квантиль

международный эконометрический журнал
на русском языке

№6
март 2009 г.

СОДЕРЖАНИЕ ВЫПУСКА

квантиль

№6, март 2009 г.

Сайт в Интернете: http://quantile.ru
Адрес электронной почты: quantile@quantile.ru
Доступ к журналу бесплатный и неограниченный

К СВЕДЕНИЮ АВТОРОВ

Рукописи для публикации в разделе «Статьи» принимаются в электронном виде по адресу *submit@quantile.ru*. Работы могут принадлежать любой прикладной сфере экономической науки. Главным требованием является интенсивное использование адекватных эконометрических методов. Рукопись должна быть написана на русском (для русскоязычных авторов) или на английском (для остальных авторов) языке в формате *Microsoft Word* или (предпочтительнее) *LaTeX*, и по объему не превышать 30 страниц формата А4 с двойным междустрочным интервалом. Работы подвергаются контролю качества членами редакционного совета и независимыми референтами. Перспективная работа может быть при необходимости возвращена автору на доработку. Редакция также приглашает к сотрудничеству экспертов по эконометрике, готовых внести вклад в методологические рубрики журнала.

Уважаемые читатели!

Отныне журнал «*Квантиль*» индексируется в электронной библиографии Американской Экономической Ассоциации EconLit, см. www.aeaweb.org/econlit. Кроме того, тематическая электронная библиотека для исследований и образования Университетская информационная система РОССИЯ внесла «*Квантиль*» в список источников, формирующих эту библиотеку, см. uisrussia.msu.ru.

Редакционный совет благодарит нижеперечисленных экономистов и статистиков, принявших, наряду с членами совета, участие в рецензировании поступивших в *«Квантиль»* статей:

Вымятнина Юлия (Европейский университет в Санкт-Петербурге)
Денисова Ирина (ЦЭФИР в Российской экономической школе)
Жукова Надежда (Стэндфордский университет)
Замулин Олег (Российская экономическая школа)
Ицхоки Олег (Гарвардский университет)
Сосунов Кирилл (Высшая школа экономики)
Шабалин Андрей (Университет штата Северная Каролина)

Редактор и редакционный совет.

Эконометрический ликбез: эффекты воздействия

Оценивание эффекта воздействия*

Рубен Ениколопов†

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 $i = 1, \ldots, N$. 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*. 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.

†Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1717. Электронная почта: *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 наличие 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

$$\sigma^2_w(x) = \mathbb{V}\left[Y(w)|X = x\right].$$

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. Стоить 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 получение 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 ($\tau_{P,} = \mathbb{E}[Y(1)-Y(0)|W=0]$).[1] Предположение о несмешиваемости обеспечивает выполнение обоих этих условий, что позволяет идентифицировать средний эффект воздействия. При этом предположении выполняется следующее равенство:

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

[1] В случае, если нарушается последние условие, сравнение средних даст верную оценку среднего эффект 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:

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

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 раздельно 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 экспериментально 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 Σ – 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

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

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: $\mathfrak{J}_M(i) = \{\ell_1(i), \ldots, \ell_M(i)\}$. Определим вмененные потенциальные исходы как

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

Простейшая мэтчинг-оценка – это

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

Неприятным свойством простейшей мэтчинг-оценки является то, что она обладает смещением порядка $O\left(N^{-1/K}\right)$, 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

$$(Y(0), Y(1)) \perp W | e(X). \tag{1}$$

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.[2] 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:

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

и

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

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

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

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):

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

[2] 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 m/M, где $m = 1, \dots, M-1$. Обозначим как J_{im} индикатор того, что объект i попал в блок m. В этом случае

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

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

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

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 наличие 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

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

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

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) избавится 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 λ_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 оценки 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 λ_i являются функцией ковариат и индикаторов групповой принадлежности для всех наблюдений. В этом случае дисперсия оценки равна

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

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

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

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

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

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 меду 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 пересечение 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 hear catherization. *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 Califormia–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 semiparametric 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.

Эффекты воздействия*

Уитни К. Ньюи†

Массачусетсский Технологический Институт, Кэмбридж, США

Данное эссе посвящено вопросам идентификации и оценивания среднего эффекта воздействия и среднего эффекта воздействия на подвергшихся воздействию.

Введение

В работах, посвященных эффектам воздействия, исследуется, как на некоторый интересующий исследователя исход, например, на заработную плату, влияют некоторые другие параметры, такие как программы по обучению персонала. Очевидно, что эффекты воздействия имеют отношение к структурным моделям, в которых интересующий исследователя параметр находится в левой части регрессионного уравнения, а переменные, отвечающие за воздействия – в правой. Действительно, как мы увидим, модель с эффектами воздействия можно рассматривать как линейную структурную модель со случайными коэффициентами. Для облегчения восприятия в данной работе будет введена специальная терминология, общепринятая для моделей с эффектами воздействия.

Объекты мы будем нумеровать с помощью индекса i, а через D_i обозначим индикатор воздействия, принимающий значение 1, если объект был подвержен этому воздействию, и значение 0 в противном случае. Например, $D_i = 1$ может означать, что индивид с номером i принимал участие в программе по обучению персонала или проходил медицинское лечение. Для того, чтобы описать эффект воздействия, необходимо ввести еще два параметра. Пусть Y_{i0} обозначает возможный исход, полученный в случае, если объект не был подвержен воздействию ($D_i = 0$), а Y_{i1} – возможный исход в случае, если объект был подвержен воздействию ($D_i = 1$). Ясно, что эти два параметра ненаблюдаемы. Один из них является «гипотетическим»: он обозначает исход, который мог бы быть получен, если бы применялось обратное воздействие. Тогда наблюдаемый исход будет выражаться как

$$Y_i = D_i Y_{i1} + (1 - D_i) Y_{i0}.$$

Эффект воздействия для объекта i определяется как

$$\beta_i = Y_{i1} - Y_{i0}.$$

Этот параметр неидентифицируем, поскольку наблюдается только один из возможных исходов. Но существуют некоторые другие параметры, которые при определенных условиях можно оценить. Один из них – это популяционный средний эффект воздействия (average treatment effect):

$$ATE \stackrel{\text{def}}{=} \mathbb{E}[\beta_i].$$

*Перевод Е. Скиба и С. Анатольева. Эссе является материалом по курсу “New Econometric Methods,” прочитанного весной 2007 г. в Массачусетсском Технологическом Институте, в рамках MIT OpenCourseWare (ocw.mit.edu). Цитировать как: Ньюи, Уитни К. (2009) «Эффекты воздействия», Квантиль, №6, стр. 15–23. Citation: Newey, Whitney K. (2009) “Treatment effects,” Quantile, No.6, pp. 15–23.

†Адрес: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Building E52, Room 262D, Cambridge, MA 02142-1347, USA. Электронная почта: *wnewey@mit.edu*

Он описывает эффект воздействия, усредненный по всей популяции объектов. Другой интересный параметр – это популяционный средний эффект воздействия на подвергшихся воздействию (average effect of treatment on treated):

$$ATT \stackrel{\text{def}}{=} \mathbb{E}[\beta_i | D_i = 1].$$

Он описывает эффект воздействия, усредненный по всей популяции подвергшихся воздействию объектов. Третий важный и интересный для исследования параметр называется локальным средним эффектом воздействия. Он будет введен нами в дальнейшем.

Для того чтобы лучше понять воздействия и разнообразные эффекты, которые они оказывают, установим связь между нашей моделью и регрессионной моделью со случайными коэффициентами. Из предыдущего уравнения для Y_i получаем:

$$Y_i = Y_{i0} + (Y_{i1} - Y_{i0})D_i = \alpha_i + \beta_i D_i,$$

$$\alpha_i = Y_{i0}, \ \ \beta_i = Y_{i1} - Y_{i0}.$$

Таким образом, Y_i можно представить в виде линейной модели, в которой эффект воздействия β_i является коэффициентом при D_i, а константа α_i и угол наклона β_i меняются в зависимости от объекта. Следовательно, ATE есть средний по всей популяции угол наклона, а ATT – угол наклона, усредненный по подмножеству всех объектов популяции, имеющих $D_i = 1$.

Определение этих случайных коэффициентов помогает понять историю возникновения теории эффектов воздействия. Коэффициент $\beta_i = Y_{i1} - Y_{i0}$ иногда называют «гипотетическим», так как он объясняет, как менялся бы Y_i при изменении D_i. В моделях спроса и предложения используется понятие «движения вдоль кривой». Оно появилось в экономике благодаря Райту (Wright, 1928), который нашел ему оригинальное объяснение в рамках этой модели. Точно так же средний эффект воздействия – это просто ожидаемое значение случайного коэффициента в линейной модели, то есть средний угол наклона кривой.

Разнообразные допущения в модели позволяют оценить ATT и ATE. В данной работе будут обсуждаться некоторые условия, при которых возможно идентифицировать эти параметры. Доказательства будут проводиться путем демонстрации того, как эти параметры можно представить в виде ожидаемых значений данных.

Начнем с самого простого случая.

Однородные эффекты воздействия

Простейший случай в этой модели – случай однородных эффектов воздействия, когда $\beta_i = \bar{\beta}$, то есть эффект воздействия один и тот же для всех объектов. Тогда ATT и ATE – это просто $\bar{\beta}$. В этом случае, обозначив $\bar{\alpha} = \mathbb{E}[\alpha_i]$ и $\varepsilon_i = \alpha_i - \bar{\alpha}$, получаем

$$Y_i = \bar{\alpha} + \bar{\beta} D_i + \varepsilon_i.$$

Тогда модель сводится к обычной линейной модели с аддитивными ошибками и постоянными коэффициентами. В противоположность этому в базовую модель также входят аддитивные ошибки, но углы наклона случайны. Обратим внимание на то, что случайность α_i эквивалентна ее представлению в виде суммы константы и возмущения: $\alpha_i = \bar{\alpha} + \varepsilon_i$.

Идентифицировать и оценить $\bar{\beta}$ и $\bar{\alpha}$ можно обычным способом, если у нас имеется инструмент Z_i, некоррелированный с ε_i и коррелированный с D_i, так что

$$0 = \mathbb{C}(Z_i, \varepsilon_i) = \mathbb{C}(Z_i, \alpha_i) = \mathbb{C}(Z_i, Y_{i0}),$$

$$\mathbb{C}(Z_i, D_i) \neq 0.$$

В этом случае коэффициенты можно идентифицировать через обычные инструментальные уравнения:

$$\bar{\beta} = \frac{\mathbb{C}(Z_i, Y_i)}{\mathbb{C}(Z_i, D_i)}.$$

Оценить эти коэффициенты можно стандартным образом, заменив популяционную ковариацию на выборочную. Если подвести итог, в данном разделе не было введено ничего нового, кроме использования терминологии из теории эффектов воздействия в стандартной модели с фиктивными переменными.

Предположение об однородности эффектов воздействия является слишком сильным во многих ситуациях. Например, можно считать, что влияние на заработную плату программ по повышению квалификации или влияние уменьшения числа учеников в классе на уровень образования одинаковы для каждого объекта. Однако непохоже, что эти предположения выполняются в действительности. Поэтому в дальнейшем будем считать, что коэффициент β_i меняется в зависимости от объекта.

Случайное распределение

Случайное распределение означает, что исход не зависит от того, был ли объект подвержен воздействию или нет. Здесь мы делаем статистическое предположение

$$\mathbb{E}[Y_{i0}|D_i] = \mathbb{E}[Y_{i0}],$$

то есть среднее значение исхода по всем неподвергшимся воздействию объектам, не зависит от статуса подверженности воздействию. То же самое условие можно записать как $\mathbb{E}[\alpha_i|D_i] = 0$. Оно является более общим, чем условие независимости, поскольку позволяет моментам Y_{i0} более высокого порядка зависеть от D_i. Тем не менее, сложно представить ситуацию, в которой будет выполняться условие независимости среднего без полной независимости.

Для того чтобы понять, что происходит при этом предположении, заметим сначала, что

$$\mathbb{E}[\beta_i|D_i]D_i = \begin{cases} 0, & D_i = 0, \\ \mathbb{E}[\beta_i|D_i = 1], & D_i = 1, \end{cases} = \mathbb{E}[\beta_i|D_i = 1]D_i.$$

Тогда в условии независимости среднего получаем:

$$\begin{aligned} \mathbb{E}[Y_i|D_i] &= \mathbb{E}[\alpha_i + \beta_i D_i|D_i] = \mathbb{E}[\alpha_i] + \mathbb{E}[\beta_i|D_i]D_i \\ &= E[\alpha_i] + \mathbb{E}[\beta_i|D_i = 1]D_i. \end{aligned}$$

В данном случае фиктивная переменная в регрессии Y_i на константу и D_i имеет в качестве углового коэффициента ATT. Если кроме того мы предположим, что среднее значение Y_{i1} не зависит от D_i, то есть

$$\mathbb{E}[Y_{i1}|D_i] = \mathbb{E}[Y_{i1}],$$

то в результате имеем $ATE = ATT$, поскольку

$$\begin{aligned} \mathbb{E}[\beta_i|D_i = 1] &= \mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 1] \\ &= \mathbb{E}[Y_{i1}] - \mathbb{E}[Y_{i0}] = \mathbb{E}[\beta_i]. \end{aligned}$$

В итоге мы получили, что когда Y_{i0} независимо от D_i в среднем, ATT представляется как дамми-коэффициент в регрессии исхода на константу и индикатор воздействия. Также мы показали, что если, кроме того, Y_{i1} независимо от D_i в среднем, ATE является тем же самым коэффициентом. Конечно, этот коэффициент можно оценить с помощью линейной регрессии Y_i на $(1, D_i)$. Более того, как и следовало ожидать, этот коэффициент в регрессии является единственным, что отличает средние значения Y_i для подвергшихся и не подвергшихся воздействию объектов.

Обсуждение

Для многих приложений случайное распределение – слишком сильное допущение. Обычно объекты могут выбирать, хотят ли они подвергаться воздействию или нет, например, покидая выборку в случае, если условия воздействия их не устраивают. Люди могут уклоняться от программ по повышению квалификации или не пользоваться медицинским обслуживанием. Если эти решения связаны с (α_i, β_i), то мы не можем говорить о независимости (α_i, β_i) и D_i. В рамках линейной модели $Y_i = \alpha_i + \beta_i D_i$ возможна эндогенность, при которой D_i может коррелировать со случайными коэффициентами α_i и β_i. Этот случай сложнее стандартного, поскольку угол наклона β_i также может коррелировать с D_i.

Существуют два подхода к этой проблеме. Первый (знакомый) подход состоит в применении инструментальных переменных. Второй подход получил название «отбор по наблюдаемым характеристикам». В этом случае благодаря анализу, условному на некоторые наблюдаемые переменные, исчезает корреляция между D_i и (α_i, β_i). Поскольку использование инструментов является наиболее известным и распространенным способом решения этой проблемы, мы сначала его и рассмотрим.

Инструментальная идентификация эффектов воздействия

В обычной линейной модели, для которой однородность эффектов воздействия – лишь частный случай, для нахождения угла наклона необходимо предположить некоррелируемость инструментов с возмущениями и их коррелируемость с D_i. Похожие условия будут использоваться для идентификации эффектов воздействия с помощью инструментов. Пусть Z_i – инструмент. Будем предполагать, что

$$\mathbb{E}[\alpha_i|Z_i] = \mathbb{E}[Y_{i0}|Z_i] = \mathbb{E}[Y_{i0}] = \mathbb{E}[\alpha_i],$$

то есть средний эффект у всех неподвергшихся воздействию объектов независим от инструментов.

Мы также рассмотрим случай, когда Z_i является фиктивной переменной, то есть когда $Z_i \in \{0,1\}$ с $P = \mathbb{P}\{Z_i = 1\}$ и $0 < P < 1$. (Вопрос: зачем мы предполагаем $0 < P < 1$?). Для дамми-инструмента существует полезная формула для ковариации между инструментом и любой другой случайной величиной W_i. В частности, можно получить, что

$$\begin{aligned}
\mathbb{C}(W_i, Z_i) &= \mathbb{E}[W_i Z_i] - \mathbb{E}[W_i]\mathbb{E}[Z_i] = \left(\frac{\mathbb{E}[W_i Z_i]}{P} - \mathbb{E}[W_i]\right) P \\
&= (\mathbb{E}[W_i|Z_i = 1] - \mathbb{E}[W_i])P \\
&= \{\mathbb{E}[W_i|Z_i = 1] - (P\mathbb{E}[W_i|Z_i = 1] + (1-P)\mathbb{E}[W_i|Z_i = 0])\}P \\
&= (\mathbb{E}[W_i|Z_i = 1] - \mathbb{E}[W_i|Z_i = 0])P(1-P).
\end{aligned}$$

То есть ковариация между W_i и Z_i представляется в виде разницы двух условных средних величин при различных значениях Z_i, умноженной на $P(1-P)$.

Из этой формулы можно сделать два полезных вывода. Во-первых, независимость среднего Y_{i0} от Z_i эквивалентна некоррелированности Y_{i0} и Z_i, что верно поскольку $\mathbb{C}(Y_{i0}, Z_i) = 0$ тогда и только тогда, когда $\mathbb{E}[Y_{i0}|Z_i = 1] = \mathbb{E}[Y_{i0}|Z_i = 0]$. Во-вторых, из этой формулы можно получить выражение для предела инструментальной оценки угла наклона:

$$\frac{\mathbb{C}(Z_i, Y_i)}{\mathbb{C}(Z_i, D_i)} = \frac{\mathbb{E}[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - E[D_i|Z_i = 0]}.$$

Этот результат называют Вальдовской формулой для инструментальных переменных, ссылаясь на работу Wald (1940), в которой Вальд использовал фиктивные переменные в качестве инструмента для решения проблемы ошибок измерения.

Непохоже, что при условии независимости в среднем α_i от Z_i инструментальная формула идентифицируют ATT, ATE или вообще что-либо полезное. Подстановка $Y_i = \alpha_i + \beta_i D_i$, и использование независимости в среднем α_i дает

$$\begin{aligned}\frac{\mathbb{C}(Z_i, Y_i)}{\mathbb{C}(Z_i, D_i)} &= \frac{\mathbb{E}[\alpha_i|Z_i = 1] - \mathbb{E}[\alpha_i|Z_i = 0] + \mathbb{E}[\beta_i D_i|Z_i = 1] - \mathbb{E}[\beta_i D_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} \\ &= \frac{\mathbb{E}[\beta_i D_i|Z_i = 1] - \mathbb{E}[\beta_i D_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}.\end{aligned}$$

Проблема заключается в коррелированности β_i и D_i, что в общем случае не позволяет их разделить. Тем не менее, существуют два особых случая, в которых кое-что важное идентифицируемо: случай случайной склонности к воздействию и случай локальных средних эффектов воздействия.

Случайная склонность к воздействию

Стандартен случай в медицинской практике, когда людям назначают лечение случайным образом, но не все его принимают. Тогда параметр Z_i отвечает за назначение, то есть $Z_i = 1$, если индивиду i назначено лечение, и $Z_i = 0$ в противном случае. В такой постановке только те проходят лечение (то есть для кого $D_i = 1$), кому это лечение было случайным образом приписано. Оказывается, что в этом случае инструменты позволяют получить ATT. Этот результат благодаря Имбенсу и Рубину привело к широкому использованию инструментов в биостатистике.

