

# Эконометрический ликбез: эффекты воздействия

## Оценивание эффекта воздействия\*

Рубен Ениколопов<sup>†</sup>

*Российская экономическая школа, Москва, Россия*

В настоящем эссе содержится краткий обзор методов оценивания среднего эффекта воздействия программ, когда интересующая нас независимая переменная является бинарной.

### 1 Введение

В эмпирических исследованиях часто оценивают эффект бинарного воздействия одной переменной на другую, контролируя на определенный набор третьих переменных. Подобные ситуации возникают, например, когда требуется оценить эффект воздействия лекарства, контролируя на характеристики пациентов, или эффект экономической программы, контролируя на характеристики объектов, на которые направлена данная программа (фирм, людей, регионов и т.д.). В подобной ситуации все наблюдения распадаются на две группы – группу активного воздействия (*treatment group*) и контрольную группу (*control group*). В первую попадают объекты, подвергшиеся воздействию интересующей нас программы, лекарства и т.п., в то время как во вторую группу попадают объекты, не подвергшиеся такому воздействию. В случае, когда включение в качестве контрольных переменных конечного набора переменных позволяет полностью нивелировать не относящиеся к эффекту программы различия между экспериментальной и контрольной группой, сравнение результатов двух групп могут быть истолкованы с точки зрения наличия причинно-следственной связи между воздействием и интересующим нас исходом.

### 2 Потенциальные исходы

Данный подход разработан в работах Рубина (Rubin, 1974, 1977, 1978).

#### 2.1 Определения

Пусть мы наблюдаем  $N$  объектов, случайно выбранных из популяции, где каждый объект обозначен индексом  $i = 1, \dots, N$ . Мы предполагаем, что для каждого объекта  $i$  определена пара потенциальных исходов:  $Y_i(0)$  для исхода в случае, если объект попадает в контрольную группу, и  $Y_i(1)$  для исхода в случае, если объект попадает в экспериментальную группу (группу активного воздействия).

Стоит сразу отметить два важные ограничения в рассматриваемой ситуации. Во-первых, предполагается, что интенсивность воздействия на объекты в группе воздействия предполагается неизменной, что позволяет объединить все объекты в одну экспериментальную группу. Во-вторых, в данном подходе для каждого объекта потенциальные исходы фиксированы и

\* Данное эссе в значительной мере основывается на статье Imbens & Wooldridge (2008), а также на лекции, прочитанной Гвидо Имбенсом в курсе “What’s New in Econometrics”, видеозапись и текст которой можно найти по адресу [www.nber.org/minicourse3.html](http://www.nber.org/minicourse3.html). Цитировать как: Ениколопов, Рубен (2009) «Оценивание эффекта воздействия», Квантиль, №6, стр. 3–14. Citation: Enikolopov, Ruben (2009) “Estimation of treatment effects,” *Quantile*, No.6, pp. 3–14.

<sup>†</sup> Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1717. Электронная почта: [REnikolopov@nes.ru](mailto:REnikolopov@nes.ru)

не зависят от того, какие другие объекты подвергаются воздействию. Данное предположение известно как предположение о стабильности эффекта воздействия на объект (*stable unit treatment value assumption, SUTVA*) и может нарушаться при наличии внешних эффектов от воздействия между объектами (см. Angrist, Imbens & Rubin, 1996).

Для каждого объекта  $i$  определен вектор характеристик  $X_i$ , называемых ковариатами. Важным свойством ковариат является то, что изучаемое воздействие не оказывает на них никакого влияния. Примером ковариат является набор характеристик, которым обладал объект до того, как на него было оказано воздействие.

Каждый объект подвержен одному из двух типов воздействия:  $W_i = 0$ , если объект  $i$  подвергнут контрольному воздействию, и  $W_i = 1$ , если объект  $i$  подвергнут активному воздействию. Таким образом, для каждого объекта  $i$  мы наблюдаем тройку  $(W_i, Y_i, X_i)$ , где  $Y_i$  – реализовавшийся исход

$$Y_i \equiv Y_i(W_i) = \begin{cases} Y_i(0), & \text{если } W_i = 0, \\ Y_i(1), & \text{если } W_i = 1. \end{cases}$$

Следует отметить, что в реальности мы наблюдаем только один из этих исходов в зависимости от того, попал ли объект в контрольную или экспериментальную группу, в то время как второй (ненаблюдаемый) исход отвечает на вопрос «что бы было, если бы объект попал в другую группу».

