриваемому году первого заключения договора займа (g).

Квадратичная ошибка оценки $ATT(g,t)$ по формуле $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_i^2}$, где $\hat{\varphi}_i$ — подстановочная оценка функции влияния для наблюдения i, то есть оценка квадратичной ошибки определялась выражением:

$$\mathrm{se}\left(\hat{\tau}_{g,t}\right) = \sqrt{\frac{1}{n_{g,t}}\sum_{i=1}^{n_{g,t}}\left(\frac{1}{\widehat{\mathbb{E}}_i\left(D_g\right)}\left(\frac{D_{i,g}-\hat{\pi}_{g,t}\left(X_{i,g-1}\right)}{1-\hat{\pi}_{g,t}\left(X_{i,g-1}\right)}\ \left(\Delta Y_{i,g-1,t}-\hat{\mu}_{0,\Delta,g-1,t}\left(X_{i,g-1}\right)\right)-D_{i,g}\hat{\tau}_{g,t}\right)\right)^2},$$

где $n_{g,t}$ — число наблюдений, отобранных для оценивания эффекта для пары (*g, t*).

Границы доверительного интервала строились по формуле $\hat{\tau}_{g,t} \pm 1{,}96 \times \mathrm{se}\left(\hat{\tau}_{g,t}\right)$.