Прежде чем показать, как инструменты позволяют получить ATT, отметим, что все объекты, у которых $Z_i = 0$, не будут подвержены воздействию, то есть из того, что $D_i = 0$ следует, что $Z_i = 0$. Тогда

$$\frac{\mathbb{C}(Z_i, Y_i)}{\mathbb{C}(Z_i, D_i)} = \frac{\mathbb{E}[\beta_i D_i|Z_i = 1] - 0}{\mathbb{E}[D_i|Z_i = 1] - 0} = \frac{\mathbb{E}[\beta_i D_i|Z_i = 1]}{\mathbb{E}[D_i|Z_i = 1]}.$$

Также заметим, что из из условия $D_i = 1$ следует условие $Z_i = 1$, то есть $\{D_i = 1\} \subset \{Z_i = 1\}$. Следовательно, $\mathbb{E}[\beta_i|D_i = 1, Z_i = 1] = \mathbb{E}[\beta_i|D_i = 1] = ATT$. С помощью рассуждений, аналогичных приведенным выше, получаем

$$D_i \mathbb{E}[\beta_i|D_i, Z_i = 1] = D_i \mathbb{E}[\beta_i|D_i = 1, Z_i = 1] = D_i \cdot ATT.$$

Используем закон повторных математических ожиданий:

$$\mathbb{E}[\beta_i D_i|Z_i = 1] = \mathbb{E}[D_i \mathbb{E}[\beta_i|D_i, Z_i = 1]|Z_i = 1] = ATT \cdot \mathbb{E}[D_i|Z_i = 1].$$

В итоге, разделив на $\mathbb{E}[D_i|Z_i = 1]$, получаем

$$\frac{\mathbb{C}(Z_i, Y_i)}{\mathbb{C}(Z_i, D_i)} = \frac{\mathbb{E}[\beta_i D_i|Z_i = 1]}{\mathbb{E}[D_i|Z_i = 1]} = \frac{ATT \cdot \mathbb{E}[D_i|Z_i = 1]}{\mathbb{E}[D_i|Z_i = 1]} = ATT.$$

Локальный средний эффект воздействия

Во втором интересном случае, когда эффект воздействия можно идентифицировать с помощью инструментальных переменных, появляются условия независимости и монотонности. Рассмотрим следующие условия:

Независимость: $D_i = \Pi(Z_i, V_i)$, и вектор (β_i, V_i) независим от Z_i.

Монотонность: $\Pi(1, V_i) \geq \Pi(0, V_i)$ и $\mathbb{P}\{\Pi(1, V_i) > \Pi(0, V_i)\} > 0$.

Условие независимости означает, что существует приведенная форма $\Pi(z, v)$ с возмущением V_i, которое может быть вектором и может входить в модель нелинейно. В качестве примера можно привести модель пересечения порога, в которой $D_i = \mathbb{I}\{Z_i + V_i > 0\}$. Условие монотонности означает, что изменение инструмента влечет изменение воздействия в одном и том же направлении. Это условие выполняется в модели пересечения порога. Приведенную форму этой модели иногда называют уравнением выбора, согласно которому объект подвергают воздействию, когда $\Pi(z, v) = 1$.

Оказывается, что в этих условиях инструменты позволяют найти среднее значение β_i по некоторому подмножеству популяции, называемое локальным средним эффектом воздействия (local average treatment effect). Этот эффект можно определить как

$$LATE = \mathbb{E}[\beta_i | \Pi(1, V_i) > \Pi(0, V_i)].$$

LATE – это средний эффект воздействия среди тех объектов, чье поведение было бы другим, если бы изменились инструменты. Этот параметр может представлять интерес, например, в модели, где Y_i – логарифм заработной платы, D_i – индикатор окончания средней школы, Z_i – фиктивная переменная, соответствующая кварталу рождения, а LATE – средний эффект от образования среди всех отчисленных из школы, которые бы остались в школе, если бы имели другой квартал рождения, и среди всех оставшихся в школе, которые были бы отчислены, если бы имели другой квартал рождения. Таким образом, инструменты оценивают среднюю полезность от получения образования для потенциально отчисленных. LATE – интересный параметр, хотя он и не отражает доходность образования для всей популяции.

Покажем, что инструменты позволяют получить LATE при условиях независимости и монотонности. Пусть $T_i = \Pi(1, V_i) - \Pi(0, V_i)$. Тогда

$$\begin{aligned}
\mathbb{E}[\beta_i D_i | Z_i = 1] - \mathbb{E}[\beta_i D_i | Z_i = 0] &= \mathbb{E}[\beta_i \Pi(1, V_i) | Z_i = 1] - \mathbb{E}[\beta_i \Pi(1, V_i) | Z_i = 0] \\
&= \mathbb{E}[\beta_i \Pi(1, V_i)] - \mathbb{E}[\beta_i \Pi(1, V_i)] = \mathbb{E}[\beta_i T_i].
\end{aligned}$$

Похожим образом можно получить, что

$$\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0] = \mathbb{E}[T_i].$$

По условию монотонности, T_i – фиктивная переменная, принимающая значения 0 или 1. Следовательно,

$$\frac{\mathbb{C}(Z_i, Y_i)}{\mathbb{C}(Z_i, D_i)} = \frac{\mathbb{E}[\beta_i T_i]}{\mathbb{E}[T_i]} = \mathbb{E}[\beta_i | T_i = 1] = \mathbb{E}[\beta_i | \Pi(1, V_i) > \Pi(0, V_i)].$$

Эмпирический пример применения LATE

Приведем эмпирический пример применения LATE из работы Ангриста и Кругера (Angrist & Krueger, 1991), в которой была получена оценка доходности образования при использовании кварталов рождения в качестве инструментов. Данные взяты из переписи населения США 1980-го года, как и в работе Дональда и Ньюи (Donald & Newey, 2001). Двухшаговый метод наименьших квадратов с тремя инструментами дает результат 0,1077 со стандартной ошибкой 0,0195, а оценивание по Фуллеру со 180 инструментами дает результат 0,1063 со стандартной ошибкой (скорректированной на множественность инструментов) 0,0143. Таким образом, получаем, что доходность обучения для «потенциально отчисленных» составляет около 11 процентов.

Отбор по наблюдаемым характеристикам

Для идентификации эффекта воздействия используется другой тип модели, в которой анализ условно на наблюдаемых (или идентифицируемых) переменных X_i заставляет эффект воздействия вести себя как если бы он задавался случайно. Аналогичный подход применялся для удаления эндогенности в линейном уравнении добавлением регрессоров. Условные переменные похожи на пропущенные регрессоры, при включении которых в регрессионное уравнение эндогенность исчезает. Делается специальное предположение:

$$\mathbb{E}[Y_{i0}|X_i, D_i] = \mathbb{E}[Y_{i0}|X_i],$$

то есть Y_{i0} независимо в среднем от D_i условно на X_i. Такое предположение похоже на используемое ранее предположение $\mathbb{E}[Y_{i0}|D_i] = \mathbb{E}[Y_{i0}]$ и является его «условной» версией.

Проблема здесь в том, что неясно, откуда взять переменные X_i. Существует несколько экономических моделей, в которых такие переменные заложены в самой модели. Однако во многих практических случаях эти переменные выбираются без ссылок на модель. Вопрос идентификации тонкий, и важно подобрать правильные X_i. Условная независимость в среднем, которая должна выполняться для X_i, может не выполняться ни для подмножества X_i, ни когда к X_i добавлены дополнительные переменные.

Такое предположение позволяет идентифицировать ATT при наличии еще одного дополнительного условия. Пусть $\mathcal{X}$ обозначает носитель X_i (наименьшее замкнутое множество, имеющее единичную вероятность), а $\mathcal{X}_0$ и $\mathcal{X}_1$ – носители X_i условно на $D_i = 0$ и $D_i = 1$ соответственно. Тогда дополнительное условие – это условие общего носителя

$$\mathcal{X} = \mathcal{X}_0 = \mathcal{X}_1.$$

Такое предположение является необходимым и достаточным условием для существования $\mathbb{E}[Y_i|X_i, D_i = 1]$ и $\mathbb{E}[Y_i|X_i, D_i = 0]$ для всех X_i. Всегда можно проверить, выполняется ли оно на практике или нет.

Из условия общего носителя и условия условной независимости в среднем получаем:

$$\begin{aligned}\mathbb{E}[Y_i|X_i, D_i = 1] - \mathbb{E}[Y_i|X_i, D_i = 0] &= \mathbb{E}[\alpha_i|X_i, D_i = 1] - \mathbb{E}[\alpha_i|X_i, D_i = 0] \\ &\quad + \mathbb{E}[\beta_i|X_i, D_i = 1] = \mathbb{E}[\beta_i|X_i, D_i = 1].\end{aligned}$$

Здесь $\mathbb{E}[\beta_i|X_i, D_i = 1]$ – это условная версия ATT. С помощью закона повторных математических ожиданий ATT определяется как математическое ожидание этой разницы по X_i условно на $D_i = 1$, то есть

$$ATT = \mathbb{E}[\beta_i|D_i = 1] = \mathbb{E}[(\mathbb{E}[Y_i|X_i, D_i = 1] - \mathbb{E}[Y_i|X_i, D_i = 0])|D_i = 1].$$

ATE также можно получить, если мы предположим, что Y_{i1} условно независимо в среднем от D_i условно на X_i. В этом случае

$$\begin{aligned}\mathbb{E}[\beta_i|X_i, D_i = 1] &= \mathbb{E}[Y_{i1}|X_i, D_i = 1] - \mathbb{E}[Y_{i0}|X_i, D_i = 1] \\ &= \mathbb{E}[Y_{i1}|X_i] - \mathbb{E}[Y_{i0}|X_i] = \mathbb{E}[\beta_i|X_i].\end{aligned}$$

Следовательно,

$$ATE = \mathbb{E}[\beta_i] = \mathbb{E}[\mathbb{E}[Y_i|X_i, D_i = 1] - \mathbb{E}[Y_i|X_i, D_i = 0]].$$

В отличие от безусловного случая, ATE является отличной от ATT функцией от распределения данных. ATE получается путем усреднения $\mathbb{E}[Y_i|X_i, D_i = 1] - \mathbb{E}[Y_i|X_i, D_i = 0]$ по всем X_i, в то время как ATT получается с помощью усреднения только при $D_i = 1$.

Оценивание ATT и ATE при этих условных ограничениях – сложная задача. Заметим, что эти параметры зависят от условных ожиданий. Обычно не предполагается, что условные ожидания имеют какую-то определенную функциональную форму. Следовательно, необходимо использовать непараметрическое регрессионное оценивание.

Непапараметрическое оценивание проводить сложно, когда у X_i большая размерность. Обычно такую ситуацию называют «проклятием размерности». Были попытки ослабить проклятие размерности с помощью «меры склонности» – функции $P(X)$, заданной как вероятность оказаться подверженным воздействию (или оказаться «выбранным») условно на X, то есть

$$P(X_i) = \mathbb{P}\{D_i = 1|X_i\} = \mathbb{E}[D_i|X_i].$$

Оказывается, что условная независимость в среднем Y_{i0} условно на X_i влечет условную независимость в среднем условно на $P(X_i)$. Таким образом, если бы $P(X_i)$ было известно, то представлялось бы возможным найти и оценить ATT и ATE, используя одномерную условную переменную, а не многомерную величину X_i. А именно, если $\mathbb{E}[Y_{i0}|X_i, D_i] = \mathbb{E}[Y_{i0}|X_i]$ и $0 < P(X_i) < 1$ с вероятностью единица, то $\mathbb{E}[Y_{i0}|P(X_i), D_i] = \mathbb{E}[Y_{i0}|P(X_i)]$, так что рассуждения, приведенные выше, позволяют получить:

$$ATT = \mathbb{E}[(\mathbb{E}[Y_i|P(X_i), D_i = 1] - \mathbb{E}[Y_i|P(X_i), D_i = 0])|D_i = 1].$$

Если, кроме того, $\mathbb{E}[Y_{i0}|X_i, D_i] = \mathbb{E}[Y_{i0}|X_i]$, то

$$ATE = \mathbb{E}[\mathbb{E}[Y_i|P(X_i), D_i = 1] - \mathbb{E}[Y_i|P(X_i), D_i = 0]].$$

Таким образом, ATT и ATE есть математические ожидания от непараметрической функции от двух переменных, $P(X_i)$ и D_i.

Если $P(X_i)$ полностью неизвестна и неспецифицирована, то никакой пользы от анализа условно на «мере склонности» нет, поскольку $P(X_i)$ тоже является функцией многомерного аргумента. Таким образом, положительный эффект от использования «меры склонности» возникает тогда, когда о $P(X)$ известно больше, чем о $\mathbb{E}[Y_i|X_i, D_i]$.

Остается доказать, что из независимости условно на X следует независимость условно на $P(X)$. Обозначим ради простоты $P_i = P(X_i)$. Мы получим результат для произвольной переменной W_i, и его можно будет применить и к Y_{i0}, и к Y_{i1}. Для того чтобы доказать, что из $\mathbb{E}[W_i|X_i, D_i] = \mathbb{E}[W_i|X_i]$ следует $\mathbb{E}[W_i|P_i, D_i] = \mathbb{E}[W_i|P_i]$, заметим, что по закону повторных математических ожиданий

$$\begin{aligned}
\mathbb{E}[W_i|P_i, D_i = 1] &= \mathbb{E}[\mathbb{E}[W_i|X_i, D_i = 1]|P_i, D_i = 1] = \mathbb{E}[\mathbb{E}[W_i|X_i]|P_i, D_i = 1] \\
&= \frac{\mathbb{E}[D_i\mathbb{E}[W_i|X_i]|P_i]}{\mathbb{E}[D_i|P_i]} = \frac{\mathbb{E}[P_i\mathbb{E}[W_i|X_i]|P_i]}{P_i} \\
&= \mathbb{E}[\mathbb{E}[W_i|X_i]|P_i] = \mathbb{E}[W_i|P_i].
\end{aligned}$$

Точно таким же способом можно получить, что $\mathbb{E}[W_i|P_i, D_i = 0] = \mathbb{E}[W_i|P_i]$, так что нужное нам утверждение следует из предыдущего равенства.

Отсутствие непрерывности в регрессии

Существуют два таких случая: в первом случае переменная воздействия меняется скачками, а во втором случае функция вероятности эффекта воздействия имеет разрывы.

Предположим, $D_i = \mathbb{I}\{X_i \geq c\}$. В этом случае $\mathbb{E}[Y_{i0}|D_i, X_i] = \mathbb{E}[Y_{i0}|X_i]$ и $\mathbb{E}[Y_{i1}|D_i, X_i] = \mathbb{E}[Y_{i1}|X_i]$ по построению. Условие общего носителя не выполняется, поскольку $\mathcal{X}_0$ и $\mathcal{X}_1$ не пересекаются.

Применим другой подход для идентификации параметров и положимся только на условие непрерывности для $\mathbb{E}[Y_{i0}|X_i = x]$ и $\mathbb{E}[Y_{i1}|X_i = x]$.

Предположение: $\mathbb{E}[Y_{i0}|X_i = x]$ и $\mathbb{E}[Y_{i1}|X_i = x]$ непрерывны по x в точке c.

Заметим, что $Y_i = Y_{i0}$ при $X_i < c$ и $Y_i = Y_{i1}$ при $X_i \geq c$. Тогда

$$\mathbb{E}[Y_{i0}|X_i = c] = \lim_{x \uparrow c} \mathbb{E}[Y_{i0}|X_i = x] = \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x],$$

$$\mathbb{E}[Y_{i1}|X_i = c] = \lim_{x \downarrow c} \mathbb{E}[Y_{i1}|X_i = x] = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x].$$

Отсюда следует, что

$$\mathbb{E}[\beta_i|X_i = c] = \mathbb{E}[Y_{i1} - Y_{i0}|X_i = c] = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x].$$

Таким образом, условный эффект воздействия $\mathbb{E}[Y_{i1} - Y_{i0}|X_i = c]$ определяется как скачок $\mathbb{E}[Y_i|X_i = x]$ в точке $x = c$.

Можно интерпретировать этот эффект по-другому. Аналогично манипуляциям выше,

$$\mathbb{E}[\beta_i|X_i = c] = \mathbb{E}[Y_i|D_i = 1, X_i = c] - \mathbb{E}[Y_i|D_i = 0, X_i = c].$$

Заметим, что $\mathbb{E}[Y_i|D_i = 1, X_i = c]$ и $\mathbb{E}[Y_i|D_i = 0, X_i = c]$ представляют собой непараметрические регрессионные функции, оцененные на границах их носителей: первая – на нижней границе, а вторая – на верхней. Значит, в данной ситуации нельзя применять стандартную ядерную регрессию, но можно применить локальную линейную регрессию.

Литература

Angrist, J.D. & A.B. Krueger (1991). Does compulsory school attendance affect schooling and earning? *Quarterly Journal of Economics* 106, 979–1014.

Donald, S. & W. Newey (2001). Choosing the number of instruments. *Econometrica* 69, 1161–1191.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11, 284–300.

Wright, P.G. (1928). *The Tariff on Animal and Vegetable Oils*. New York: MacMillan.

Treatment effects

Whitney K. Newey

Massachusetts Institute of Technology, Cambridge, USA

This essay discusses the issues of identification and estimation of the average treatment effect and the average effect of treatment on the treated.

Оценивание методом «разность разностей»*

Джеффри М. Вулдридж†

Университет штата Мичиган, Ист Лэнсинг, США

Настоящее эссе представляет собой обзор оценивания методом «разность разностей», в котором сначала излагается базовая методология, затем более детально обсуждаются последние достижения в области инференции, и в заключение рассматриваются новые методы оценивания эффектов воздействия в различных нелинейных и полупараметрических моделях.

1 Введение и обзор базовой методологии

Начиная с работы Ashenfelter & Card (1985), оценивание методом «разность разностей» при анализе экономической политики получило широкое распространение. В простейшей постановке наблюдаются некоторые исходы для двух групп и двух временных периодов. Одна из групп подвержена воздействию, или участвует в некоторой программе, во втором периоде, но не в первом. Вторая группа не подвержена воздействию ни в одном из периодов. В случае, когда одни и те же объекты внутри групп наблюдаются в каждом периоде, среднее изменение исхода во второй (контрольной) группе вычитается из среднего изменения исхода в первой (опытной) группе. Это устраняет смещение при сравнении исходов в опытной и контрольной группах только во втором периоде, которое может быть следствием постоянных различий между этими группами, а также смещение при сравнении во времени, которое может быть вызвано временными трендами, никак не связанными с программой.