*Мера склонности (propensity score)* определена как условная вероятность того, что объект будет подвергнут активному воздействию:

$$e(x) = \mathbb{P}\{W = 1|X = x\} = \mathbb{E}[W|X = x].$$

Также определим для  $w \in \{0, 1\}$  регрессии среднего и дисперсии

$$\mu_w(x) = \mathbb{E}[Y(w)|X = x]$$

и

$$\sigma_w^2(x) = \mathbb{V}[Y(w)|X = x].$$

## 2.2 Оцениваемый параметр

Для каждого объекта определен эффект воздействия  $\tau_i = Y_i(1) - Y_i(0)$ . В качестве параметра, который мы пытаемся оценить, чаще всего используется популяционный средний эффект воздействия

$$\tau_P = \mathbb{E}[Y(1) - Y(0)].$$

Также в качестве оцениваемого параметра может использоваться популяционный средний эффект воздействия на подвергшихся воздействию

$$\tau_{P,T} = \mathbb{E}[Y(1) - Y(0)|W = 1],$$

а также выборочные, а не популяционные, средние эффекты. Стоит отметить, что разница между этими параметрами имеет значение только в случае неоднородности эффекта воздействия. В том случае, когда эффект воздействия предполагается одинаковым для всех объектов, разница между этими оцениваемыми параметрами полностью нивелируется. К сожалению, в реальных ситуациях предположение об однородности эффектов практически всегда оказывается слишком сильным.

### 2.3 Предположения

Чтобы было возможно идентифицировать средний эффект воздействия, делаются следующие два предположения:

**Предположение 1. Несмешиваемость (unconfoundedness):** Пара  $(Y(0), Y(1))$  независима от  $W$  условно на  $X$ , т.е.  $(Y(0), Y(1)) \perp W|X$ .

Предположение о независимости означает, что, условно на ковариатах, распределение объектов по группам не зависит от потенциальных исходов для данного объекта и, в частности, не зависит от эффекта воздействия для данного объекта. Таким образом, исключаются случаи в которых, например, пациенты, которые больше нуждаются в получении лекарства (контролируя на ковариаты), будут с большей вероятностью включены в экспериментальную группу. Тот факт, что независимость предполагается условной на ковариатах, позволяет использовать данный подход в случаях, когда на распределение по подгруппам оказывают влияние определенные наблюдаемые характеристики, которые также могут быть связаны с потенциальными исходами.

Данное предположение также известно как предположение об условной независимости (Lechner, 1999) или отбор по наблюдаемым характеристикам (Heckman & Robb, 1985). По сути, данное предположение очень близко к предположению экзогенности в стандартных регрессионных моделях.

**Предположение 2. Пересечение (overlap):**  $0 < \mathbb{P}\{W = 1|X\} < 1$ .

Интуитивно, данное предположение означает, что не должно существовать такого значения ковариат, при котором мы можем однозначно утверждать, что объект будет принадлежать либо к контрольной группе, либо к группе воздействия. В случае нарушения данного предположения мы не можем оценить средний эффект воздействия для всей популяции, хотя мы можем оценить его для подмножества объектов, для которых данное предположение выполняется.

### 2.4 Общий подход

Для оценки эффекта воздействия тем или иным способом приходится сравнивать исходы в контрольной и экспериментальной группах. В чистом виде такое сравнение дает следующий результат:

$$\begin{aligned} \Delta &= \mathbb{E}[Y|W = 1] - \mathbb{E}[Y|W = 0] \\ &= \mathbb{E}[Y(1) - Y(0)|W = 1] + \mathbb{E}[Y(0)|W = 1] - \mathbb{E}[Y(0)|W = 0]. \end{aligned}$$

Первый член в этом выражении,  $\mathbb{E}[Y(1) - Y(0)|W = 1]$ , – это средний эффект воздействия на подвергшихся воздействию. Второй член,  $\mathbb{E}[Y(0)|W = 1] - \mathbb{E}[Y(0)|W = 0]$ , – это смещение из-за отбора (selection bias), который указывает, насколько отличались бы исходы в контрольной и экспериментальной группе, даже если никакого активного воздействия не было бы произведено. Таким образом, сравнение исходов в контрольной и экспериментальной группах позволяет оценить средний эффект воздействия в случае, если, во-первых, нет смещения из-за отбора, и, во-вторых, эффект воздействия на подвергшихся воздействию не отличается от эффекта воздействия на не подвергшихся воздействию ( $\tau_P = \mathbb{E}[Y(1) - Y(0)|W = 0]$ ).<sup>1</sup> Предположение о несмешиваемости обеспечивает выполнение обоих этих условий, что позволяет идентифицировать средний эффект воздействия. При этом предположении выполняется следующее равенство:

$$\mu_w(x) = \mathbb{E}[Y(w)|X = x] = \mathbb{E}[Y(w)|W = w, X = x] = \mathbb{E}[Y|W = w, X = x].$$

<sup>1</sup>В случае, если нарушается последнее условие, сравнение средних даст верную оценку среднего эффекта воздействия на подвергшихся воздействию, но не среднего эффекта воздействия.