При наличии повторяющихся выборок за два периода времени модель для типичного представителя каждой из четырех групп можно записать следующим образом:

$$y = \beta_0 + \beta_1 dB + \delta_0 d2 + \delta_1 d2 \cdot dB + u, \tag{1}$$

где y – представляющий интерес исход, $d2$ – фиктивная переменная для второго периода, а dB – фиктивная переменная для опытной группы. Фиктивная переменная dB улавливает возможные различия между опытной и контрольной группами до осуществления программы. Фиктивная переменная для второго периода $d2$ улавливает факторы, которые бы вызвали изменения в y даже при отсутствии воздействия или программы. Представляющий интерес коэффициент δ_1 находится при переменной взаимодействия $d2 \cdot dB$, которая совпадает с фиктивной переменной, равной единице для наблюдений в опытной группе во втором периоде. Оценка δ_1 методом «разность разностей» (РР-оценка) – это обычная МНК-оценка для уравнения (1) на основе случайных выборок по четырем группам. Ее можно записать в виде

$$\hat{\delta}_1 = (\bar{y}_{B,2} - \bar{y}_{B,1}) - (\bar{y}_{A,2} - \bar{y}_{A,1}),$$

где A обозначает контрольную группу. Инференция даже при умеренных размерах выборки для каждой из четырех групп очень проста, и ее легко сделать робастной к гетероскедастичности по группам или временным периодам в рамках модели регрессии.

*Перевод Б. Гершмана. Эссе является предварительной версией материала, подготовленного для *Курса прикладной эконометрики* Гвидо В. Имбенса и Джеффри М. Вулдриджа, издательство Кембриджского университета. Цитировать как: Вулдридж, Джеффри М. (2009) «Оценивание методом „разность разностей“», Квантиль, №6, стр. 25–47. Citation: Wooldridge, Jeffrey M. (2009) "Difference-in-differences estimation," Quantile, No.6, pp. 25–47.

†Адрес: Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA. Электронная почта: *wooldri1@msu.edu*

В некоторых случаях более убедительный анализ эффекта программы доступен при тщательном выборе опытной и контрольной групп. Предположим, например, что штат осуществляет программу в области здравоохранения для пожилых людей, скажем, в возрасте 65 лет и старше, и зависимая переменная y – это некий показатель здоровья. Одна из возможностей состоит в том, чтобы использовать данные только по жителям штата, в котором реализуется программа, как до, так и после ее внедрения, и в качестве контрольной группы взять жителей в возрасте до 65 лет (или, например, в возрасте от 55 до 64 лет), а в качестве опытной группы – жителей в возрасте 65 лет и старше. Потенциальная проблема такого РР-анализа состоит в том, что другие факторы, не связанные с новой программой штата, могут повлиять на уровень здоровья пожилых людей по сравнению с молодыми, например, изменения политики в области здравоохранения на федеральном уровне. Иная стратегия РР-анализа заключается в использовании другого штата для формирования контрольной группы, то есть в рассмотрении пожилых жителей штата, в котором программа отсутствует, в качестве контрольной группы. В данном случае проблема в том, что *изменения* в уровне здоровья пожилых людей могут систематически различаться по штатам, скажем, из-за различий в уровнях дохода и богатства, не связанных с реализацией программы.

Более устойчивого анализа по сравнению с обеими РР-стратегиями, описанными выше, можно добиться, сравнивая РР-оценку для штата, где была реализована программа, с аналогичной оценкой для контрольного штата. Если обозначить два временных периода за 1 и 2, штат, в котором проводится программа, за B, а группу пожилого населения за E, то расширенная версия уравнения (1) примет вид

$$y = \beta_0 + \beta_1 dB + \beta_2 dE + \beta_3 dB \cdot dE + \delta_0 d2 + \delta_1 d2 \cdot dB + \delta_2 d2 \cdot dE + \delta_3 d2 \cdot dB \cdot dE + u. \quad (2)$$

Теперь интерес представляет коэффициент δ_3 при переменной тройного взаимодействия $d2 \cdot dB \cdot dE$. МНК-оценку $\hat{\delta}_3$ можно записать в виде

$$\hat{\delta}_3 = [(\bar{y}_{B,E,2} - \bar{y}_{B,E,1}) - (\bar{y}_{B,N,2} - \bar{y}_{B,N,1})] - [(\bar{y}_{A,E,2} - \bar{y}_{A,E,1}) - (\bar{y}_{A,N,2} - \bar{y}_{A,N,1})], \quad (3)$$

где индекс A снова относится к штату, где программа отсутствует, а индекс N – к группе молодых жителей. Оценку (3) обычно называют оценкой методом «*разность в разностях разностей*» (РРР-оценкой). Первый член в квадратных скобках – это РР-оценка, полученная только для штата B при использовании молодого населения в качестве контрольной группы и данных для обоих периодов. Для уверенности в том, что РР-оценка не просто улавливает различие трендов в уровнях здоровья между пожилыми и молодыми, из РР-оценки вычитается аналогичная оцененная разность трендов для контрольного штата (второй член в квадратных скобках). Если систематические различия в трендах между пожилым и молодым населением в штате A отсутствуют, РРР-оценка будет близкой к РР-оценке, полученной только по данным для штата B.

При реализации оценивания в рамках модели регрессии легко получить стандартную ошибку для $\hat{\delta}_3$, в том числе робастную к гетероскедастичности. Как и в случае РР-оценки, можно легко добавить в уравнение (2) дополнительные регрессоры, а также проводить инференцию, робастную к произвольной форме гетероскедастичности.

Только что изложенный стандартный подход предполагает, что вся вариация при инференции происходит из-за выборочных ошибок при оценивании средних для каждой комбинации группы и периода. Этот подход имеет долгую историю в статистике, поскольку он эквивалентен дисперсионному анализу (ANOVA). Коротко говоря, сначала в соответствии с временными периодами и статусом подверженности воздействию определяются различные подгруппы. Затем для каждой подгруппы берутся случайные выборки и подсчитываются соответствующие разности выборочных средних.

Недавно были предложены различные подходы, основанные на других источниках вариации, нежели выборочная ошибка, возможно в дополнение к выборочной ошибке при оценивании средних. Bertrand, Duflo & Mullainathan (2004) (BDM), Donald & Lang (2007), Hansen

(2007a,b) и Abadie, Diamond & Hainmueller (2007) рассматривают дополнительные источники вариации. На самом деле, в большинстве случаев предполагается, что эти дополнительные источники поглощают выборочную ошибку при оценивании средних по группам и временным периодам. Один из взглядов на дополнительные источники вариации, как в Abadie, Diamond & Hainmueller (2007), состоит в учете неопределенности, связанной с качеством контрольных групп. Обоснованность такого подхода зависит от конкретного приложения. Например, если удается естественным образом определить контрольную и опытную группы и получить случайные выборки, сложно объяснить, почему неприменима модель дисперсионного анализа. С другой стороны, если имеется определенная опытная группа, но отсутствует естественная контрольная группа, имеет смысл учитывать неопределенность, связанную с выбором контрольной группы. Такой подход предложен в Abadie, Diamond & Hainmueller (2007) и кратко рассматривается далее в разделе 6.

Прежде чем перейти к общей постановке, полезно понять, является ли необходимым или желательным включение в РР-анализ чего-то иного, помимо выборочной ошибки. В разделе 2 описан подход, недавно предложенный в Donald & Lang (2007) (DL). Как мы увидим, подход DL не позволяет проводить инференцию в базовом случае сравнения средних для двух групп, даже при случайном распределении воздействия. В частности, хотя DL-оценка – это обычная разность средних из дисперсионного анализа, инференция недоступна. Кроме того, подход DL неприменим к стандартным ситуациям РР- и РРР-оценивания, изложенным выше. Можно ли утверждать, что в этих случаях нельзя построить доверительные интервалы? Во многих ситуациях подобный вывод кажется крайностью.

Возьмем пример из Meyer, Viscusi & Durbin (1995) (MVD), которые оценивают эффект размера выплат по безработице на продолжительность времени, проводимого на пособии. MVD располагают данными за два периода времени – до и после повышения потолка покрываемых доходов. Опытная группа включает работников с высоким доходом, а контрольная – работников с низким доходом, то есть тех, кто не должен был быть затронут изменением потолка. По данным для штата Кентукки при совокупном размере выборки 5626 наблюдений MVD получают РР-оценку эффекта данного вмешательства, равную 19,2%-му повышению продолжительности времени, проводимого на пособии. t-статистика равна 2,76 и, как и сама оценка, мало меняется при добавлении контрольных переменных. MVD также изучают данные для штата Мичиган. Применяя тот же РР-подход, они находят почти идентичный эффект – 19,1%. Тем не менее, при «всего лишь» 1524 наблюдениях t-статистика равна 1,22. Похоже, в этом примере имеет место большая выборочная ошибка при подсчете средних, и нельзя получить точную оценку при отсутствии достаточно больших выборок. Неясно, какую выгоду дает заключение о том, что, поскольку эффект вмешательства точно идентифицируем, инференция в таких случаях недоступна.

2 Подход DL и метод минимального расстояния при малом числе групп

Начнем этот раздел с изложения подхода Donald & Lang (2007) (DL). Постановка DL применима к общему случаю РР-анализа, когда число групп (контрольных и опытных) достаточно мало, и для каждой из них доступны довольно большие выборки. Это такая же постановка, как и при традиционном подходе, рассмотренном в разделе 1, но предполагаемая структура модели отличается, что ведет к иному подходу к инференции.

Для иллюстрации вопросов, поднятых в DL, рассмотрим случай с единственным регрессором, меняющимся только по группам:

$$\begin{aligned} y_{gm} &= \alpha + \beta x_g + c_g + u_{gm} \qquad (4) \\ &= \delta_g + \beta x_g + u_{gm},\ m = 1, ..., M_g;\ g = 1, ..., G. \qquad (5) \end{aligned}$$

В простейшем случае x_g – единственный индикатор программы. Ключевой характеристикой уравнения (4) является наличие переменной c_g, эффекта группы или кластера. Иначе модель можно записать в виде (5) с общим коэффициентом наклона β, но зависящей от группы константой δ_g. DL рассматривают модель в виде (4), предполагая, что c_g не зависит от x_g и имеет нулевое среднее. Они используют данную формулировку для выявления проблем, связанных с применением стандартной инференции к уравнению (4) при рассмотрении c_g как части составной ошибки $v_{gm} = c_g + u_{gm}$. Известно, что это неудачная мысль даже в случае большого G и малых M_g, поскольку стандартный подход не учитывает корреляцию ошибок внутри каждой группы. Более того, как следует из обсуждения результатов в Hansen (2007a), описанных в разделе 3, при малом G даже инференция, устойчивая к кластеризации, часто дает плохие результаты.

Один из способов увидеть проблему при применении стандартной инференции – заметить, что при $M_g = M$ для всех $g = 1, ..., G$ обычная МНК-оценка $\hat{\beta}$ совпадает с межгрупповой оценкой, полученной из регрессии

$$\bar{y}_g \text{ на } 1, x_g,\ g = 1, ..., G.$$

Условно на x_g, $\hat{\beta}$ наследует свое распределение от $\{\bar{v}_g : g = 1, ..., G\}$, внутригрупповых средних составных ошибок $v_{gm} \equiv c_g + u_{gm}$. Наличие c_g означает, что новые наблюдения внутри группы не дают дополнительной информации для оценивания β, помимо их эффекта на групповое среднее $\bar{y}_g$. По сути, имеются только G полезных носителей информации.

Если добавить несколько сильных предположений, у данной проблемы существует решение, но только при $G > 2$. Вдобавок к предпосылке о том, что $M_g = M$ для всех g, предположим, что $c_g|x_g \sim N(0, \sigma_c^2)$ и $u_{gm}|x_g, c_g \sim N(0, \sigma_u^2)$. Тогда $\bar{v}_g$ не зависит от x_g и $\bar{v}_g \sim N(0, \sigma_c^2 + \sigma_u^2/M)$ для всех g. Поскольку предполагается независимость по g, уравнение

$$\bar{y}_g = \alpha + \beta x_g + \bar{v}_g,\ g = 1, ..., G, \tag{6}$$

удовлетворяет предположениям классической линейной модели. Следовательно, можно проводить инференцию на основе t_{G-2}-распределения для тестирования гипотез о β при условии, что $G > 2$. Если G очень мало, условия на значимость t-статистики при использовании t_{G-2}-распределения гораздо более строгие, чем для $t_{M_1+M_2+...+M_G-2}$-распределения, которое применялось бы в случае простой объединенной регрессии, как при стандартном РР-анализе.

Если x_g – $1 \times K$ вектор, чтобы использовать t_{G-K-1}-распределение для инференции, необходимо условие $G > K + 1$. Оно допускает случай множественности временных периодов и опытных групп, если предположить независимость выборок по всем группам и временным периодам.

Как отмечают DL, осуществление правильной инференции при наличии c_g – это *не* просто вопрос корректировки стандартных ошибок, получаемых при применении обычного МНК, на корреляцию внутри кластеров, которая все равно не работает в случае малого G. При одинаковом размере групп имеется только одна оценка: объединенная регрессия, модель со случайными эффектами и межгрупповая регрессия (6) – все они дают *одну и ту же* оценку $\hat{\beta}$. Регрессия в (6) при использовании t_{G-K-1}-распределения позволяет проводить инференцию с надлежащим размером теста.

Метод DL можно использовать и без предположения о нормальности u_{gm}, если общий размер групп M большой: согласно центральной предельной теореме, $\bar{u}_g$ будет асимптотически нормально распределена при весьма общих предпосылках. Тогда, поскольку c_g нормально распределена, можно считать $\bar{v}_g$ асимптотически нормальной с постоянной дисперсией. Более того, даже если размеры групп меняются, для очень больших размеров групп $\bar{u}_g$ будет пренебрежимо малой частью $\bar{v}_g$, поскольку $\mathbb{V}(\bar{v}_g) = \sigma_c^2 + \sigma_u^2/M_g$. Если c_g нормально распределена и доминирует $\bar{v}_g$, анализ классической линейной модели (6) будет приблизительно верным.

Модель DL находит наиболее широкую сферу применения, когда среднее значение ошибок, $\bar{u}_g$, можно игнорировать, то есть либо σ_u^2 мала по сравнению с σ_c^2, либо M_g велики, либо и то и другое. На самом деле, применение подхода DL с различными размерами групп или негауссовыми u_{gm} равносильно игнорированию ошибки оценивания выборочных средних $\bar{y}_g$. Иными словами, это равносильно тому, как если бы простая модель регрессии $\mu_g = \alpha+\beta x_g+c_g$ анализировалась при классических предположениях линейной модели (где $\bar{y}_g$ используется вместо неизвестного внутригруппового среднего μ_g). Такой подход сильно отличается от традиционного, когда c_g отбрасывается и для построения доверительных интервалов и проведения инференции используется выборочная вариация в $\bar{y}_g$. Подход DL кажется разумным, когда при подсчете $\bar{y}_g$ выборку составляет вся популяция для группы g, а $\bar{y}_g$ рассматривается как зависимая, объясняемая переменная.

Если $1 \times L$ вектор z_{gm}, меняющийся внутри группы, присутствует в исходной модели, как это имеет место во многих РР-приложениях, можно использовать усредненное уравнение

$$\bar{y}_g = \alpha + x_g\beta + \bar{z}_g\gamma + \bar{v}_g,\ g = 1, ..., G, \tag{7}$$

при условии, что $G > K + L + 1$. Если c_g не зависит от $(x_g, \bar{z}_g)$, имеет гомоскедастичное нормальное распределение, и размеры групп велики, инференция может быть реализована на основе $t_{G-K-L-1}$-распределения. Это довольно стандартный способ проверки устойчивости результатов, полученных по дезагрегированным данным, но часто он осуществляется для несколько больших значений G (скажем, $G = 50$). Когда некоторые регрессоры меняются внутри кластера, использовать усредненные данные, вообще говоря, неэффективно. Но использование средних означает, что стандартные ошибки не нужно корректировать на внутригрупповую корреляцию. Как обсуждается в следующем разделе, если G достаточно большое и размеры выборок не слишком велики, робастная к кластеризации инференция может быть приемлемой.

Для малого G и больших M_g инференция, реализуемая при рассмотрении (6) или (7) как классической линейной модели, будет очень консервативной в отсутствие кластеризации. Возможно, в некоторых случаях желательно учитывать этот источник вариации, но он отсекает некоторые широко применяемые методы анализа эффекта программ. Предположим, например, что имеются две популяции (например, мужчины и женщины, два разных города, или опытная и контрольная группы) со средними μ_g, $g = 1, 2$, и требуется получить доверительный интервал для их разности. Почти во всех случаях имеет смысл рассматривать данные как две случайные выборки, по одной для каждой подгруппы популяции. При случайной выборке для каждой группы и предположении о нормальности и равенстве дисперсий в популяциях, обычная тест-статистика на равенство средних имеет точное $t_{M_1+M_2-2}$-распределение при нулевой гипотезе о равных популяционных средних. Иначе говоря, можно построить точный 95%-ый доверительный интервал для разности популяционных средних. Даже при умеренных размерах M_1 и M_2, $t_{M_1+M_2-2}$-распределение близко к стандартному нормальному. Кроме того, можно ослабить предположение о нормальности для проведения асимптотической инференции, а t-статистику легко модифицировать для случая различных популяционных дисперсий. При контролируемом эксперименте стандартный анализ разности средних часто весьма убедителен. Тем не менее, изучать оценку в постановке DL невозможно, поскольку $G = 2$. Проблема очевидна из (5): по сути, имеется три параметра – δ_1, δ_2 и β, – но только два наблюдения.

DL критикуют Card & Krueger (1994) за сравнение средних изменений в заработной плате работников заведений быстрого питания по двум штатам, поскольку Card & Krueger не учитывают эффект штата (Нью-Джерси и Пеннсильвании) c_g в составной ошибке v_{gm}. Но критика DL в случае $G = 2$ выглядит немного странной. В постановке DL разность средних оценивает величину

$$\mu_2 - \mu_1 = (\delta_2 + \beta) - \delta_1 = (\alpha + c_2 + \beta) - (\alpha + c_1) = \beta + (c_2 - c_1).$$

При предположениях DL $c_2 - c_1$ имеет нулевое среднее, так что оценивание этой величины не должно привести к смещению при анализе эффекта программы. DL предполагают, что β – интересующий параметр, но, если эксперимент основан на рандомизации (что также справедливо в DL, так как предполагается независимость c_g и x_g), включение c_g в оцениваемый эффект безболезненно. Короче говоря, в постановке DL анализ в Card & Krueger не является систематически смещенным; DL просто утверждают, что инференция невозможна. Но если естественный эксперимент, лежащий в основе PP-анализа, надежен, то групповые эффекты просто должны быть составной частью оцениваемых средних.

Кроме того, не очевидно, следует ли использовать подход DL даже при достаточном большом числе степеней свободы. Допустим, например, что есть $G = 4$ группы, две из которых контрольные ($x_1 = x_2 = 0$) и две – опытные ($x_3 = x_4 = 1$). Подход DL предполагает подсчет средних для каждой группы, $\bar{y}_g$, и оценивание регрессии $\bar{y}_g$ на $1, x_g$, $g = 1, ..., 4$. Инференция основана на t_2-распределении. Оценку $\hat{\beta}$ в данном случае можно записать в виде

$$\hat{\beta} = \frac{\bar{y}_3 + \bar{y}_4}{2} - \frac{\bar{y}_1 + \bar{y}_2}{2}. \tag{8}$$

Объединенная регрессия для дезагрегированных данных дает взвешенную среднюю $(p_3\bar{y}_3 + p_4\bar{y}_4) - (p_1\bar{y}_1 + p_2\bar{y}_2)$, где $p_1 = M_1/(M_1 + M_2)$, $p_2 = M_2/(M_1 + M_2)$, $p_3 = M_3/(M_3 + M_4)$, и $p_4 = M_4/(M_3 + M_4)$ – относительные доли в контрольных и опытных группах, соответственно. Когда $\hat{\beta}$ имеет вид (8), неясно, зачем использовать t_2-распределение для инференции. Каждое $\bar{y}_g$ обычно подсчитывается по большой выборке – $M_g = 30$ или около того обычно достаточно для асимптотической нормальности стандартизированного среднего, – а значит, при должной стандартизации $\hat{\beta}$ имеет асимптотически стандартное нормальное распределение при весьма общих условиях.

В данном приложении подход DL отвергает обычную инференцию на основе внутригрупповых средних, подсчитанных по большим выборкам, поскольку могут не выполняться условия $\mu_1 = \mu_2$ и $\mu_3 = \mu_4$. Другими словами, контрольная группа может быть разнородной, как и опытная группа. Но этот факт сам по себе не делает неверной стандартную инференцию для модели (8). На самом деле, если *определить* объект интереса как

$$\tau = \frac{\mu_3 + \mu_4}{2} - \frac{\mu_1 + \mu_2}{2}, \tag{9}$$

то есть своего рода средний эффект воздействия, то оценка $\hat{\beta}$ состоятельна для τ и (при должной нормировке) асимптотически нормальна по мере роста M_g.

Уравнение (9) намекает на иной способ взглянуть на случай малого G и больших M_g. В этом конкретном приложении два параметра, α и β, оцениваются на основе четырех моментов, которые можно оценить по данным. МНК-оценки из (6) в этом случае являются оценками минимального расстояния (МР) – см. Wooldridge (2002, Глава 14) – при ограничениях $\mu_1 = \mu_2 = \alpha$ и $\mu_3 = \mu_4 = \alpha + \beta$. Если взять единичную матрицу 4×4 в качестве взвешивающей, можно получить $\hat{\beta}$ как в (8) и $\hat{\alpha} = (\bar{y}_1 + \bar{y}_2)/2$. При применении МР-подхода ясно, что имеются два сверхидентифицирующих ограничения, которые легко тестировать. Но даже их отвержение всего лишь означает, что по крайней мере одна пара средних среди всех контрольных и опытных групп различается.

При больших размерах групп даже при небольших G можно сформулировать задачу в рамках модели минимального расстояния, как это делается, например, в Loeb & Bound (1996), у которых имелось $G = 36$ возрастных групп и множество наблюдений для каждой группы. Для каждой группы g запишем

$$y_{gm} = \delta_g + z_{gm}\gamma_g + u_{gm},\ m = 1, ..., M_g, \tag{10}$$

предполагая случайность выборок внутри групп и независимость выборок по группам. Предположим также справедливость всех стандартных условий для состоятельности (при $M_g \to$

∞) и $\sqrt{M_g}$-асимптотической нормальности МНК-оценки; см, например, Wooldridge (2002, Глава 4). Наличие переменных x_g группового уровня в «структурной» модели можно рассматривать как наложение ограничений на константы δ_g в отдельных моделях для групп в (10). В частности,

$$\delta_g = \alpha + x_g\beta,\ g = 1, ..., G, \tag{11}$$

где x_g – фиксированные наблюдаемые характеристики разнородных групп. При K характеристиках для определения α и β необходимо условие $G \geq K+1$. Если M_g достаточно велики для точного оценивания δ_g, простая двухшаговая стратегия оценивания напрашивается сама собой. Сначала надо оценить $\hat{\delta}_g$ и $\hat{\gamma}_g$ из МНК-регрессий для каждой группы. Если $G = K+1$, то обычно можно однозначно выразить $\hat{\theta} \equiv (\hat{\alpha}, \hat{\beta}')'$ в терминах $G \times 1$ вектора $\hat{\delta}$: $\hat{\theta} = X^{-1}\hat{\delta}$, где X – $(K+1) \times (K+1)$ матрица с g-й строкой $(1, x_g)$. Если $G > K+1$, то на втором шаге можно использовать подход минимального расстояния, изложенный в Wooldridge (2002, Раздел 14.6). Если в качестве взвешивающей матрицы I_G взять единичную матрицу $G \times G$, то оценку минимального расстояния можно подсчитать из МНК-регрессии

$$\hat{\delta}_g \text{ на } 1, x_g,\ g = 1, ..., G. \tag{12}$$

При асимптотике с $M_g = \rho_g M$, где $0 < \rho_g \leq 1$ и $M \to \infty$, оценка минимального расстояния $\hat{\theta}$ состоятельна и $\sqrt{M}$-асимптотически нормальна. Тем не менее, эта конкретная оценка минимального расстояния является асимптотически неэффективной, если не делать сильных предположений. Поскольку выборки предполагаются независимыми, не значительно сложнее получить эффективную МР-оценку, также известную как оценка «минимального хи-квадрат».

Рассмотрим сначала случай, когда z_{gm} отсутствует на первом шаге оценивания, то есть $\hat{\delta}_g$ – это просто $\bar{y}_g$, выборочное среднее для группы g. Пусть $\hat{\sigma}_g^2$ обозначает обычную выборочную дисперсию для группы g. Поскольку $\bar{y}_g$ независимы по g, эффективная МР-оценка использует диагональную взвешивающую матрицу. Ради легкости вычисления, оценку минимального хи-квадрат можно получить, используя взвешенный МНК (ВМНК) в (12), где в качестве весов для группы g берется $M_g/\hat{\sigma}_g^2$ (группы с большим количеством данных и меньшей дисперсией получают больший вес). Что удобно, стандартные t-статистики из ВМНК-регрессии имеют асимптотическое стандартное нормальное распределение при больших размерах групп M_g. При фиксированном G, ВМНК – просто вычислительный метод; стандартная асимптотика для ВМНК-оценки предполагает $G \to \infty$. Подход минимального расстояния работает при малом G, если $G \geq K+1$ и каждое M_g достаточно велико, чтобы нормальность была хорошим приближением распределения (надлежащим образом нормированных) выборочных средних для каждой группы. В стандартной РР-постановке из раздела 1, МР-подход сводится к взвешиванию наименьших квадратов, где в качестве весов используются величины, обратные к оцененным внутригрупповым дисперсиям.

Если z_{gm} присутствует на первом шаге оценивания, в качестве весов при оценивании методом минимального хи-квадрат берутся величины, обратные асимптотическим дисперсиям для g оценок констант в G отдельных регрессиях. При больших M_g можно сделать их полностью робастными к гетероскедастичности в $\mathbb{E}(u_{gm}^2|z_{gm})$, используя оценку дисперсии в «сэндвичной» форме из White (1980). Даже при предположении о гомоскедастичности внутри групп хотелось бы как минимум разрешить разные σ_g^2. При наличии $\widehat{Avar}(\hat{\delta}_g)$, которые являются просто квадратами сообщаемых стандартных ошибок для $\hat{\delta}_g$, можно использовать веса $1/\widehat{Avar}(\hat{\delta}_g)$ в вычислительно простой ВМНК-процедуре. По-прежнему для получения диагональной взвешивающей матрицы при МР-оценивании используется предположение о независимости по g.

Важным побочным продуктом ВМНК-регрессии является статистика минимального хи-квадрат, которую можно использовать для тестирования $G-K-1$ сверхидентифицирующих ограничений. Эту статистику легко подсчитать как взвешенную сумму квадратов остатков, скажем, SSR_w. При нулевой гипотезе в (11), $SSR_w \overset{a}{\sim} \chi^2_{G-K-1}$ по мере роста размеров групп M_g. Если H_0 отвергается при достаточно малом уровне значимости, x_g не хватает для характеризации изменений констант между группами. Если H_0 не отвергается, появляется некоторая уверенность в правильности спецификации, и можно проводить инференцию, используя стандартное нормальное распределение для t-статистик при тестировании гипотез о линейных комбинациях популяционных средних.

МР-подход также применим, если предположить, что $\gamma_g = \gamma$ для всех g. Получить сами $\hat{\delta}_g$ легко. Сначала следует оценить объединенную регрессию

$$y_{gm} \text{ на } d1_g, d2_g, ..., dG_g, z_{gm},\ m = 1, ..., M_g;\ g = 1, ..., G, \tag{13}$$

где $d1_g, d2_g, ..., dG_g$ – групповые фиктивные переменные. Использование оценки $\hat{\delta}_g$ из объединенной регрессии (13) в МР-оценивании осложняется тем фактом, что $\hat{\delta}_g$ более не являются асимптотически независимыми; на самом деле, $\hat{\delta}_g = \bar{y}_g - \bar{z}_g\hat{\gamma}$, где $\hat{\gamma}$ – вектор общих регрессионных коэффициентов, и присутствие $\hat{\gamma}$ порождает корреляцию оценок констант. Пусть $\hat{V}$ – $G\times G$ оцененная (асимптотическая) дисперсионная матрица $G\times 1$ вектора $\hat{\delta}$. Тогда МР-оценка имеет вид $\hat{\theta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\hat{\delta}$, а ее асимптотическая дисперсия равна $(X'\hat{V}^{-1}X)^{-1}$. Если оценивается МНК-регрессия (12) или ее ВМНК-версия, обычные стандартные ошибки будут неправильными, поскольку они не учитывают межгрупповую корреляцию между оценками.

Если сверхидентифицирующие ограничения отвергаются, по сути это означает, что $\delta_g = \alpha + x_g\beta + c_g$, где c_g можно интерпретировать как отклонение от ограничений в (11) для группы g. По мере того как G растет относительно K, вероятность отвержения ограничений растет. Одна из возможностей состоит в применении подхода DL, то есть в анализе МНК-регрессии (12) в контексте классической линейной модели (КЛМ), где инференция основана на t_{G-K-1}-распределении. Почему КЛМ-анализ оправдан в данном случае? Поскольку $\hat{\delta}_g = \delta_g + O_p(M_g^{-1/2})$, можно игнорировать ошибку оценивания в $\hat{\delta}_g$ для больших M_g (то же предположение о «больших» M_g лежит в основе МР-подхода.) Это равносильно оцениванию уравнения $\delta_g = \alpha + x_g\beta + c_g,\ g = 1, ..., G$, с помощью МНК. Если c_g имеют нормальное распределение, классический анализ применим, так как предполагается независимость c_g и x_g. Этот подход желателен, когда невозможно или не хочется искать регрессоры на уровне групп, которые полностью идентифицируют δ_g. Он основан на предположении о том, что прочие факторы в c_g систематически не связаны с x_g, что разумно, если, скажем, x_g – случайно назначаемая на групповом уровне программа, как в случае, рассмотренном в Angrist & Lavy (2002).

3 РР-анализ при большом числе групп

Обратимся теперь к случаю РР-оценивания, когда число групп «велико». Введем также временное измерение в явном виде и положим, что интервал времени достаточно долгий. Такая структура является хорошим описанием многих РР-моделей, когда группы представлены географическими регионами и данные по этим регионам доступны для большого числа временных периодов.

При большом числе периодов и групп полезна общая модель, рассмотренная в BDM (2004) и Hansen (2007a). На индивидуальном уровне уравнение имеет вид

$$y_{igt} = \lambda_t + \alpha_g + \mathbf{x}_{gt}\boldsymbol{\beta} + \mathbf{z}_{igt}\boldsymbol{\gamma}_{gt} + v_{gt} + u_{igt},\ i = 1, ..., M_{gt}, \tag{14}$$

где i индексирует объект, g – группу, а t – время. В этой модели имеется полный набор временных эффектов λ_t, групповых эффектов α_g, регрессоров на уровне группы/периода

x_{gt} (это переменные воздействия), регрессоров индивидуального уровня $\mathbf{z}_{igt}$, ненаблюдаемых эффектов на уровне группы/периода v_{gt}, и ошибок на индивидуальном уровне u_{igt}. Интерес представляет оценивание $\boldsymbol{\beta}$.

Полезно записать (14) с константами, меняющимися по группам и времени:

$$y_{igt} = \delta_{gt} + \mathbf{z}_{igt}\boldsymbol{\gamma}_{gt} + u_{igt},\ i = 1, ..., M_{gt}, \tag{15}$$

что дает модель на индивидуальном уровне, тогда как константы и регрессионные коэффициенты могут различаться в зависимости от пары (g, t). В этом случае

$$\delta_{gt} = \lambda_t + \alpha_g + \mathbf{x}_{gt}\boldsymbol{\beta} + v_{gt}. \tag{16}$$

Можно воспринимать (16) как регрессионную модель на уровне группы/временного периода.

Как обсуждается в BDM, обычный метод оценивания и инференции для (14) – игнорировать v_{gt}, то есть считать независимыми наблюдения на индивидуальном уровне. Когда v_{gt} присутствует, подобная инференция может оказаться весьма обманчивой. BDM и Hansen (2007a) допускают серийную корреляцию для $\{v_{gt} : t = 1, 2, ..., T\}$ и предполагают независимость по группам g.

Далее будем считать для простоты, что (16) представляет конечный интерес. Регрессоры $\mathbf{x}_{gt}$ наблюдаются, λ_t учитываются с помощью годовых фиктивных переменных, а α_g – с помощью групповых фиктивных переменных. Тогда проблема в том, что не наблюдаются δ_{gt}. Но для оценивания δ_{gt} можно использовать данные на индивидуальном уровне, если размеры выборок на уровне группы/периодов M_{gt} достаточно велики. При случайных выборках внутри каждой пары (g, t) естественная оценка δ_{gt} получается из МНК-оценивания (15) для каждой пары (g, t) при предположении, что $\mathbb{E}(\mathbf{z}'_{igt}u_{igt}) = \mathbf{0}$. В большинстве РР-приложений это предположение выполняется почти по определению, поскольку для улучшения оценки δ_{gt} включаются контрольные переменные на индивидуальном уровне. Если напрашивается конкретная форма гетероскедастичности и предполагается, что $\mathbb{E}(u_{it}|\mathbf{z}_{igt}) = 0$, можно использовать ВМНК-процедуру. Иногда хочется предположить однородность регрессионных коэффициентов – скажем, $\boldsymbol{\gamma}_{gt} = \boldsymbol{\gamma}_g$ или даже $\boldsymbol{\gamma}_{gt} = \boldsymbol{\gamma}$, – и для наложения подобных ограничений данные объединяются. В любом случае, предположим, что M_{gt} достаточно велики, чтобы можно было игнорировать ошибку оценивания в $\hat{\delta}_{gt}$; будем считать, что вариация в уравнении (16) исходит от v_{gt}. Hansen (2007a) рассматривает корректировку инференции, которая учитывает выборочную ошибку в $\hat{\delta}_{gt}$, но методы становятся более сложными. Альтернативой является применение метода минимального расстояния, как в разделе 2. МР-подход по сути устраняет v_{gt} из уравнения (16) и рассматривает $\delta_{gt} = \lambda_t + \alpha_g + \mathbf{x}_{gt}\boldsymbol{\beta}$ как множество ограничений, накладываемых на δ_{gt}. Инференция на основе МР-оценки использует только выборочную вариацию в оценках $\hat{\delta}_{gt}$, которые независимы по всем парам (g, t), если оцениваются по отдельности, или коррелированы, если данные объединяются.

Поскольку ошибка оценивания в $\hat{\delta}_{gt}$ игнорируется, рассмотрим просто уравнение для панельных данных

$$\hat{\delta}_{gt} = \lambda_t + \alpha_g + \mathbf{x}_{gt}\boldsymbol{\beta} + v_{gt},\ t = 1, ..., T,\ g = 1, ..., G, \tag{17}$$

где ошибка сохраняет вид v_{gt}, поскольку $\hat{\delta}_{gt}$ и δ_{gt} считаются взаимозаменяемыми. При таком предположении можно напрямую применить результаты из BDM и Hansen (2007a) к этому уравнению. А именно, если оценить (17) с помощью МНК, что подразумевает полные эффекты времени и группы наряду с x_{gt}, эта оценка имеет удовлетворительные свойства по мере роста G и T, если $\{v_{gt} : t = 1, 2, ..., T\}$ – слабо зависимый (с перемешиванием) временной ряд для всех g. Симуляции в BDM и Hansen (2007a) показывают, что инференция, робастная к кластеризации, где каждый кластер – это множество временных периодов, работает достаточно хорошо, если $\{v_{gt}\}$ следует стационарному $\mathsf{AR}(1)$-процессу и G умеренно велико.

Hansen (2007b), замечая, что МНК-оценка (оценка с фиксированными эффектами) для (17) неэффективна, если v_{gt} серийно некоррелированы (и, возможно, гетероскедастичны), предлагает доступную ОМНК-оценку. Как хорошо известно, если T недостаточно велико, оценка параметров дисперсионной матрицы $\Omega_g = \mathbb{V}(\mathbf{v}_g)$, где $\mathbf{v}_g$ – $T \times 1$ вектор ошибок для каждого g, затруднительна после удаления групповых эффектов. Иными словами, применение остатков из модели с фиксированными эффектами $\hat{v}_{gt}$ для оценки Ω_g может давать серьезное смещение при малых T. Solon (1984) обратил внимание на эту проблему для гомоскедастичной $\mathsf{AR}(1)$ модели. Конечно, смещение исчезает, когда $T \to \infty$, и программные пакеты, такие как STATA, имеющие встроенную команду для оценивания модели с фиксированными эффектами и $\mathsf{AR}(1)$-ошибками, используют обычную оценку $\mathsf{AR}(1)$-коэффициента $\hat{\rho}$, полученную из регрессии

$$\hat{v}_{gt} \text{ на } \hat{v}_{g,t-1},\ t = 2, ..., T,\ g = 1, ..., G.$$

Как обсуждается в Wooldridge (2003) и Hansen (2007b), один из способов учета смещения в $\hat{\rho}$ – использовать полностью робастную оценку дисперсионной матрицы. Но симуляции Хансена показывают, что этот подход весьма неэффективен по сравнению с его предложением, заключающемся в корректировке смещения оценки $\hat{\rho}$ и ее последующем использовании при доступном ОМНК-оценивании (Хансен работает с общей $\mathsf{AR}(p)$ моделью). Хансен доказывает множество привлекательных теоретических свойств своей оценки. Итеративная процедура с корректировкой смещения имеет то же самое асимптотическое распределение, что и $\hat{\rho}$, в случае, когда оценка $\hat{\rho}$ должна хорошо работать: при G и T, стремящимся к бесконечности. Что наиболее важно при применении к РР-задачам, доступная ОМНК-оценка на основе итеративной процедуры имеет то же самое асимптотическое распределение, что и ОМНК-оценка при $G \to \infty$ и фиксированном T. Когда G и T велики, нет необходимости в итерировании для достижения эффективности.

Далее Хансен показывает, что даже при больших G и T, когда нескорректированные AR-коэффициенты дают асимптотическую эффективность, скорректированные на смещение оценки дают улучшения с точки зрения асимптотики более высоких порядков. Одно из ограничений результатов Хансена состоит в предположении о строгой экзогенности регрессоров $\{\mathbf{x}_{gt} : t = 1, ..., T\}$. Известно, что при использовании МНК, то есть обыкновенной оценки с фиксированными эффектами, строгая экзогенность не требуется для состоятельности при $T \to \infty$. ОМНК, используя корреляцию по разным временным периодам, склонен усугублять смещение, которое исходит от отсутствия строгой экзогенности. При анализе эффектов программ это повод для беспокойства, если программа может вводиться и отменяться с течением времени, поскольку тогда необходимо понять, связано ли решение о реализации или приостановке программы с ее результатами в прошлом.