Поскольку последнее выражение содержит только наблюдаемые величины, мы можем идентифицировать  $\mu_w(x)$ . Таким образом, для того, чтобы оценить средний эффект воздействия, мы можем вначале оценить средний эффект воздействия для подмножества с ковариатами  $X = x$ , а затем усреднить по всем таким подмножествам:

$$\begin{aligned}\tau(x) &= \mathbb{E}[Y(1) - Y(0)|X = x] = \mathbb{E}[Y(1)|X = x] - \mathbb{E}[Y(0)|X = x] \\ &= \mathbb{E}[Y(1)|W = 1, X = x] - \mathbb{E}[Y(0)|W = 0, X = x] \\ &= \mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x], \\ \tau_P &= \mathbb{E}[\tau(x)].\end{aligned}$$

### 3 Оценивание среднего эффекта воздействия

Существует целый ряд различных способов оценивания среднего эффекта воздействия. Основными подходами являются регрессионный, на основе мэтчинга, на основе меры склонности, а также их разные комбинации.

#### 3.1 Регрессии

Данный подход основан на получении состоятельной оценки двух регрессионных функций  $\mu_w(x)$ ,  $w \in \{0, 1\}$ . Пусть мы получили такие оценки  $\hat{\mu}_w(x)$ . В таком случае средний эффект воздействия оценивается как их разность, усредненная по эмпирическому распределению ковариат:

$$\hat{\tau}_{reg} = \frac{1}{N} \sum_{i=1}^N (\hat{\mu}_1(X) - \hat{\mu}_0(X)).$$

Таким образом, оценки регрессионных функций используются для вычисления вмененного (*imputed*) значения ненаблюдаемого потенциального исхода для каждого из объектов ( $Y_i(0)$  для объектов из экспериментальной группы и  $Y_i(1)$  для объектов из контрольной группы).

Наиболее простым способом оценки  $\mu_w(x)$  является параметрическое оценивание, например, при помощи линейной регрессии (Rubin, 1977). В частности, если мы используем регрессионную функцию  $\mu_w(x) = \beta'x + \tau\mu$ , то средний эффект воздействия может быть оценен при помощи простой регрессии  $Y_i = \alpha + \beta'X_i + \tau W_i + \varepsilon_i$ . Можно обобщить данный подход, оценивая отдельно регрессии для двух групп  $\mu_w(x) = \beta'_w x$ . Можно также использовать различные непараметрические методы оценивания регрессионной функции (Imbens, Newey & Ridder, 2005; Chen, Hong & Tarozzi, 2005; Heckman, Ichimura & Todd, 1997, 1998).

#### 3.2 Мэтчинг

Как и в случае регрессионного подхода, в оценках, использующих метод мэтчинга, при оценке эффекта воздействия для каждого из объектов вычисляются вмененные значения ненаблюдаемых потенциальных исходов. Однако, в случае мэтчинга в качестве вмененного значения используются наблюдаемые исходы объектов из другой группы воздействия, обладающих наиболее похожими значениями ковариат. Таким образом, для каждого объекта из экспериментальной группы ставится в соответствие один или несколько объектов из контрольной группы, с максимально похожими значениями ковариат. В качестве вмененного значения  $Y_i(0)$  для этого объекта используется усредненный исход этих объектов из контрольной группы. Аналогично, для каждого объекта из контрольной группы находится объект из экспериментальной группы с максимально похожим значением ковариат, и их усредненный исход используется в качестве вмененного значения  $Y_i(1)$ .

В простейшем случае дискретных ковариат можно добиться полного совпадения ковариат для различных объектов, и в качестве вмененного значения используется среднее значение исходов для всех объектов из противоположной группы воздействия, обладающих точно такими же значениями ковариат. Когда хотя бы одна ковариата является непрерывной, в общем случае добиться полного совпадения значений ковариат для разных объектов невозможно. В этом случае вводится метрика на пространстве ковариат и каждому объекту ставится в соответствие один или несколько объектов из противоположной группы, для которых расстояние до объекта минимально. При таком подходе исследователю необходимо выбрать используемую метрику, а также количество поставленных в соответствие объектов из противоположной группы. В подавляющем большинстве приложений используется метрика Махаланобиса, в соответствии с которой расстояние между точками вычисляется как  $d(x, y) = (x - y)' \Sigma^{-1} (x - y)$ , где  $\Sigma$  – ковариационная матрица ковариат. Основным достоинством этой метрики является то, что она не зависит от используемых единиц измерения.

Таким образом, единственным параметром, который остается выбрать исследователю, становится количество поставленных в соответствие объектов из противоположной группы. К сожалению, однозначных рекомендаций по выбору количества поставленных в соответствие объектов нет. Увеличивая их количество, мы увеличиваем эффективность оценки, одновременно увеличивая ее смещение. Abadie & Imbens (2006) показывают, что потери в эффективности от небольшого количества поставленных в соответствие объектов не очень велики, так что во многих приложениях оказывается достаточно ставить в соответствие лишь один объект. В прикладных исследованиях хорошим тоном считается проверка результатов на устойчивость по отношению к данному параметру.

Говоря более формально, оценка методом мэтчинга происходит следующим образом. Пусть у нас есть выборка  $\{(Y_i, X_i, W_i)\}_{i=1}^N$ . Обозначим через  $\ell_m(i)$  индекс объекта, находящегося на  $m$ -ом месте по расстоянию от объекта  $i$  среди объектов в противоположной группе. Формально,  $\ell_m(i)$  удовлетворяет следующим условиям:  $W_{\ell_m(i)} = 1 - W_i$ , и

$$\sum_{j: W_j = 1 - W_i} I \{ \|X_j - X_i\| < \|X_{\ell_m(i)} - X_i\| \} = m,$$

где  $I\{\cdot\}$  – индикатор-функция, принимающая значение 1, если выражение в фигурных скобках истинно, и 0 в противном случае. Обозначим через  $\mathfrak{J}_M(i)$  множество индексов для первых  $M$  объектов, поставленных в соответствие объекту  $i$ :  $\mathfrak{J}_M(i) = \{\ell_1(i), \dots, \ell_M(i)\}$ . Определим вмененные потенциальные исходы как

$$\hat{Y}_i(0) = \begin{cases} Y_i, & \text{если } W_i = 0, \\ M^{-1} \sum_{j \in \mathfrak{J}_M(i)} Y_j, & \text{если } W_i = 1, \end{cases}$$

$$\hat{Y}_i(1) = \begin{cases} M^{-1} \sum_{j \in \mathfrak{J}_M(i)} Y_j, & \text{если } W_i = 0, \\ Y_i, & \text{если } W_i = 1. \end{cases}$$

Простейшая мэтчинг-оценка – это

$$\hat{\tau}_M = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i(1) - \hat{Y}_i(0)).$$

Неприятным свойством простейшей мэтчинг-оценки является то, что она обладает смещением порядка  $O(N^{-1/K})$ , где  $K$  – количество непрерывных ковариат, использованных в мэтчинге (Abadie & Imbens, 2006). Таким образом, при  $K > 2$  оценка перестает быть  $N^{1/2}$ -состоятельной. Однако сочетание мэтчинга с регрессионными методами позволяет избавиться от этого смещения (см. ниже).

### 3.3 Мера склонности

Оценки с использованием меры склонности основаны на следующем результате (Rosenbaum & Rubin, 1983a). Пусть предположение о несмешиваемости верно. Тогда

$$(Y(0), Y(1)) \perp W | e(X). \quad (1)$$

Таким образом, чтобы избавиться от смещения в оценке, вместо контроля на полный набор ковариат достаточно контролировать лишь на меру склонности, условную вероятность того, что объект будет подвергнут активному воздействию. Существует несколько способов реализации данного метода, описанные ниже.

На практике оценивание с использованием меры склонности происходит в два этапа. На первом этапе оценивается сама мера склонности. Чаще всего это происходит путем оценивания пробит- или логит-модели, где зависимой величиной является индикаторная переменная, принимающая значение 1, если объект попал в экспериментальную группу, и 0, если объект попал в контрольную группу.<sup>2</sup> Затем предсказанное значение из оцененной регрессии используется в качестве меры склонности для дальнейшего оценивания среднего эффекта воздействия. Подобный метод оценивания реализован в пакете STATA (см. Becker & Ichino, 2002).

#### 3.3.1 Мера склонности и взвешивание

Использование меры склонности для взвешивания основано на следующих равенствах:

$$\mathbb{E} \left[ \frac{WY}{e(X)} \right] = \mathbb{E} [Y(1)]$$

и

$$\mathbb{E} \left[ \frac{(1-W)Y}{1-e(X)} \right] = \mathbb{E} [Y(0)],$$

из которых следует, что

$$\tau_P = \mathbb{E} \left[ \frac{WY}{e(X)} - \frac{(1-W)Y}{1-e(X)} \right].$$

Используя метод аналогий, можно использовать это неравенство для построения оценки

$$\tilde{\tau} = \frac{1}{N} \sum_{i=1}^N \left[ \frac{W_i Y_i}{\hat{e}(X_i)} - \frac{(1-W_i) Y_i}{1-\hat{e}(X_i)} \right].$$