При большом G и малом T можно оценить дисперсионную матрицу Ω_g без ограничений и далее применить ОМНК – этот подход впервые был предложен в Kiefer (1980) и недавно изучен в Hausman & Kuersteiner (2005). Он эквивалентен отбрасыванию эффекта периода в уравнении с вычтенным средним по времени и реализации полного ОМНК (это также позволяет избежать вырожденности дисперсионной матрицы ошибок с вычтенным средним по времени). Hausman & Kuersteiner показывают, что подход Кифера работает достаточно хорошо, когда $G = 50$ и $T = 10$, хотя при $G = 50$ и $T = 20$ наблюдаются существенные искажения размера.

В частности, в случае, когда M_{gt} не особенно велики, игнорирование ошибки оценивания в $\hat{\delta}_{gt}$ может вызывать беспокойство. Один из простых способов избежать этого – агрегировать уравнение (14) по объектам, что дает

$$\bar{y}_{gt} = \lambda_t + \alpha_g + \mathbf{x}_{gt}\boldsymbol{\beta} + \bar{\mathbf{z}}_{gt}\boldsymbol{\gamma} + v_{gt} + \bar{u}_{gt},\ t = 1, .., T,\ g = 1, ..., G.$$

Конечно, это уравнение можно оценить как модель с фиксированными эффектами, и полностью робастная инференция доступна по методу из Hansen (2007a), поскольку составная ошибка $\{r_{gt} \equiv v_{gt} + \bar{u}_{gt}\}$ слабо зависима. При большом G и малом T можно использовать ОМНК-оценку с фиксированными эффектами, используя дисперсионную матрицу без ограничений. Трудности с применением конкретной модели временного ряда для ошибки возникают из-за наличия $\bar{u}_{gt}$. При различных M_{gt} $\mathbb{V}(\bar{u}_{gt})$ почти наверняка гетероскедастична (конечно, она может быть таковой и при равных M_{gt}). Так что даже если специфицировать, скажем, AR(1)-модель $v_{gt} = \rho v_{g,t-1} + e_{gt}$, дисперсионная матрица для $\mathbf{r}_g$ становится более сложной. Один из возможных путей – просто предположить, что составная ошибка r_{gt} следует простой модели, реализовать методы Хансена, но затем применять полностью робастную инференцию.

Подход из Donald & Land (2007) применим в данной постановке, если использовать анализ для конечных выборок для объединенной регрессии (17). Тем не менее, DL предполагают, что ошибки $\{v_{gt}\}$ некоррелированы во времени, так что, несмотря на то что для небольших G и T в t-распределении мало степеней свободы, их подход не учитывает вариацию из-за серийной корреляции $\{v_{gt} : t = 1, ..., T\}$.

4 Панельные данные на индивидуальном уровне

Панельные данные на индивидуальном уровне позволяют применять мощные методы для оценки эффектов программ. В простейшем случае имеются два временных периода и бинарный индикатор программы, w_{it}, равный единице, если объект i участвует в программе в момент t. Простая эффективная модель имеет вид

$$y_{it} = \alpha + \eta d2_t + \tau w_{it} + c_i + u_{it},\ t = 1, 2, \tag{18}$$

где $d2_t = 1$ при $t = 2$ и нуль в противном случае, c_i – наблюдаемый эффект, а u_{it} – случайные ошибки. Коэффициент τ – эффект воздействия. Простая процедура оценивания состоит во взятии первых разностей для удаления c_i:

$$(y_{i2} - y_{i1}) = \eta + \tau(w_{i2} - w_{i1}) + (u_{i2} - u_{i1}),$$

или

$$\Delta y_i = \eta + \tau \Delta w_i + \Delta u_i. \tag{19}$$

Если $\mathbb{E}(\Delta w_i \Delta u_i) = 0$, то есть изменения в статусе участия в программе некоррелированы с изменениями случайных ошибок, МНК-оценки уравнения (19) состоятельны. В самом распространенном случае $w_{i1} = 0$ для всех i, то есть никто не участвует в программе в начальный период времени. Тогда МНК-оценка имеет вид

$$\hat{\tau} = \Delta \bar{y}_{treat} - \Delta \bar{y}_{control} \tag{20}$$

и представляет собой РР-оценку с отличием лишь в том, что берутся разности средних по времени для тех же объектов. Эту же самую оценку можно получить, не вводя разнородность, а просто записав уравнение для y_{it} с полным набором эффектов групп и времени. Кроме того, (20) дает иные оценки, нежели регрессия y_{i2} на $1, y_{i1}$, w_{i2}, то есть когда y_{i1} используется в качестве контрольной переменной в кросс-секционной регрессии. Оценки могут быть близкими, но их состоятельность основана на разных предположениях.

Полезно доказать несостоятельность оценки при включении y_{i1} в качестве регрессора при стандартных предположениях в модели с ненаблюдаемыми эффектами, когда $w_{i1} \equiv 0$. В этом случае можно записать:

$$\begin{aligned} y_{i1} &= \alpha + c_i + u_{i1}, \\ y_{i2} &= (\alpha + \eta) + \tau w_{i2} + c_i + u_{i2}. \end{aligned}$$

Стандартными предположениями – см. Wooldridge (2002, Глава 10) – являются

$$\mathbb{E}(u_{it}|c_i, w_{i2}) = 0, \ t = 1, 2, \tag{21}$$

$$\mathbb{E}(u_{i2}u_{i1}|w_{i2}, c_i) = 0, \tag{22}$$

то есть участие в программе (во втором периоде) строго экзогенно условно на c_i и ошибки условно серийно некоррелированы. Пусть $\tilde{\tau}$ – МНК-оценка для регрессии y_{i2} на $1, y_{i1}$, w_{i2}, $i = 1, ..., N$. Тогда

$$\tilde{\tau} = \left(\sum_{i=1}^{N} \hat{r}_{i2}^2\right)^{-1} \left(\sum_{i=1}^{N} \hat{r}_{i2}\Delta y_i\right),$$

где $\hat{r}_{i2}$ – остатки регрессии w_{i2} на $1, y_{i1}$, $i = 1, ..., N$. После подстановки $\Delta y_i = \eta + \tau w_{i2} + \Delta u_i$ и простых вычислений получаем

$$\tilde{\tau} = \tau + \left(N^{-1}\sum_{i=1}^{N} \hat{r}_{i2}^2\right)^{-1} \left(N^{-1}\sum_{i=1}^{N} \hat{r}_{i2}\Delta u_i\right).$$

Запишем линейную проекцию w_{i2} на $1, y_{i1}$ в виде $w_{i2} = \pi_0 + \pi_1 y_{i1} + r_{i2}$. Тогда предел по вероятности элемента в числителе равен

$$-\pi_1\mathbb{E}(y_{i1}\Delta u_i) = -\pi_1\mathbb{E}(y_{i1}u_{i2}) + \pi_1\mathbb{E}(y_{i1}u_{i1}) = \pi_1\mathbb{E}(y_{i1}u_{i1}).$$

Далее, $\mathbb{E}(y_{i1}u_{i1}) = \sigma_{u_1}^2$. Таким образом,

$$\operatorname{plim}(\tilde{\tau}) = \tau + \pi_1\frac{\sigma_{u_1}^2}{\sigma_{r_2}^2}.$$

Кроме того,

$$\pi_1 = \frac{\mathbb{C}(y_{i1}, w_{i2})}{\mathbb{V}(y_{i1})} = \frac{\mathbb{C}(c_i, w_{i2})}{\sigma_c^2 + \sigma_{u_1}^2}$$

и имеет тот же знак, что $\mathbb{C}(c_i, w_{i2})$. Чтобы использовать этот результат, предположим, что w_{i2} означает участие в программе профессиональной подготовки, а y_{it} – трудовой доход. Тогда, если участие в программе отрицательно связано с ненаблюдаемыми характеристиками c_i, ведущими к более высокому доходу, $\pi_1 < 0$, и оценка $\tilde{\tau}$ смещена вниз. Таким образом, включение y_{i1} в качестве дополнительного регрессора ведет к недооценке эффекта в среднем. Если более производительные работники самоотбираются в программу, эффект будет переоценен. При предпосылках (21) и (22) простая оценка (20) состоятельна (и даже несмещена).

Конечно, предположение (22) может быть нежелательным, но с двумя периодами неясно, как можно отличить временную корреляцию, вызванную наличием c_i в (18) и корреляцию между u_{i1} и u_{i2}.

В общем случае с большим числом периодов и произвольным порядком действия программы можно использовать модель

$$y_{it} = \lambda_t + \tau w_{it} + \mathbf{x}_{it}\boldsymbol{\gamma} + c_i + u_{it}, \ t = 1, ..., T, \tag{23}$$

которая учитывает агрегированные временные эффекты и включает контрольные переменные $\mathbf{x}_{it}$. Оценивание модели с фиксированными эффектами и взятие первых разностей для удаления c_i – стандартные процедуры при условии, что индикатор программы w_{it} строго экзогенен: корреляция между w_{it} и u_{ir} для любых t и r ведет к несостоятельности обеих оценок, хотя оценка с фиксированными эффектами обычно имеет меньшее смещение, когда можно предположить экзогенность в текущем периоде, $\mathbb{C}(w_{it}, u_{it}) = 0$. Строгая экзогенность может нарушаться, если реализация программы меняется в зависимости от прежних исходов y_{it}. В случае, когда $w_{it} = 1$ при $w_{ir} = 1$ для $r < t$, строгая экзогенность обычно является разумным предположением.

Уравнение (23) допускает зависимость назначения программы от уровня индивидуального эффекта c_i, но, кроме того, w_{it} может коррелировать с индивидуальными временными трендами зависимой переменной. В таком случае имеет место модель с коррелированным случайным трендом:

$$y_{it} = c_i + g_i t + \lambda_t + \tau w_{it} + \mathbf{x}_{it}\boldsymbol{\gamma} + u_{it},\ t = 1, ..., T, \tag{24}$$

где g_i – коэффициент тренда для индивида i. Общий анализ допускает произвольную корреляцию между (c_i, g_i) и w_{it}, что требует $T \geq 3$. После взятия первых разностей получаем

$$\Delta y_{it} = g_i + \eta_t + \tau \Delta w_{it} + \Delta \mathbf{x}_{it}\boldsymbol{\gamma} + \Delta u_{it},\ t = 2, ..., T, \tag{25}$$

где $\eta_t = \lambda_t - \lambda_{t-1}$ – это новый набор временных эффектов. Уравнение (25) можно оценить, снова взяв разности или используя оценку с фиксированными эффектами. Выбор зависит от характера серийной корреляции $\{\Delta u_{it}\}$ (при предположении о строгой экзогенности программы и регрессоров). Если Δu_{it} почти не коррелируют, предпочтительна оценка с фиксированными эффектами. Если исходные ошибки $\{u_{it}\}$ почти не коррелируют, предпочтительно применение оценки с фиксированными эффектами к (24) с целью удаления линейных трендов зависимой переменной, переменной воздействия и регрессоров. Полностью робастная инференция с использованием робастных к кластеризации оценок весьма проста. Конечно, может возникнуть желание разрешить эффекту программы меняться во времени, что легко сделать, добавив переменную взаимодействия между временными фиктивными переменными и индикатором программы.

Используя модель гипотетических исходов из литературы об эффектах воздействия, можно применять стандартные подходы на основе панельных данных,. Для каждой пары (i, t), обозначим за $y_{it}(1)$ и $y_{it}(0)$ гипотетические исходы и предположим, что регрессоры отсутствуют. Один из способов формулировки предположения о несмешанности воздействия заключается в том, что для постоянной во времени разнородности $\mathbf{c}_i$ выполняются условия

$$\begin{aligned} \mathbb{E}[y_{it}(0)|\mathbf{w}_i, \mathbf{c}_i] &= \mathbb{E}[y_{it}(0)|\mathbf{c}_i] && (26)\\ \mathbb{E}[y_{it}(1)|\mathbf{w}_i, \mathbf{c}_i] &= \mathbb{E}[y_{it}(1)|\mathbf{c}_i], && (27)\end{aligned}$$

где $\mathbf{w}_i = (w_{i1}, ..., w_{iT})$ – последовательность всех воздействий. Подобного рода предположение о строгой экзогенности условно на латентных переменных упоминалось ранее. Оно позволяет воздействию быть коррелированным с постоянной во времени разнородностью, но не со случайными изменениями гипотетических исходов в какой-либо период времени. Далее предположим, что ожидаемый эффект воздействия зависит только от времени:

$$\mathbb{E}[y_{it}(1)|\mathbf{c}_i] = \mathbb{E}[y_{it}(0)|\mathbf{c}_i] + \tau_t,\ t = 1, ..., T. \tag{28}$$

Записывая $y_{it} = (1 - w_{it})y_{it}(0) + w_{it}y_{it}(1)$ и используя (26), (27) и (28), получим

$$\mathbb{E}(y_{it}|\mathbf{w}_i, \mathbf{c}_i) = \mathbb{E}[y_{it}(0)|\mathbf{c}_i] + w_{it}\{\mathbb{E}[y_{it}(1)|\mathbf{c}_i] - \mathbb{E}[y_{it}(0)|\mathbf{c}_i]\} = \mathbb{E}[y_{it}(0)|\mathbf{c}_i] + \tau_t w_{it}.$$

Если предположить аддитивную структуру $\mathbb{E}[y_{it}(0)|\mathbf{c}_i]$, а именно

$$\mathbb{E}[y_{it}(0)|\mathbf{c}_i] = \alpha_{t0} + c_{i0},$$

то получим

$$\mathbb{E}(y_{it}|\mathbf{w}_i, \mathbf{c}_i) = \alpha_{t0} + c_{i0} + \tau_t w_{it}, \tag{29}$$

уравнение для оценивания, к которому применимы хорошо известные методы. Поскольку $\{w_{it} : t = 1, ..., T\}$ строго экзогенны условно на c_{i0}, можно применять модель с фиксированными эффектами или брать первые разности при полном наборе временных фиктивных переменных. Стандартный анализ предполагает $\tau_t = \tau$, но, разумеется, можно легко разрешить эффектам программы меняться во времени.

Можно брать математическое ожидание условно еще и на регрессорах $\mathbf{x}_{it}$ и предположить линейность, скажем, $\mathbb{E}[y_{it}(0)|\mathbf{x}_{it}, \mathbf{c}_i] = \alpha_{t0} + \mathbf{x}_{it}\boldsymbol{\gamma}_0 + c_{i0}$. Если (26) принимает вид

$$\mathbb{E}[y_{it}(0)|\mathbf{w}_i, \mathbf{x}_i, \mathbf{c}_i] = \mathbb{E}[y_{it}(0)|\mathbf{x}_{it}, \mathbf{c}_i], \tag{30}$$

и похожую форму имеет (27), то в уравнение для оценивания (29) просто добавляется член $\mathbf{x}_{it}\boldsymbol{\gamma}_0$. Более интересные модели получаются, если разрешить зависимость эффекта воздействия от разнородности. Предположим, вдобавок к предположению об игнорируемости в (30) (и эквивалентному условию для $y_{it}(1)$), что

$$\mathbb{E}[y_{it}(1) - y_{it}(0)|\mathbf{x}_{it}, \mathbf{c}_i] = \tau_t + a_i + (\mathbf{x}_{it} - \boldsymbol{\xi}_t)\boldsymbol{\delta}, \tag{31}$$

где a_i – функция от $\mathbf{c}_i$, нормализованная так, чтобы $\mathbb{E}(a_i) = 0$ и $\boldsymbol{\xi}_t = \mathbb{E}(\mathbf{x}_{it})$. Уравнение (31) позволяет эффекту воздействия зависеть от времени, ненаблюдаемой разнородности и наблюдаемых регрессоров. Тогда

$$\mathbb{E}(y_{it}|\mathbf{w}_i, \mathbf{x}_i, \mathbf{c}_i) = \alpha_{t0} + \tau_t w_{it} + \mathbf{x}_{it}\boldsymbol{\gamma}_0 + w_{it}(\mathbf{x}_{it} - \boldsymbol{\xi}_t)\boldsymbol{\delta} + c_{i0} + a_i w_{it}.$$

Это модель с коррелированными случайными коэффициентами, поскольку коэффициент при w_{it} равен $(\tau_t + a_i)$, с математическим ожиданием τ_t. В общем случае хотелось бы разрешить w_{it} коррелировать с a_i и c_{i0}. При малом T и большом N мы не пытаемся оценить a_i (и c_{i0}). Но расширенное внутри-преобразование устраняет $a_i w_{it}$. Упростим ситуацию, предполагая $\tau_t = \tau$ и опуская все иные регрессоры. Тогда регрессия, которая, казалось бы, страдает от проблемы шумовых параметров, позволяет состоятельно оценить τ: следует регрессировать y_{it} на фиктивные переменные для лет, индивидов и переменные взаимодействия последних с w_{it}. Иными словами, следует оценить уравнение

$$\hat{y}_{it} = \hat{\alpha}_{t0} + \hat{c}_{i0} + \hat{\tau}_i w_{it}.$$

Хотя $\hat{\tau}_i$ обычно являются плохими оценками $\tau_i = \tau + a_i$, их среднее является хорошей оценкой для τ :

$$\hat{\tau} = N^{-1}\sum_{i=1}^{N}\hat{\tau}_i.$$

Стандартную ошибку можно найти, используя метод из Wooldridge (2002, Раздел 11.2) или применяя бутстрап.

Wooldridge (2005) дает простое достаточное условие, при котором обычная оценка с фиксированными эффектами, которая ведет себя как если бы эффекты воздействия были постоянными, состоятельно оценивает средний эффект воздействия. В дополнение к предположению о несмешанности достаточно предположить, что

$$\mathbb{E}(\tau_i|\ddot{w}_{it}) = \mathbb{E}(\tau_i) = \tau,\ t = 1, ..., T, \tag{32}$$

где $\ddot{w}_{it} = w_{it} - \bar{w}_i$. В сущности, эффекты воздействия на индивидуальном уровне могут коррелировать со средней склонностью быть подверженным воздействию $\bar{w}_i$, но не с отклонениями для какого-либо определенного периода времени.

Предположение (32) не является совсем общим, и интересно иметь простой способ определить, является ли эффект воздействия разнородным среди объектов. Для этого можно использовать корреляцию между τ_i и воздействием. Вспоминая, что $\tau_i = \tau + a_i$, полезным предположением (которое не обязательно должно выполняться для получения теста) является следующее:

$$\mathbb{E}(a_i|w_{i1},...,w_{iT}) = \mathbb{E}(a_i|\bar{w}_i) = \rho(\bar{w}_i - \mu_{\bar{w}_i}),$$

где прочие регрессоры отброшены. Тогда можно оценить уравнение (с регрессорами)

$$y_{it} = \alpha_{t0} + \tau w_{it} + \mathbf{x}_{it}\boldsymbol{\gamma}_0 + w_{it}(\mathbf{x}_{it} - \bar{\mathbf{x}}_t)\boldsymbol{\delta} + \rho w_{it}(\bar{w}_i - \bar{w}) + c_{i0} + e_{it}$$

как стандартную модель с фиксированными эффектами. Далее к $\hat{\rho}$ применим обычный t-тест, робастный к гетероскедастичности и серийной корреляции. Если нулевая гипотеза отвергается, это не означает, что обычная оценка с фиксированными эффектами несостоятельна, но такое возможно.

5 Полупараметрический и непараметрический подходы

Вернемся к постановке с двумя группами и двумя периодами. Athey & Imbens (2006) (AI) обобщают стандартную РР-модель в нескольких направлениях. Чтобы соответствовать обозначениям в AI, будем называть два временных периода $t = 0$ и $t = 1$, а две группы – $g = 0$ и $g = 1$. Пусть $Y_i(0)$ означает гипотетический исход при отсутствии воздействия, а $Y_i(1)$ – при его наличии. AI предполагают, что

$$Y_i(0) = h_0(U_i, T_i), \tag{33}$$

где T_i – временной период и

$$h_0(u,t) \text{ строго возрастает по } u \text{ для } t = 0, 1. \tag{34}$$

Случайная величина U_i отражает все ненаблюдаемые характеристики индивида i. Из уравнения (33) следует, что исход для индивида с $U_i = u$ будет одинаковым в данный период независимо от принадлежности к группе. Предположение о строгой монотонности (34) отсекает случай дискретной зависимой переменной, но Athey & Imbens (2006) приводят границы при слабой монотонности и показывают, как при дополнительных предположениях можно получить точечную идентификацию.

Распределение U_i может меняться по группам, но не во времени внутри групп, так что

$$D(U_i|T_i, G_i) = D(U_i|G_i). \tag{35}$$

Из этой предпосылки следует, что внутри группы популяционное распределение стабильно во времени.

Стандартную РР-модель в этих терминах можно записать в виде

$$h_0(u,t) = u + \delta \cdot t$$

и

$$U_i = \alpha + \gamma G_i + V_i,\ V_i \perp (G_i, T_i),$$

хотя в силу линейности можно обойтись предположением $\mathbb{E}(V_i|G_i, T_i) = 0$. Если эффект воздействия постоянен среди индивидов, $\tau = Y_i(1) - Y_i(0)$, то

$$Y_i = \alpha + \beta T_i + \gamma G_i + \tau G_i T_i + V_i, \tag{36}$$

где $Y_i = (1 - G_i T_i) Y_i(0) + G_i T_i Y_i(1)$ – реализация исхода. Так как $\mathbb{E}(V_i|G_i, T_i) = 0$, параметры в (36) можно оценить с помощью МНК.

Athey & Imbens называют расширение обычной РР-модели моделью «*изменения изменений*» (ИИ). Они не только показывают, как получить средний эффект воздействия, но и доказывают, что распределение гипотетического исхода условно на воздействии

$$D(Y_i(0)|G_i = 1, T_i = 1)$$

идентифицируемо. Распределение $D(Y_i(1)|G_i = 1, T_i = 1)$ идентифицируемо по данным, поскольку $Y_i = Y_i(1)$ при $G_i = T_i = 1$. Дополнительное условие, которое используют AI, состоит в том, что носитель распределения $D(U_i|G_i = 1)$ содержится в носителе $D(U_i|G_i = 0)$:

$$\mathbb{U}_1 \subseteq \mathbb{U}_0. \tag{37}$$

Пусть $F^0_{gt}(y)$ обозначает кумулятивную функцию распределения $D(Y_i(0)|G_i = g, T_i = t)$ для $g = 0, 1$ и $t = 0, 1$, а $F_{gt}(y)$ – КФР для наблюдаемого исхода Y_i условно на $G_i = g$ и $T_i = t$. По определению, $F_{gt}(y)$ обычно идентифицируема по данным в предположении случайных выборок для каждой пары (g, t). AI показывают, что при (33), (34), (35) и (37)

$$F^{(0)}_{11}(y) = F_{10}(F^{-1}_{00}(F_{01}(y))), \tag{38}$$

где $F^{-1}_{00}(\cdot)$ – обратная функция для F_{00}, которая существует при предположении о строгой монотонности. Заметим, что все КФР в правой части (38) можно оценить по данным; это просто КФР наблюдаемых исходов условно на разных парах (g, t). Поскольку $F^{(1)}_{11}(y) = F_{11}(y)$, можно оценить распределения обоих гипотетических исходов условно на воздействии, $G_i = T_i = 1$.

Средний эффект воздействия в рамках ИИ-модели имеет вид

$$\tau_{CIC} = \mathbb{E}[Y(1)|G = 1, T = 1] - \mathbb{E}[Y(0)|G = 1, T = 1] = \mathbb{E}(Y_{11}(1)) - \mathbb{E}(Y_{11}(0)),$$

где индекс i отброшен, $Y_{gt}(1)$ – случайная величина, имеющая распределение $D(Y(1)|G = g, t)$, а $Y_{gt}(0)$ – случайная величина с распределением $D(Y(0)|G = g, t)$. При тех же предположениях, что и выше,

$$\tau_{CIC} = \mathbb{E}(Y_{11}) - \mathbb{E}[F^{-1}_{01}(F_{00}(Y_{10}))],$$

где Y_{gt} – случайная величина с распределением $D(Y|G = g, t)$. При наличии случайных выборок для каждой подгруппы, состоятельная оценка τ_{CIC} имеет вид

$$\hat{\tau}_{CIC} = N^{-1}_{11}\sum_{i=1}^{N_{11}} Y_{11,i} - N^{-1}_{10}\sum_{i=1}^{N_{10}} \hat{F}^{-1}_{01}(\hat{F}_{00}(Y_{10,i}))$$

при состоятельных оценках КФР $\hat{F}_{00}$ и $\hat{F}_{01}$ для контрольных групп в начальный и последующий моменты времени, соответственно. Теперь $Y_{11,i}$ обозначает реализацию наблюдаемого исхода для группы $g = 1, t = 1$; аналогично для $Y_{10,i}$. Athey & Imbens (2006) установили слабые условия, при которых $\hat{\tau}_{CIC}$ является $\sqrt{N}$-асимптотически нормальной (естественно, число наблюдений должно расти внутри каждой группы). В случае, когда распределения Y_{10} и Y_{00} совпадают, оценка представляет собой простую разность средних для опытной группы по времени.

Описанный подход применим как в случае повторяющихся кросс-секций, так и в случае панельных данных. Athey & Imbens (2006) обсуждают, какие предположения можно ослабить в случае панельных данных и какие доступны альтернативные стратегии оценивания. В частности, если U_{i0} и U_{i1} представляют ненаблюдаемые характеристики объекта i в начальный и последующий периоды времени, соответственно, то (35) можно привести к виду

$$D(U_{i0}|G_i) = D(U_{i1}|G_i),$$

что позволяет структуру ненаблюдаемых компонент $U_{it} = C_i + V_{it}$, где V_{it} имеет одно и то же распределение в каждом периоде.

Как обсуждают AI, при панельных данных доступны другие подходы к оцениванию. Altonji & Matzkin (2005) используют предположения о взаимозаменяемости для идентификации средних частных эффектов. Для иллюстрации их подхода предположим, что гипотетические исходы удовлетворяют предположению об игнорируемости

$$\mathbb{E}(Y_{it}(g)|W_{i1}, ..., W_{iT}, U_i) = h_{tg}(U_i),\ t = 1, ..., T,\ g = 0, 1. \tag{39}$$

Эффект воздействия для объекта i в период t равен $h_{t1}(U_i) - h_{t0}(U_i)$, а средний эффект воздействия –

$$\tau_t = \mathbb{E}[h_{t1}(U_i) - h_{t0}(U_i)],\ t = 1, ..., T.$$

Предположим, что

$$D(U_i|W_{i1}, ..., W_{iT}) = D(U_i|\bar{W}_i), \tag{40}$$

то есть только интенсивность воздействия коррелирует с разнородностью. При условиях (39) и (40) можно показать, что

$$\mathbb{E}(Y_{it}|W_i) = \mathbb{E}[\mathbb{E}(Y_{it}|W_i, U_i)|W_i] = \mathbb{E}(Y_{it}|W_{it}, \bar{W}_i).$$

Ключевая особенность состоит в том, что $\mathbb{E}(Y_{it}|W_i)$ не зависит от $\{W_{i1}, ..., W_{iT}\}$ произвольным образом; это функция только от $(W_{it}, \bar{W}_i)$. Если W_{it} непрерывны или принимают множество значений, для оценки $\mathbb{E}(y_{it}|W_{it}, \bar{W}_i)$ можно применить методы локального сглаживания. В случае эффектов воздействия оценивание очень простое, поскольку $(W_{it}, \bar{W}_i)$ может принимать только $2T$ значений. Средний эффект воздействия можно оценить как

$$\hat{\tau}_t = N^{-1} \sum_{i=1}^{n} (\hat{\mu}_t^Y(1, \bar{W}_i) - \hat{\mu}_t^Y(0, \bar{W}_i)).$$

Если объединить данные по t (и по i) и оценить линейную регрессию Y_{it} на $1, d2_t, ..., dT_t$, W_{it}, $\bar{W}_i$, $t = 1, ..., T$, $i = 1, ..., N$, получим обычную оценку с фиксированными эффектами $\hat{\tau}_{FE}$ как коэффициент при W_{it}. Wooldridge (2005) описывает другие случаи и сравнивает эту стратегию с другими подходами. С помощью условного ММП для логит-модели можно оценить параметры, но обычно не средние эффекты воздействия, и требуется условная независимость. В пробит-модели Чемберлена с коррелированными случайными коэффициентами разнородность выражается в том, что $U_i|W_i \sim N(\xi_0 + \boldsymbol{\xi}_1 W_{i1} + ... + \boldsymbol{\xi}_T W_{iT}, \eta^2)$, и это позволяет идентифицировать средние эффекты воздействия без предположения о взаимозаменяемости, но сохраняя предположение о распределении (и функциональной форме вероятности для зависимой переменной).

Для основного случая двух периодов, когда воздействие не происходит в начальный период времени ни для одного из объектов, но происходит для некоторых объектов во второй период, Heckman, Ichimura & Todd (1997) (HIT) и Abadie (2005) предлагают методы как для повторяющихся кросс-секций, так и для панельных данных, которые накладывают предположение о несмешанности на изменения во времени. Пусть $Y_t(w)$ обозначает гипотетический исход в момент времени t, $t = 0, 1$, соответствующий статусу подверженности воздействию w, $w = 0, 1$. Поскольку в начальный период воздействие отсутствует, $Y_0(1) = Y_0(0)$, что просто означает, что в начальный период времени гипотетические исходы отсутствуют.

Как в HIT (1997) и Abadie (2005), рассмотрим сначала оценивание

$$\tau_{att} = \mathbb{E}[Y_1(1) - Y_1(0)|W = 1].$$

Поскольку нет объектов, подверженных воздействию в начальный период времени, $W = 1$ означает наличие воздействия во втором периоде. Чтобы оценить τ_{att}, ключевым предположением о несмешанности является

$$\mathbb{E}[Y_1(0) - Y_0(0)|X, W] = \mathbb{E}[Y_1(0) - Y_0(0)|X], \tag{41}$$

так что, условно на X, статус подверженности воздействию не связан с изменениями во времени при отсутствии программы. Вдобавок, сделаем предположение о пересечении:

$$0 < \mathbb{P}\{W = 1|X = x\} < 1 \tag{42}$$

для всех x. Для оценки τ_{att} можно ослабить (42) до $\mathbb{P}\{W = 1|X = x\} < 1$. Используя регрессионный подход, можно оценить τ_{att}, сперва оценив $\mathbb{E}[Y_1(1) - Y_1(0)|X, W = 1]$. Как показано в HIT, это математическое ожидание идентифицируемо при условии (41). Пусть $Y_1 = (1 - W) \cdot Y_1(0) + W \cdot Y_1(1)$ – наблюдаемый исход для $t = 1$, а $Y_0 = Y_0(0) = Y_0(1)$ – в момент $t = 0$. Тогда

$$\begin{aligned}
&\{\mathbb{E}(Y_1|X, W = 1) - \mathbb{E}(Y_1|X, W = 0)\} - \{\mathbb{E}(Y_0|X, W = 1) - \mathbb{E}(Y_0|X, W = 0)\} = \\
&\{\mathbb{E}[Y_1(1)|X, W = 1] - \mathbb{E}[Y_1(0)|X, W = 0]\} - \{\mathbb{E}[Y_0(1)|X, W = 1] - \mathbb{E}[Y_0(0)|X, W = 0]\} = \\
&\{\mathbb{E}[Y_1(1)|X, W = 1] - \mathbb{E}[Y_1(0)|X, W = 1]\} + \{\mathbb{E}[Y_1(0)|X, W = 1] - \mathbb{E}[Y_1(0)|X, W = 0]\} - \\
&- \{\mathbb{E}[Y_0(1)|X, W = 1] - \mathbb{E}[Y_0(0)|X, W = 0]\} = \\
&\{\mathbb{E}[Y_1(1) - Y_1(0)|X, W = 1] - \mathbb{E}[Y_0(1) - Y_0(0)|X, W = 1]\} + \\
&+ \{\mathbb{E}[Y_1(0) - Y_0(0)|X, W = 1] - \mathbb{E}[Y_1(0) - Y_0(0)|X, W = 0]\} = \\
&\mathbb{E}[Y_1(1) - Y_1(0)|X, W = 1] - \mathbb{E}[Y_0(1) - Y_0(0)|X, W = 1],
\end{aligned}$$

где последнее равенство выполняется в силу (41). Но $Y_0(1) = Y_0(0)$, то есть

$$\begin{aligned}
&\{\mathbb{E}(Y_1|X, W = 1) - \mathbb{E}(Y_1|X, W = 0)\} - \{\mathbb{E}(Y_0|X, W = 1) - \mathbb{E}(Y_0|X, W = 0)\} = \\
&\mathbb{E}[Y_1(1) - Y_1(0)|X, W = 1].
\end{aligned} \tag{43}$$

Каждое из четырех математических ожиданий по левую сторону (43) можно оценить, имея случайные выборки для двух периодов. Например, можно использовать гибкие параметрические модели или даже непараметрическое оценивание, чтобы оценить $\mathbb{E}(Y_1|X, W = 1)$ по данным для объектов, подверженных воздействию в момент $t = 1$.

Анализ для

$$\tau_{ate} = \mathbb{E}[Y_1(1) - Y_1(0)]$$

аналогичный. Теперь добавим к (41) дополнительное предположение о несмешиваемости:

$$\mathbb{E}[Y_1(1) - Y_0(1)|X, W] = \mathbb{E}[Y_1(1) - Y_0(1)|X], \tag{44}$$

то есть о том, что статус подверженности воздействию не связан с эффектом воздействия. При (41) и (44)

$$\begin{aligned}
&\{\mathbb{E}(Y_1|X, W = 1) - \mathbb{E}(Y_1|X, W = 0)\} - \{\mathbb{E}(Y_0|X, W = 1) - \mathbb{E}(Y_0|X, W = 0)\} = \\
&\mathbb{E}[Y_1(1) - Y_1(0)|X],
\end{aligned} \tag{45}$$

так что теперь можно оценить средний эффект воздействия условно на X, используя оценки условных средних для четырех групп в соответствии с индикаторами периода и статуса подверженности воздействию. Если требуется τ_{ate}, можно просто усреднить разности подогнанных значений по регрессорам (вспомним, что имеются разные случайные выборки в каждый период времени). Оценка принимает вид

$$\hat{\tau}_{ate,reg} = N_1^{-1} \sum_{i=1}^{N_1} (\hat{\mu}_{11}(X_i) - \hat{\mu}_{10}(X_i)) - N_0^{-1} \sum_{i=1}^{N_0} (\hat{\mu}_{01}(X_i) - \hat{\mu}_{00}(X_i)), \tag{46}$$

где $\hat{\mu}_{tw}(x)$ – оцененная регрессионная функция для периода времени t и статуса подверженности воздействию w, N_1 – общее число наблюдений для $t = 1$, а N_0 – общее число наблюдений для нулевого периода. Базовый РР-подход с регрессорами просто предполагает линейность и постоянство коэффициентов по t и статусу подверженности воздействию. Конечно, эти предположения можно легко ослабить в контексте обычной модели линейной регрессии. Строго говоря, (45) ведет к τ_{ate} (после усреднения относительно распределения X) только если распределение регрессоров не меняется во времени. Конечно, одна из причин включения регрессоров – учет возможных изменений структуры рассматриваемых популяций во времени. Обычный РР-подход избегает этой проблемы, предполагая, что эффект воздействия не зависит от регрессоров. Уравнение (46) позволяет эффектам воздействия различаться в зависимости от X, но два средних значения обязательно подсчитываются для различных временных периодов.

Abadie (2005) показывает, как с помощью взвешивания на основе оцениваемой вероятности воздействия можно получить τ_{att} в случае повторяющихся кросс-секций при условии стационарности. Оценку можно записать способом, похожим на (46). В частности,

$$\hat{\tau}_{att,ps} = N_1^{-1} \sum_{i=1}^{N_1} \frac{(W_i - \hat{p}(X_i))Y_{i1}}{\hat{\rho}(1 - \hat{p}(X_i))} - N_0^{-1} \sum_{i=1}^{N_0} \frac{(W_i - \hat{p}(X_i))Y_{i0}}{\hat{\rho}(1 - \hat{p}(X_i))}, \tag{47}$$

где $\{Y_{i1} : i = 1, ..., N_1\}$ – данные для $t = 1$, а $\{Y_{i0} : i = 1, ..., N_0\}$ – для $t = 0$. Уравнение (47) имеет очевидную интерпретацию. Первое среднее – стандартная взвешенная оценка на основе оцениваемой вероятности воздействия, если использовать только $t = 1$ и предположить несмешиваемость в уровнях; см., например, Dehejia & Wahba (1999) и Wooldridge (2002, Глава 18). Второе среднее – та же оценка, но на данных для $t = 0$ (конечно, эта оценка сама по себе несет мало смысла, поскольку в период $t = 0$ нет индивидов, подверженных воздействию). Как и в (46), оцененный эффект воздействия – это разность двух оценок в разные периоды времени.