Однако сумма весов в подобной оценке на конечных выборках может отличаться от единицы, что является нежелательным свойством. От этого свойства легко избавиться, скорректировав соответствующим образом веса, в результате чего получаем следующую оценку среднего эффекта воздействия (Hirano, Imbens & Ridder, 2003):

$$\hat{\tau}_{weight} = \left( \sum_{i=1}^N \frac{W_i}{\hat{e}(X_i)} \right)^{-1} \sum_{i=1}^N \frac{W_i Y_i}{\hat{e}(X_i)} - \left( \sum_{i=1}^N \frac{(1-W_i)}{1-\hat{e}(X_i)} \right)^{-1} \sum_{i=1}^N \frac{(1-W_i) Y_i}{1-\hat{e}(X_i)}.$$

<sup>2</sup>В более продвинутых методах может производиться непараметрическая оценка меры склонности (например, см. Hirano, Imbens & Ridder, 2003).

### 3.3.2 Мера склонности и блокирование

В данном методе после оценивания меры склонности все наблюдения разбиваются на  $M$  блоков, внутри которых объекты обладают примерно одинаковой вероятностью быть подвергнутым активному воздействию. В простейшем случае единичный интервал разбивается на  $M$  блоков одинакового размера с границами  $t/M$ , где  $t = 1, \dots, M - 1$ . Обозначим как  $J_{im}$  индикатор того, что объект  $i$  попал в блок  $t$ . В этом случае

$$J_{im} = I \{(m - 1)/M < e(X_i) \leq t/M\}.$$

Обозначим число объектов из группы  $w \in \{0, 1\}$ , попавших в блок  $t$ , как  $N_{wt}$ . Внутри каждого блока эффект воздействия оценивается так, как если бы распределение по группам происходило случайным образом:

$$\hat{\tau}_m = \frac{1}{N_{1m}} \sum_{i=1}^N J_{im} W_i Y_i - \frac{1}{N_{0m}} \sum_{i=1}^N J_{im} (1 - W_i) Y_i.$$

Средний эффект воздействия оценивается как

$$\hat{\tau}_{block} = \sum_{i=1}^M \hat{\tau}_m \frac{N_{1m} + N_{0m}}{N}.$$

В данном подходе единственным параметром, который необходимо выбрать исследователю, является количество блоков. На практике оказывается, что уже пять блоков оказывается достаточным количеством, чтобы избавиться от большей части смещения в оценке (Cochran, 1968; Rosenbaum & Rubin, 1983b; Dehejia & Wahba, 1999). Количество блоков может быть увеличено при наличии достаточного количества наблюдений, особенно если внутри блоков распределение ковариат у контрольной и экспериментальной групп существенно отличается.

## 3.4 Смешанные методы

### 3.4.1 Мэтчинг и регрессия

Наиболее интересным из смешанных методов является сочетание мэтчинга и регрессии. В данном подходе в простую оценку, получаемую при помощи мэтчинга, вносится поправка, учитывающая разницу в ковариатах между поставленными в соответствие объектами. Формально, в качестве вмененных исходов используются

$$\begin{aligned} \hat{Y}_i(0) &= \begin{cases} Y_i, & \text{если } W_i = 0, \\ M^{-1} \sum_{j \in \mathfrak{J}_M(i)} (Y_j + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_j)), & \text{если } W_i = 1, \end{cases} \\ \hat{Y}_i(1) &= \begin{cases} M^{-1} \sum_{j \in \mathfrak{J}_M(i)} (Y_j + \hat{\mu}_1(X_i) - \hat{\mu}_1(X_j)), & \text{если } W_i = 0, \\ Y_i, & \text{если } W_i = 1, \end{cases} \end{aligned}$$

после чего эти вмененные исходы используются для построения оценки среднего эффекта воздействия так же, как и при простом мэтчинге. Основным достоинством данного метода по сравнению с простым мэтчингом является то, что он позволяет (асимптотически) избавиться от смещения, делая оценку  $N^{1/2}$ -состоятельной (Abadie & Imbens, 2006). Данный метод реализован в пакете STATA (см. Abadie, Drukker, Herr & Imbens, 2003).

### 3.4.2 Взвешивание и регрессия

Оценивание с использованием взвешивания может быть переформулировано как оценивание методом наименьших квадратов функции

$$Y_i = \alpha + \tau W_i + \varepsilon_i$$

с весами, равными

$$\lambda_i = \sqrt{\frac{W_i}{e(X_i)} + \frac{1 - W_i}{1 - e(X_i)}}.$$

В данной формулировке мы можем легко обобщить данный подход и включить дополнительные ковариаты, чтобы увеличить точность оценок. В этом случае оценивается следующая регрессия:

$$Y_i = \alpha + \beta' X_i + \tau W_i + \varepsilon_i$$

с использованием тех же самых весов  $\lambda_i$ . Данный подход описан в Robins & Ritov (1997) и Hirano & Imbens (2001). Одним из достоинств этого и большинства других смешанных методов является так называемая «двойная устойчивость» – полученная оценка является состоятельной, если хотя бы один из используемых подходов верен. В данном случае оценка состоятельна, если регрессионная модель или мера склонности верно специфицированы.