При наличии панельных данных можно брать разности для тех же объектов во времени. Например, (45) можно записать в форме

$$\mathbb{E}(\Delta Y|X, W = 1) - \mathbb{E}(\Delta Y|X, W = 0) = \mathbb{E}[Y_1(1) - Y_1(0)|X],$$

где $\Delta Y = Y_1 - Y_0$ – изменение в наблюдаемом исходе для типичного представителя популяции. Но это просто означает, что можно применить регрессионную корректировку или методы на основе оцениваемой вероятности воздействия к изменениям в Y. Регрессионная корректировка требует оценивания $\mu_1^{\Delta}(X) \equiv \mathbb{E}(\Delta Y|X, W = 1)$ и $\mu_0^{\Delta}(X) \equiv \mathbb{E}(\Delta Y|X, W = 0)$ с использованием опытной и контрольной групп, соответственно, а затем τ_{ate} как

$$N^{-1} \sum_{i=1}^{N} (\hat{\mu}_1^{\Delta}(X_i) - \hat{\mu}_0^{\Delta}(X_i)),$$

где N – общее число индивидов в панельных данных. Abadie (2005) показывает, что

$$\tau_{att} = (\mathbb{P}\{W = 1\})^{-1} \mathbb{E}\left[\frac{(W - p(X))\Delta Y}{1 - p(X)}\right], \tag{48}$$

где $p(X) = \mathbb{P}\{W = 1|X\}$ – вероятность воздействия. Величины в (48) наблюдаемы или, в случае $p(X)$ и $\rho = \mathbb{P}\{W = 1\}$, оцениваемы. Как в Hirano, Imbens & Ridder (2003), для оценки $p(X)$ можно использовать гибкую логит-модель; в качестве $\hat{\rho}$ используется доля объектов, подверженных воздействию. Тогда

$$\hat{\tau}_{att,ps} = N^{-1} \sum_{i=1}^{N} \frac{(W_i - \hat{p}(X_i))\Delta Y_i}{\hat{\rho}(1 - \hat{p}(X_i))}$$

состоятельна и $\sqrt{N}$-асимптотически нормальна. HIR обсуждают оценивание дисперсии. Wooldridge (2007) приводит простую корректировку, доступную в случае, когда $\hat{p}(\cdot)$ рассматривается как параметрическая модель.

При условии (44) получаем выражение, взвешенное на обратную вероятность

$$\tau_{ate} = \mathbb{E}\left[\frac{(W - p(X))\Delta Y}{p(X)(1 - p(X))}\right],$$

берущее начало в работе Horvitz & Thompson (1952). Чтобы оценить средний эффект воздействия для определенной популяции, необходимо предположение о полном пересечении (42), и тогда

$$\hat{\tau}_{ate,ps} = N^{-1}\sum_{i=1}^{N}\frac{(W_i - \hat{p}(X_i))\Delta Y_i}{\hat{p}(X_i)(1 - \hat{p}(X_i))}. \tag{49}$$

Hirano, Imbens & Ridder (2003) изучают эту оценку в деталях, когда $\hat{p}(x)$ – серийная логит-оценка. Если воспринимать эту оценку параметрически, простая корректировка делает легким проведение правильной инференции для $\hat{\tau}_{ate}$. Пусть $\hat{K}_i$ – слагаемое из (49) за вычетом $\hat{\tau}_{ate}$, а $\hat{D}_i = h(X_i)(W_i - \Lambda(h(X_i)\hat{\gamma}))$ – градиент (вектор-строка) логит-оценивания. Подсчитаем остатки $\hat{R}_i$ МНК-регрессии $\hat{K}_i$ на $\hat{D}_i$, $i = 1, ..., N$. Тогда, состоятельная оценка $Avar\sqrt{N}(\hat{\tau}_{ate,ps} - \tau_{ate})$ – это просто выборочная дисперсия $\hat{R}_i$. Она никогда не превышает дисперсию, которая была бы получена при игнорировании оценивания $p(x)$ и простом использовании выборочной дисперсии самих $\hat{K}_i$.

Методы, которые сочетают регрессионную корректировку и взвешивание с помощью оцениваемой вероятности воздействия, также применимы к разностям при наличии панельных данных. Подробности можно найти в Imbens (2004), Wooldridge (2007), и Imbens & Wooldridge (2009).

6 Методы синтетической контрольной группы для сравнительных кейсов

В разделе 3 обсуждались РР-методы, игнорирующие выборочную вариацию при подсчете в средних по группам и временным периодам (в более общем случае, в регрессионных коэффициентах). Abadie, Diamond & Hainmueller (2007), основываясь на работе Abadie & Gardeazabal (2003), утверждают, что при анализе программ на общем уровне нет неточности оценивания: цель – определить эффект программы для всей популяции, скажем, штата, а совокупные величины измеряются без ошибки (или с очень маленькой ошибкой).

Конечно, один из источников вариации в каждом исследовании с применением данных, меняющихся во времени, – изменение исходов во времени, даже если те являются совокупными величинами, измеряемыми без ошибки. Методология кейс-стади – один из таких примеров: часто регрессии для временных рядов по одному объекту, такому как штат, применяются для определения эффекта программы (изменения скоростных ограничений, программы по контролю курения табака и так далее) на совокупный исход. Но недостатком таких кейсов является то, что в них не используется контрольная группа для учета совокупных эффектов, не имеющих ничего общего с определенной программой в штате.

В контексте исследований «случай-контроль», когда временной ряд доступен для конкретного объекта – опытной группы – часто бывает много потенциальных контрольных групп. Скажем, в примере с программой по контролю курения табака каждый штат США является потенциальной контрольной группой для Калифорнии (при условии, что он не проводил аналогичную программу). ADH изучают такую постановку и подчеркивают важность вариации, связанной с выбором подходящих контрольных групп. Они обращают внимание на то, что даже при отсутствии выборочной ошибки при анализе программ на уровне штата необходимо учитывать подобную неопределенность.

Поход ADH состоит в том, чтобы выбирать синтетическую контрольную группу из набора возможных контрольных групп. Например, в случае контроля курения табака в Калифорнии ADH выделяют 38 штатов, в которых подобные программы не проводились в исследуемый временной период. Вместо того чтобы использовать стандартный анализ с фиксированными эффектами (который фактически предполагает, что каждый штат одинаково хорош в качестве контрольной группы)б ADH предлагают выбирать взвешенное среднее потенциальных контрольных групп. Конечно, выбор подходящей контрольной группы или групп часто делается неформально, включая сопоставление на основе значений регрессоров в предпрограммный период. ADH формализуют процедуру выбора оптимальным образом выбирая веса, и предлагают способы проведения инференции.

Рассмотрим простой пример с двумя временными периодами: одним до реализации программы и одним после. Пусть y_{it} – исход для объекта i в периоде t, где $i = 1$ – объект, подверженный воздействию во втором периоде. Предположим, имеется J возможных контрольных групп, индексируемых как $\{2, ..., J+1\}$. Пусть $\mathbf{x}_i$ – наблюдаемые регрессоры для объекта i, которые не затрагиваются (или не были бы затронуты) программой; $\mathbf{x}_i$ может содержать регрессоры периода $t = 2$, если на них не влияет программа. Обычно эффект программы можно оценить как

$$y_{12} - \sum_{j=2}^{J+1} w_j y_{j2},$$

где w_j – неотрицательные веса, в сумме дающие единицу. Вопрос заключается в следующем: как выбрать веса, то есть синтетическую контрольную группу, чтобы получить лучшую оценку эффекта программы? ADH предлагают выбирать веса, минимизирующие, в простейшем случае, расстояние между $(y_{11}, \mathbf{x}_1)$ и $\sum_{j=2}^{J+1} w_j \cdot (y_{j1}, \mathbf{x}_j)$ или некоторую линейную комбинацию элементов $(y_{11}, \mathbf{x}_1)$ и $(y_{j1}, \mathbf{x}_j)$. Оптимальные веса – которые различаются в зависимости от определения расстояния – дают синтетическую контрольную группу, чьи исходы в период перед программой и регрессоры в период проведения программы «наиболее близки». При наличии более чем двух периодов можно использовать средние значения исходов в период перед программой или взвешенные средние, которые дают больший вес более недавним исходам в предпрограммный период.

ADH предлагают методы перестановки для проведения инференции, которые требуют оценки гипотетических эффектов воздействия для каждого региона (потенциальной контрольной группы), используя тот же метод синтетической контрольной группы, что и для региона, в котором проводилась программа. Таким способом можно сравнить, является ли оценка эффекта воздействия при использовании метода синтетической контрольной группы существенно больше оценки этого эффекта при случайном выборе региона. Инференция является точной даже в случае, когда совокупные исходы оцениваются с ошибкой по данным на индивидуальном уровне.

Литература

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72, 1–19.

Abadie, A. & J. Gardeazabal (2003). Economic costs of conflict: A case study of the Basque Country. *American Economic Review* 93, 113–132.

Abadie, A., A. Diamond & J. Hainmueller (2007). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *NBER Technical Working Paper* No. 335.

Altonji, J.G. & R.L. Matzkin (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73, 1053–1102.

Angrist, J.D. & V. Lavy (2002). The effect of high school matriculation awards: Evidence from randomized trials. *NBER Working Paper* No. 9389.

Ashenfelter, O. & D. Card (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics* 67, 648–660.

Athey, S. & G.W. Imbens (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica* 74, 431–497.

Bertrand, M., E. Duflo & S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 249–275.

Card, D. & A.B. Krueger (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* 84, 772–793.

Dehejia, R.H. & S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of American Statistical Association* 94, 1053–1062.

Donald, S.G. & K. Lang (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics* 89, 221–233.

Hansen, C.B. (2007a). Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141, 597–620.

Hansen, C.B. (2007b). Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics* 140, 670–694.

Hausman, J.A. & G. Kuersteiner (2005). Difference in difference meets generalized least squares: Higher order properties of hypotheses tests. Working Paper No. 2005-010, Boston University.

Heckman, J., H. Ichimura & P. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies* 64, 605–654.

Hirano, K., G.W. Imbens & G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71, 1161–1189.

Horvitz, D. & D. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association* 47, 663–685.

Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* 86, 1–29.

Imbens, G.W. & J.M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, forthcoming.

Kiefer, N.M. (1980). Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance. *Journal of Econometrics* 14, 195–202.

Loeb, S. & J. Bound (1996). The effect of measured school inputs on academic achievement: Evidence from the 1920s, 1930s and 1940s birth cohorts. *Review of Economics and Statistics* 78, 653–664.

Meyer, B.D., W.K. Viscusi & D.L. Durbin (1995). Workers' compensation and injury duration: Evidence from a natural experiment. *American Economic Review* 85, 322–340.

Solon, G. (1984). Estimating autocorrelations in fixed-effects models. *NBER Technical Working Paper* No. 032.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Wooldridge, J.M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review* 93, 133–138.

Wooldridge, J.M. (2005). Fixed effects and related estimators for correlated random-coefficient and treatment effect panel data models. *Review of Economics and Statistics* 87, 385–390.

Wooldridge, J.M. (2007). Inverse probability weighted M-estimation for general missing data problems. *Journal of Econometrics* 141, 1281–1301.

Difference-in-differences estimation

Jeffrey M. Wooldridge

Michigan State University, East Lansing, USA

This article provides an overview of difference-in-differences estimation, starting with a review of the basic methodology, discussing in some detail recent advances in inference, and concluding with new methods for estimating treatment effects in various nonlinear and semiparametric models.

Эконометрический ликбез: ограниченные зависимые переменные

Оценивание моделей дискретного выбора и моделей с цензурированием*

Эрик Бьорн†

Университет Осло, Осло, Норвегия

В настоящих заметках содержится обзор вопросов спецификации модели, функции правдоподобия и структуры задач максимального правдоподобия для моделей дискретного выбора и моделей с цензурированием. Первая часть касается оценивания в случае одного уравнения с одномерными (кросс-секционными) наблюдениями. Другая часть расширяет постановку на случай двух уравнений. Последняя часть рассматривает расширение на ситуацию панельных данных.

1 Одномерные (кросс-секционные) данные

1.1 Отправная точка

Отправной точкой является следующее уравнение:

$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \sigma\varepsilon_i, \quad \varepsilon_i \sim \mathsf{IIN}(0,1), \quad i = 1,\ldots,N, \tag{1}$$

где i обозначает номер наблюдения, IIN символизирует «одинаково, независимо и нормально распределены», y_i^* – значение эндогенной переменной для наблюдения i, $\boldsymbol{x}_{it}$ – вектор-строка наблюдаемых ковариат (экзогенных переменных), $\boldsymbol{\beta}$ – вектор-столбец коэффициентов, σ – положительная константа, и ε_i – ненаблюдаемый случайный шум. Мы не наблюдаем $(y_i^*, \boldsymbol{x}_i)$ для всех i. *Различия между тремя моделями, представленными ниже, определяются тем, как наблюдаются пары* (y_i^*, x_i). Эти три модели представляют из себя вариации на одну тему: *модели с ограниченной наблюдаемостью эндогенных переменных.* Мы будем обозначать наблюдения как (y_i, x_i). Нашей целью является несмещенная инференция о $\boldsymbol{\beta}$, т.е. об эффекте изменений в $\boldsymbol{x}_i$ на y_i^*, исходя из множества наблюдений $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$.

1.2 Что мы наблюдаем?

Мы рассмотрим три случая, отличающихся по тому, как наблюдается латентная эндогенная переменная y_i^*.

Случай 1: Случай дискретного выбора. Мы наблюдаем $\boldsymbol{x}_i$ и

$$y_i = \begin{cases} 1 \quad \text{для} \quad y_i^* > 0 \quad \Leftrightarrow \quad -\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \\ 0 \quad \text{для} \quad y_i^* \le 0 \quad \Leftrightarrow \quad -\varepsilon_i \ge \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \end{cases} \qquad i = 1,\ldots,N. \tag{2}$$

*Перевод С. Анатольева. Цитировать как: Бьорн, Эрик (2009) «Оценивание моделей дискретного выбора и моделей с цензурированием», Квантиль, №6, стр. 49–57. Citation: Biørn, Erik (2009) “Estimation of discrete choice and censoring models,” Quantile, No.6, pp. 49–57.

†Адрес: Department of Economics, University of Oslo, P.O. Box 1095 Blindern, 0317 Oslo, Norway. Электронная почта: *erik.biorn@econ.uio.no*

Формально, y_i – это ступенчатая функция от y_i^*, со *ступенькой в нуле*. Если мы в общем случае определим функцию $z = \mathbb{I}\{\mathcal{A}\}$, равную единице если событие $\mathcal{A}$ верно и нулю если событие $\mathcal{A}$ неверно, мы сможем записать (2) компактно как

$$y_i = \mathbb{I}\{y_i^* > 0\} = \mathbb{I}\{-\varepsilon_i < \boldsymbol{x}_i\boldsymbol{\beta}/\sigma\}, \qquad i = 1, \ldots, N.$$

Случай 2: Случай с цензурированием. В этом Случае мы предполагаем, что знаем больше об y_i^*, чем в Случае 1. Мы наблюдаем $\boldsymbol{x}_i$ и

$$y_i = \max\{y_i^*, 0\} = \begin{cases} y_i^* & \text{для} \quad y_i^* > 0 \quad \Leftrightarrow \quad -\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \\ 0 & \text{для} \quad y_i^* \le 0 \quad \Leftrightarrow \quad -\varepsilon_i \ge \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, \end{cases} \qquad i = 1, \ldots, N. \tag{3}$$

Особенность этого Случая в том, что y_i *наблюдаема частично непрерывно (для* $y_i^* = y_i > 0$*), а частично дискретно (для* $y_i^* \le 0$, $y_i = 0$*)*. Формально, y_i – непрерывная функция от y_i^*, с *изломом в нуле*. Наблюдения по y_i характеризуются *нагромождением нулей*.

Случай 3: Случай с отсечением. В этом Случае мы знаем меньше, чем в Случае 2, но больше, чем в Случае 1, только если наблюдается положительное y_i. У нас нет наблюдений для всех N значений i. Наблюдения определяются следующим образом:

$$\begin{array}{lll} (y_i, \boldsymbol{x}_i) = (y_i^*, \boldsymbol{x}_i) \text{ и наблюдаема,} & \text{если} \quad y_i^* > 0 \quad \Leftrightarrow \quad -\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, & \\ & & i = 1, \ldots, N. \\ (y_i, \boldsymbol{x}_i) \text{ ненаблюдаема,} & \text{если} \quad y_i^* \le 0 \quad \Leftrightarrow \quad -\varepsilon_i \ge \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}, & \end{array} \tag{4}$$

В этом Случае процесс, определяющий, будут ли у нас наблюдения для конкретного значения i или нет, случаен, и *этот выбор есть результат решений респондентов, определяемых рассматриваемой моделью*, (1).

1.3 Вероятностная структура откликов в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Точечные вероятности двух возможных исходов для y_i, условно на $\boldsymbol{x}_i$, равны

$$\begin{array}{l} \mathbb{P}\{y_i = 1 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\left\{-\varepsilon_i < \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\middle|\boldsymbol{x}_i\right\} = \Phi\left(\dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{L}_{1i} \\ \mathbb{P}\{y_i = 0 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\left\{-\varepsilon_i \ge \dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\middle|\boldsymbol{x}_i\right\} = 1 - \Phi\left(\dfrac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{L}_{0i}, \end{array} \qquad i = 1, \ldots, N. \tag{5}$$

где $\Phi(\cdot)$ – КФР (кумулятивная функция распределения) распределения $\mathsf{N}(0, 1)$, $\equiv$ следует интерпретировать как равенство по определению, и где первый индекс у функций $\mathcal{L}$ обозначает «Режим 1» когда $y_i = 1$, и «Режим 0» когда $y_i = 0$, соответственно. (Заметим, что мы здесь используем тот факт, что $-\varepsilon_i$ и ε_i имеют одну и ту же функцию плотности, поскольку нормальное распределение симметрично.)

Часть функции правдоподобия, «относящуюся» к наблюдению i, можно переписать как

$$\mathcal{L}_i \equiv \mathcal{L}_{1i}^{y_i}\mathcal{L}_{0i}^{1-y_i} \equiv \begin{cases} \mathcal{L}_{1i} & \text{для} \quad y_i = 1, \\ \mathcal{L}_{0i} & \text{для} \quad y_i = 0, \end{cases} \qquad i = 1, \ldots, N. \tag{6}$$

Случай 2: Регрессионная модель с цензурированием

Рассмотрим вначале Режим 1, в котором y_i непрерывна и имеет ту же КФР, что и у y_i^*:

$$\Phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(\frac{y_i^* - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \tag{7}$$

и плотность, выводимую дифференцированием (7) по y_i для $y_i > 0$:

$$\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) = \frac{1}{\sigma}\phi\left(\frac{y_i^* - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{M}_{1i}, \tag{8}$$

где $\phi(\cdot) \equiv \Phi'(\cdot)$.

Рассмотрим теперь Режим 0, в котором y_i наблюдается дискретно. Этот Режим совпадает с откликом $y_i = 0$ в Случае 1. Тогда у y_i нет плотности, а есть вероятностная масса, которую можно получить из КФР y_i^* следующим образом (см. вторую часть (5)):

$$\mathbb{P}\{y_i = 0 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\{y_i^* \leq 0 \,|\, \boldsymbol{x}_i\} = \mathbb{P}\left\{-\varepsilon_i \geq \frac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\middle| \boldsymbol{x}_i\right\} = 1 - \Phi\left(\frac{\boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) \equiv \mathcal{M}_{1i}. \tag{9}$$

Часть функции правдоподобия, «относящаяся» к наблюдению i, теперь выглядит как

$$\mathcal{M}_i \equiv \begin{cases} \mathcal{M}_{1i} & \text{для} \quad y_i > 0, \\ \mathcal{M}_{0i} & \text{для} \quad y_i = 0, \end{cases} \qquad i = 1, \ldots, N. \tag{10}$$

Эта функция, таким образом, является смесью функций плотности и КФР.