### 3.4.3 Блокирование и регрессия

Оценка с использованием блокирования может быть записана как среднее оценок по каждому из блоков, которые оцениваются в регрессиях

$$Y_i = \alpha_m + \tau_m W_i + \varepsilon_i,$$

с использованием только наблюдений из блока  $m$ . Данный подход также легко обобщается с целью учесть эффект ковариат и увеличить точность оценки. В этом случае эффект внутри каждого из блоков оценивается как

$$Y_i = \alpha_m + \beta'_m X_i + \tau_m W_i + \varepsilon_i,$$

также с использованием только наблюдений из блока  $m$ . Затем оценки, полученные на разных блоках, усредняются, как и в обычном методе блокирования.

### 3.4.4 Мера склонности и регрессия

В некоторых прикладных работах мера склонности учитывается лишь в качестве дополнительного контроля в регрессии исхода  $Y_i$  на индикаторе принадлежности объекта к экспериментальной группе  $W_i$ . Асимптотические свойства подобных оценок не изучены. Кроме того, не совсем понятны преимущества данного метода по сравнению с простой регрессией на ковариаты, используемые при оценке меры склонности. По указанным выше причинам использование данного метода не рекомендуется.

### 3.4.5 Мера склонности и мэтчинг

Еще одним методом оценивания эффекта воздействия является мэтчинг, в котором в качестве единственной ковариаты используется мера склонности. Поскольку в этом случае размерность ковариат равна единице, получающаяся оценка является  $N^{1/2}$ -состоятельной и асимптотически нормальной. Если истинное значение меры склонности известно, то дисперсия этой оценки может быть оценена методом, предложенном в Abadie & Imbens (2006) и описанному в следующем разделе. К сожалению, дисперсия оценки, использующей оценку меры склонности, а не ее истинное значение, на данный момент не известна, что затрудняет использование этого метода на практике.



### 3.5 Оценивание дисперсии

Наиболее распространенным подходом к оцениванию дисперсии является бутстрап (Efron & Tibshirani, 1993; Horowitz, 2001). В большинстве случаев этот метод дает разумные оценки дисперсии. Однако для оценок с использованием мэтчинга при фиксированном числе объектов, ставящихся в соответствие, бутстрап дает неверные оценки (Abadie & Imbens, 2008).

Альтернативный метод получения оценок дисперсии описан в Abadie & Imbens (2006). Он основан на том, что большинство описанных выше оценок могут быть записаны в виде

$$\hat{\tau} = \sum_{i=1}^N \lambda_i Y_i, \text{ где } \sum_{i:W_i=1} \lambda_i = 1 \text{ и } \sum_{i:W_i=0} \lambda_i = -1,$$

и веса  $\lambda_i$  являются функцией ковариат и индикаторов групповой принадлежности для всех наблюдений. В этом случае дисперсия оценки равна

$$\mathbb{V}[\hat{\tau}|X_1, \dots, X_N, W_1, \dots, W_N] = \sum_{i=1}^N \lambda_i^2 \sigma_{W_i}^2(X_i).$$

Таким образом, для нахождения оценки дисперсии мы должны построить оценку  $\hat{\sigma}_{W_i}^2(X_i)$  для всех наблюдений. Для построения этих оценок Abadie & Imbens (2006) предлагают использовать мэтчинг. Идея заключается в том, что каждому объекту  $i$  мы ставим в соответствие объект  $\nu(i)$  из той же самой группы (т.е.  $W_i = W_{\nu(i)}$ ) с максимально похожими ковариатами. В качестве оценки  $\hat{\sigma}_{W_i}^2(X_i)$  используется

$$\hat{\sigma}_{W_i}^2(X_i) = \frac{(Y_i - Y_{\nu(i)})^2}{2}.$$

Получающаяся оценка  $\hat{\sigma}_{W_i}^2(X_i)$  для конкретного  $i$  не является состоятельной, однако агрегированная оценка

$$\hat{V} = \sum_{i=1}^N \lambda_i^2 \hat{\sigma}_{W_i}^2(X_i)$$

является состоятельной оценкой для дисперсии оценки среднего эффекта воздействия.

## 4 Оценка предположений

В предыдущем разделе описан ряд методов, позволяющих получить оценку среднего эффекта воздействия при условии выполнения предположений о несмешиваемости и пересечении. Ниже мы рассмотрим, каким образом можно попытаться оценить степень достоверности этих предположений.

### 4.1 Несмешиваемость

Предположение о несмешиваемости является принципиально нетестируемым, так как оно касается ненаблюдаемых характеристик (нереализованных исходов). Однако существует методы, которые хоть и не доказывают данное предположение, но делают его более достоверным (Heckman & Hotz, 1989; Rosenbaum, 1987). Подобные методы основаны на тестировании гипотезы о равенстве нулю эффекта воздействия в тех случаях, когда этот эффект должен быть равен нулю. В случае, если эта гипотеза отвергается, предположение о несмешиваемости становится более уязвимым. Существуют два основных класса подобных тестов.

К первому классу относятся тесты, оценивающие эффект воздействия, которое заведомо не должно отразиться на исходе. Данный подход основывается на наличие двух или

более контрольных групп, одна из которых используется в качестве фиктивной экспериментальной группы. Если сравнение двух контрольных групп указывает на наличие эффекта воздействия, то это означает, что как минимум одна из контрольных групп не является валидной и обладает системным смещением. Если гипотеза о нулевом эффекте воздействия не отвергается, предположение о несмешиваемости все равно может не выполняться, так как сравниваемые контрольные группы могут обладать одинаковым смещением. Именно поэтому следует выбирать для сравнения такие контрольные группы, относительно которых можно предполагать, что они обладают различным смещением. Например, если мы изучаем воздействие некоторой экономической программы на индивидов, мы можем разделить людей, не подвергшихся воздействию данной программы, на тех, кто не мог участвовать в программе по условиям программы, и тех, кто сам решил не принимать в ней участия. Наличие значимых различий между такими контрольными группами может свидетельствовать о наличии смещения, вызванного эффектом самоотбора.

Формально: пусть у нас имеется индикатор принадлежности к группе  $G_i \in \{-1, 0, 1\}$ , где значения  $\{-1, 0\}$  соответствуют контрольным группам:

$$W_i = \begin{cases} 1, & \text{если } G_i \in \{-1, 0\}, \\ 0, & \text{если } G_i = 1. \end{cases}$$

Вместо нетестируемого предположения о несмещенности мы можем проверить более сильное утверждение

$$(Y(0), Y(1)) \perp G|X,$$

из которого следует не только несмешиваемость, но и условие

$$Y \perp G|X, G \in \{-1, 0\},$$

которое можно проверить при помощи тестирования гипотезы о том, что

$$\mathbb{E}[\mathbb{E}[Y|G = -1, X] - \mathbb{E}[Y|G = 0, X]] = 0.$$

Ко второму классу относятся тесты, тестирующие эффект воздействия, где в качестве исхода используется характеристика, которая заведомо не зависит от изучаемого воздействия. В большинстве случаев в качестве такого псевдо-исхода используется характеристика, которая определяется до момента активного воздействия. В частности, в качестве такой характеристики может выступать лагированное значение интересующего нас исхода. В случае, если оцениваемый эффект воздействия отличен от нуля, это означает, что объекты в контрольной группе значимо отличаются от объектов в экспериментальной группе даже в отсутствие воздействия. Если же эффект отсутствует, то это хоть и не доказывает предположение о несмешиваемости, но делает его более достоверным.

## 4.2 Пересечение

В отличие от предположения о несмешиваемости, предположение о пересечении может быть оценено напрямую. Для оценки степени пересечения можно прежде всего сравнить сводную статистику распределения ковариат у контрольной и экспериментальной групп. Одним из стандартных индикаторов является нормализованная разность между средними значениями ковариаты в контрольной и экспериментальной группах:

$$\Delta_X = \frac{\overline{X_1} - \overline{X_0}}{\sqrt{S_1^2 + S_0^2}},$$

где  $S_1^2$  и  $S_0^2$  – выборочные дисперсии  $X$  в экспериментальной и контрольной группах, соответственно. В случае хорошего пересечения нормализованная разность не должна превышать единицу ни по одной из ковариат.

Сравнение ковариат по отдельности во многих случаях оказывается недостаточным, так как оно не учитывает особенностей совместного распределения ковариат (т.е. определенное сочетание ковариат может встречаться только в экспериментальной или только в контрольной группе). Для того чтобы оценить пересечение в совместном распределении ковариат, можно смотреть на распределение оцененной меры склонности. В случае хорошего пересечения наблюдения не должны быть слишком близкими ни к нулю, ни к единице.

В том случае, если оказывается, что предположение о пересечении вызывает сомнения, существует ряд способов, позволяющих в той или иной степени исправить ситуацию. Одним из наиболее распространенных методов является урезание выборки. В этом подходе из рассмотрения исключаются наблюдения, попадающие в области, где не наблюдается хорошего пересечения. При таком подходе страдает внешняя валидность метода, так как эффект оценивается лишь на определенной подвыборке наблюдений. Однако при этом сохраняется внутренняя валидность, поскольку на этой подвыборке мы получаем достоверную оценку эффекта. На практике это означает, что после оценивания меры склонности из рассмотрения исключаются наблюдения со слишком большим или слишком маленьким значением меры склонности. Crump, Hotz, Imbens & Mitnik (2008) показывают, что достаточно хороших результатов можно добиться, исключив наблюдения со значением меры склонности меньше 0,1 и больше 0,9.

## 5 Заключение

В данном эссе мы рассмотрели ряд методов, позволяющих оценить средний эффект воздействия. На практике наиболее важным является проверка уместности предположений о несмешиваемости и пересечении. В том случае, если эти предположения выглядят убедительно, выбор конкретного метода оценивания оказывается не столь важным. Хорошим стилем является проверка получаемых результатов на устойчивость относительно метода оценивания.

В практических приложениях все большую популярность приобретают методы, основанные на мэтчинге. В течение долгого времени одним из препятствий на пути их развития была вычислительная сложность, которая становится все менее важной по мере развития компьютерной техники, хотя использование подобных методов при большом количестве наблюдений и ковариат по-прежнему остается затруднительным.

## Список литературы

- Angrist, J.D., G.W. Imbens & D.B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of American Statistical Association* 91, 444–455.
- Abadie, A., D. Drukker, H. Herr & G. Imbens (2003). Implementing matching estimators for average treatment effects in STATA. *Stata Journal* 4, 290–311.
- Abadie, A. & G. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* 74, 235–267.
- Abadie, A. & G. Imbens (2008). On the failure of the bootstrap for matching estimators. *Econometrica* 76, 1537–1557.
- Becker, S. & A. Ichino (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal* 2, 358–377.
- Chen, X., H. Hong & A. Tarozzi (2005). Semiparametric efficiency in GMM models of non-classical measurement errors, missing data and treatment effects. Препринт, New York University.
- Cochran, W.G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 24, 295–314.

- Crump, R., V.J. Hotz, G. Imbens & O.Mitnik (2008). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, в печати.
- Dehejia, R. & S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of American Statistical Association* 94, 1053–1062.
- Efron, B. & R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
- Heckman, J. & J. Hotz (1989). Alternative methods for evaluating the impact of training programs (with discussion). *Journal of American Statistical Association* 84, 862–874.
- Heckman, J., H. Ichimura & P. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies* 64, 605–654.
- Heckman, J., H. Ichimura & P. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies* 65, 261–294.
- Heckman, J. & R. Robb (1985). Alternative methods for evaluating the impact of interventions. Глава в *Longitudinal Analysis of Labor Market Data* (под редакцией Heckman, J.J. & B.S. Singer). Cambridge University Press.
- Hirano, K. & G. Imbens (2001). Estimation of causal effects using propensity score weighting: An application of data on right heart catheterization. *Health Services and Outcomes Research Methodology* 2, 259–278.
- Hirano, K., G. Imbens & G. Ridder (2003). Mean-squared-error calculations for average treatment effects. Препринт, University of California–Berkeley.
- Horowitz, J.L. (2001). The bootstrap. Глава 52 в *Handbook of Econometrics* (под редакцией Heckman, J.J. & E.E. Leamer), том 5. Elsevier Science.
- Lechner, M. (1999). Earnings and employment effects of continuous off-the-job training in east Germany after unification. *Journal of Business & Economic Statistics* 17, 74–90.
- Imbens, G., W. Newey & G. Ridder (2005). Recent developments in the econometrics of program evaluation. NBER препринт №14251.
- Imbens, G. & J. Wooldridge (2008). Recent developments in the econometrics of program evaluation. NBER препринт №14251.
- Robins, J. & Y. Ritov (1997). Towards a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* 16, 285–319.
- Rosenbaum, P. (1987). The role of a second control group in an observational study (with discussion). *Statistical Science* 2, 292–316.
- Rosenbaum, P. & D. Rubin (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.
- Rosenbaum, P. & D. Rubin (1983b). Assessing the sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of Royal Statistical Society Series B* 45, 212–218.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* 66, 688–701.
- Rubin, D. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* 2, 1–26.
- Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6, 34–58.

## Estimation of treatment effects

Ruben Enikolopov

*New Economic School, Moscow, Russia*

The essay contains a short survey of methods used for estimation of average treatment effects, when the independent variable of interest is binary.