1.4 Задача максимизации правдоподобия в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Поскольку наши N наблюдений независимы, полная функция правдоподобия является произведением функций правдоподобия в (6) для всех наблюдений, что дает

$$\mathcal{L} \equiv \prod_{i=1}^{N} \mathcal{L}_i \equiv \prod_{i:y_i=1} \mathcal{L}_{1i} \prod_{i:y_i=0} \mathcal{L}_{0i}. \tag{11}$$

Максимизируя $\mathcal{L}$, или, что несколько проще, максимизируя

$$\ln(\mathcal{L}) \equiv \sum_{i=1}^{N} \ln(\mathcal{L}_i) = \sum_{i:y_i=1} \ln(\mathcal{L}_{1i}) + \sum_{i:y_i=0} \ln(\mathcal{L}_{0i}) \equiv H\left(\frac{\boldsymbol{\beta}}{\sigma}\right)$$

по $\boldsymbol{\beta}/\sigma$, получаем оценки максимального правдоподобия этих параметров. Эту задачу приходится решать численно.

Решение данной максимизиционной задачи подразумевает интегрирование: КФР распределения $\mathsf{N}(0,1)$, $\Phi(\cdot)$, определяется как некий интеграл. Заметим, что *мы не можем оценить уровень вектора коэффициентов $\boldsymbol{\beta}$, а можем лишь отношение этого вектора к стандартному отклонению возмущения в (1).* Интерпретация: можно измерить (идентифицировать) лишь отклик на переменные в $\boldsymbol{x}_i$ в «единицах стандартного отклонения». Это следствие того, что наблюдаемы лишь качественные свойства y_i^*.

Случай 2: Регрессионная модель с цензурированием

Поскольку наши N наблюдений независимы, полная функция правдоподобия является произведением функций правдоподобия в (10) для всех наблюдений, что дает

$$\mathcal{M} \equiv \prod_{i=1}^{N} \mathcal{M}_i = \prod_{i:y_i>0} \mathcal{M}_{1i} \prod_{i:y_i=0} \mathcal{M}_{0i}. \tag{12}$$

Максимизируя $\mathcal{M}$, или, что несколько проще, максимизируя

$$\ln(\mathcal{M}) \equiv \sum_{i=1}^{N} \ln(\mathcal{M}_i) = \sum_{i:y_i>0} \ln(\mathcal{M}_{1i}) + \sum_{i:y_i=0} \ln(\mathcal{M}_{0i}) \equiv G(\boldsymbol{\beta}, \sigma)$$

по $(\boldsymbol{\beta}, \sigma)$, получаем оценки максимального правдоподобия этих параметров.

Еще раз отметим, что решение подразумевает интегрирование: КФР $\mathsf{N}(0,1)$-распределения, $\Phi(\cdot)$, реализующегося в Режиме 0, определяется как некий интеграл. *В данном случае мы уже можем оценить абсолютное значение вектора коэффициентов $\boldsymbol{\beta}$ вместе с параметром σ.* Это происходит благодаря тому, что, в отличие от полностью дискретного Случая 1, y_i^* на некотором отрезке наблюдаема как количественная (непрерывная) переменная. Этого достаточно для раздельной идентификации $\boldsymbol{\beta}$ и σ.

2 Первое расширение: модель из двух уравнений

Следующая модель, которую мы рассмотрим, состоит из двух уравнений в форме (1):

$$\begin{array}{l} y_{1i}^* = \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + \sigma_1\varepsilon_{1i}, \\ y_{2i}^* = \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 + \sigma_2\varepsilon_{2i}, \end{array} \qquad \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{13}$$

$$\begin{aligned} y_{1i} &= \begin{cases} y_{1i}^* & \text{если} \quad y_{1i}^* > 0, \\ 0 & \text{если} \quad y_{1i}^* \le 0, \end{cases} \\ y_{2i} &= \begin{cases} y_{2i}^* & \text{если} \quad y_{1i}^* > 0, \\ 0 & \text{если} \quad y_{1i}^* \le 0, \end{cases} \end{aligned} \tag{14}$$

где (y_{1i}, y_{2i}) цензурированы в нуле в зависимости от знака y_{1i}^*: строго положительные значения (y_{1i}, y_{2i}) наблюдаемы только если $y_{1i} > 0$.

Из (13)–(14) находим, что

$$\begin{aligned} \mathbb{E}(y_{1i}|y_{1i} > 0) &= \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 - \sigma_1\mathbb{E}[-\varepsilon_{1i}| - \varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)] \\ &\equiv \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 - \sigma_1\mathbb{E}[\varepsilon_{1i}|\varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)], \end{aligned} \tag{15}$$

$$\begin{aligned} \mathbb{E}(y_{2i}|y_{1i} > 0) &= \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[-\varepsilon_{2i}| - \varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)] \\ &\equiv \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[\varepsilon_{2i}|\varepsilon_{1i} < \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)], \end{aligned} \tag{16}$$

$$\begin{aligned} \mathbb{E}(y_{2i}|y_{1i} = 0) &= \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[-\varepsilon_{2i}| - \varepsilon_{1i} > \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)] \\ &\equiv \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \sigma_2\mathbb{E}[\varepsilon_{2i}|\varepsilon_{1i} > \boldsymbol{x}_{1i}(\boldsymbol{\beta}_1/\sigma_1)]. \end{aligned} \tag{17}$$

Поскольку предположение о нормальности в (13) влечет за собой

$$\mathbb{E}[\varepsilon_{ji}|\varepsilon_{ji}<a] = -\frac{\phi(a)}{\Phi(a)}, \quad \mathbb{E}[\varepsilon_{ji}|\varepsilon_{ji}>a] = \frac{\phi(a)}{1-\Phi(a)}, \quad j = 1, 2;\ a \in (-\infty, +\infty),$$

$$\mathbb{E}[\varepsilon_{ji}|\varepsilon_{ki}] = \rho\,\varepsilon_{ki}, \quad j, k = 1, 2;\ j \ne k,$$

то из (15)–(17) следует в результате использования правила повторных ожиданий, что

$$\mathbb{E}(y_{1i}|y_{1i} > 0) = \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + \sigma_1\lambda_{Ai}, \tag{18}$$

$$\mathbb{E}(y_{2i}|y_{1i} > 0) = \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 + \rho\sigma_2\lambda_{Ai}, \tag{19}$$

$$\mathbb{E}(y_{2i}|y_{1i} = 0) = \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 - \rho\sigma_2\lambda_{Bi}, \tag{20}$$

где

$$\lambda_{Ai} = \frac{\phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}{\Phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}, \qquad \lambda_{Bi} = \frac{\phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}{1 - \Phi\left(\dfrac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)}. \tag{21}$$

Теперь можно легко объяснить двухшаговую процедуру Хекмана.

Шаг 1: Провести пробит-анализ на первом уравнении (13), задействуя только знак y_{1i} как качественную переменную и наблюдаемые значения $\boldsymbol{x}_{1i}$. Это дает $\widehat{\boldsymbol{\beta}_1/\sigma_1}$, откуда можно найти $\widehat{\lambda}_{Ai}$ и $\widehat{\lambda}_{Bi}$, используя (21).

Шаг 2:
(i) Для *нецензурированных* наблюдений, т.е. тех, для которых $y_{1i} > 0$, прорегрессировать y_{1i} на $\boldsymbol{x}_{1i}$ и $\widehat{\lambda}_{Ai}$, используя (18). Это дает $(\widehat{\boldsymbol{\beta}}_1, \widehat{\sigma}_1)$.
(ii) Для *нецензурированных* наблюдений, т.е. тех, для которых $y_{1i} > 0$, прорегрессировать y_{2i} на $\boldsymbol{x}_{2i}$ и $\widehat{\lambda}_{Ai}$, используя (19). Это дает $(\widehat{\boldsymbol{\beta}}_2, \widehat{\rho\sigma}_2)$. Или же для *цензурированных* наблюдений, т.е. тех, для которых $y_{1i} = 0$, прорегрессировать y_{2i} на $\boldsymbol{x}_{2i}$ и $\widehat{\lambda}_{Bi}$, используя (20). Это дает $(\widehat{\boldsymbol{\beta}}_2, \widehat{\rho\sigma}_2)$.

Обозначим через $f(u_{1i}, u_{2i}; \sigma_1, \sigma_2, \rho)$ плотность возмущений $(u_{1i}, u_{2i}) = (\sigma_1\varepsilon_{1i}, \sigma_2\varepsilon_{2i})$ в (13). Если бы цензурирования не было, то $f(y_{1i} - \boldsymbol{x}_{1i}\boldsymbol{\beta}_1, y_{2i} - \boldsymbol{x}_{2i}\boldsymbol{\beta}_2; \sigma_1, \sigma_2, \rho)$ была бы плотностью (y_{1i}, y_{2i}) (условно на $\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}$) на всей области (y_{1i}, y_{2i}). Но (y_{1i}, y_{2i}) распределены не непрерывно на той области, где цензурирование действует. В этой ситуации правдоподобия для наблюдений выводятся следующим образом.

Обозначим через $\mathcal{M}_i$ часть функции правдоподобия, относящуюся к наблюдению i. Имеем: $\mathcal{M}_i = f(y_{1i} - x_{1i}\beta_1, y_{2i} - x_{2i}\beta_2; \sigma_1, \sigma_2, \rho)$ для наблюдений с $y_{1i} > 0$ и $\mathcal{M}_i = \mathbb{P}\{y^*_{1i} \leq 0\} = \mathbb{P}\{-\varepsilon_{1i} > \boldsymbol{x}_{1i}\boldsymbol{\beta}_1/\sigma_1\} = 1 - \Phi(\boldsymbol{x}_{1i}\boldsymbol{\beta}_1/\sigma_1)$ для наблюдений с $y_{i1} = 0$.

В предположении о независимости наблюдений все это вместе означает, что полная функция правдоподобия имеет вид

$$\mathcal{M} = \prod_{i:y_{1i}>0} f(y_{1i} - \boldsymbol{x}_{1i}\boldsymbol{\beta}_1, y_{2i} - \boldsymbol{x}_{2i}\boldsymbol{\beta}_2; \sigma_1, \sigma_2, \rho) \prod_{i:y_{1i}=0} \left[1 - \Phi\left(\frac{\boldsymbol{x}_{1i}\boldsymbol{\beta}_1}{\sigma_1}\right)\right], \tag{22}$$

где $\prod_{i:y_{1i}>0}$ и $\prod_{i:y_{1i}=0}$ символизируют произведения по тем значениям i, для которых $y_{1i} > 0$, и по тем значениям i, для которых $y_{1i} = 0$, соответственно. Максимизация $\mathcal{M}$ по неизвестным параметрам дает оценки максимального правдоподобия.

3 Второе расширение: панельные данные

3.1 Отправная точка

Отправной точкой при расширении модели первого раздела на случай (сбалансированных) панельных даных является следующее уравнение:

$$y^*_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i + \sigma\varepsilon_{it}, \quad \varepsilon_{it} \sim \mathsf{IIN}(0,1), \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \tag{23}$$

Мы, правда, не наблюдаем $(y^*_{it}, \boldsymbol{x}_{it})$ для всех (i, t). Предполагается, что α_i – латентный индивидуальный эффект, который можно рассматривать как фиксированный и полностью неизвестный и неструктурированный, или же как случайный и порожденный распределением вероятностей с определенными свойствами. Мы будем придерживаться его интерпретации как случайного эффекта, хотя в определенных местах и будем действовать условно на α_i, что можно рассматривать как подражание случаю с фиксированными эффектами. Нашей целью является несмещенная инференция о $\boldsymbol{\beta}$, т.е., об эффекте изменений в $\boldsymbol{x}_{it}$ на y^*_{it}, исходя из множества наблюдений $\{\{y_{it}, \boldsymbol{x}_{it}\}_{i=1}^N\}_{t=1}^T$.

3.2 Что мы наблюдаем?

Мы рассмотрим три случая, отличающихся по тому, как наблюдается латентная эндогенная переменная y^*_{it}.

Случай 1: Случай дискретного выбора. Мы наблюдаем $\boldsymbol{x}_{it}$ и

$$y_{it}=\begin{cases}1 \quad \text{для} \quad y_{it}^*>0 \quad \Leftrightarrow \quad -\varepsilon_{it} < \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}, \\ 0 \quad \text{для} \quad y_{it}^*\le 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \ge \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma},\end{cases} \qquad \begin{matrix} i=1,\ldots,N, \\ t=1,\ldots,T, \end{matrix} \tag{24}$$

где i индексирует индивида, а t – период времени. Формально, y_{it} – это ступенчатая функция от y_{it}^*, со *ступенькой в нуле*. Можно записать (24) компактно как

$$y_{it}=\mathbb{I}\{y_{it}^*>0\}=\mathbb{I}\{-\varepsilon_{it} < (\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i)/\sigma\}, \qquad i=1,\ldots,N,\ t=1,\ldots,T.$$

Случай 2: Случай с цензурированием. В этом Случае мы знаем больше об y_{it}^*, чем в Случае 1. Мы наблюдаем $\boldsymbol{x}_{it}$ и

$$y_{it}=\max\{y_{it}^*,0\}=\begin{cases}y_{it}^* \quad \text{для} \quad y_{it}^*>0 \quad \Leftrightarrow \quad -\varepsilon_{it} < \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}, \\ 0 \quad \text{для} \quad y_{it}^*\le 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \ge \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma},\end{cases} \qquad \begin{matrix} i=1,\ldots,N, \\ t=1,\ldots,T. \end{matrix} \tag{25}$$

Особенность этого Случая в том, что y_{it} *наблюдаема частично непрерывно (для* $y_{it}^*=y_{it}>0$*), а частично дискретно (для* $y_{it}^*\le 0$, $y_{it}=0$*)*. Формально, y_{it} – непрерывная функция от y_{it}^*, с *изломом в нуле*. Наблюдения по y_{it} характеризуются *нагромождением нулей.*

Случай 3: Случай с отсечением. В этом Случае мы знаем меньше, чем в Случае 2, но больше, чем в Случае 1, только для тех индивидов в те периоды времени, когда они ответили положительно. У нас нет наблюдений для всех NT комбинаций (i,t). Наблюдения определяются следующим образом:

$$\begin{matrix}(y_{it},\boldsymbol{x}_{it})=(y_{it}^*,\boldsymbol{x}_{it}) \text{ и наблюдаема}, & \text{если} \quad y_{it}^*>0 \quad \Leftrightarrow \quad -\varepsilon_{it} < \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}, \\ (y_{it},\boldsymbol{x}_{it}) \text{ ненаблюдаема}, & \text{если} \quad y_{it}^*\le 0 \quad \Leftrightarrow \quad -\varepsilon_{it} \ge \dfrac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma},\end{matrix} \qquad \begin{matrix} i=1,\ldots,N, \\ t=1,\ldots,T. \end{matrix}$$

В этом Случае процесс, определяющий, будут ли у нас наблюдения для конкретных комбинаций (i,t) или нет, случаен, и *этот выбор есть результат индивидуальных решений респондентов, определяемых рассматриваемой моделью*, (1). Выборка представляет собой набор несбалансированных панельных данных, и *вид несбалансированности определяется эндогенно.*

3.3 Вероятностная структура откликов в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Точечные вероятности двух возможных исходов для y_{it}, условно не только на $\boldsymbol{x}_{it}$, но также на индивидуальном эффекте α_i, равны

$$\begin{aligned}\mathbb{P}\{y_{it}=1\,|\,\boldsymbol{x}_{it},\alpha_i\}&=\mathbb{P}\left\{-\varepsilon_{it} < \frac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}\middle|\boldsymbol{x}_{it},\alpha_i\right\} \\ &=\Phi\left(\frac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}\right)\equiv\mathcal{L}_{1it}(\alpha_i), \\ \mathbb{P}\{y_{it}=0\,|\,\boldsymbol{x}_{it},\alpha_i\}&=\mathbb{P}\left\{-\varepsilon_{it} \ge \frac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}\middle|\boldsymbol{x}_{it},\alpha_i\right\} \\ &=1-\Phi\left(\frac{\boldsymbol{x}_{it}\boldsymbol{\beta}+\alpha_i}{\sigma}\right)\equiv\mathcal{L}_{0it}(\alpha_i),\end{aligned} \qquad \begin{matrix} i=1,\ldots,N, \\ t=1,\ldots,T. \end{matrix} \tag{26}$$

Вновь первый индекс у функций $\mathcal{L}$ обозначает «Режим 1» когда $y_{it} = 1$, и «Режим 0» когда $y_{it} = 0$, соответственно.

Согласно определению имеем

$$\mathcal{L}_{it}(\alpha_i) \equiv \mathcal{L}_{1it}(\alpha_i)^{y_{it}} \mathcal{L}_{0it}(\alpha_i)^{1-y_{it}} \equiv \begin{cases} \mathcal{L}_{1it}(\alpha_i) & \text{для} \quad y_{it} = 1, \quad i = 1, \dots, N, \\ \mathcal{L}_{0it}(\alpha_i) & \text{для} \quad y_{it} = 0, \quad t = 1, \dots, T. \end{cases} \tag{27}$$

Часть функции правдоподобия, «относящуюся» к индивиду i, условную на α_i, можно переписать из-за независимости $y_{i1}, \dots, y_{iT}$ ввиду (23) как

$$\mathcal{L}_i(\alpha_i) \equiv \prod_{t=1}^{T} \mathcal{L}_{1it}(\alpha_i)^{y_{it}} \mathcal{L}_{0it}(\alpha_i)^{1-y_{it}} \equiv \prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i). \tag{28}$$

Здесь $t : y_{it} = 1$ и $t : y_{it} = 0$ внизу оператора взятия произведения $\prod$ означают, для индивида i, взятие произведений по всем t таким, что $y_{it} = 1$, и по всем t таким, что $y_{it} = 0$, соответственно.

Соответствующая часть *маржинальной* функции правдоподобия выводится следующим образом. Предположим, что α_i имеет функцию плотности $g(\alpha_i; \gamma)$, где γ – вектор неизвестных параметров, включающий среднее и стандартное отклонение. Тогда маржинальным аналогом (28) является

$$\mathcal{L}_i^* \equiv \int_{-\infty}^{+\infty} \mathcal{L}_i(\alpha_i)\, g(\alpha_i; \gamma) d\,\alpha_i \equiv \int_{-\infty}^{+\infty} \left[\prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i) \right] g(\alpha_i; \gamma) d\,\alpha_i. \tag{29}$$

Случай 2: Регрессионная модель с цензурированием

Рассмотрим вначале Режим 1, в котором y_{it} непрерывна и имеет ту же КФР, что и у y_{it}^*:

$$\Phi\left(\frac{y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) = \Phi\left(\frac{y_{it}^* - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) \tag{30}$$

и плотность, выводимую дифференцированием (30) по y_{it} для $y_{it} > 0$:

$$\frac{1}{\sigma}\phi\left(\frac{y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) = \frac{1}{\sigma}\phi\left(\frac{y_{it}^* - \boldsymbol{x}_{it}\boldsymbol{\beta} - \alpha_i}{\sigma}\right) \equiv \mathcal{M}_{1it}(\alpha_i). \tag{31}$$

Рассмотрим теперь Режим 0, в котором y_{it} наблюдается дискретно. Этот Режим совпадает с откликом $y_{it} = 0$ в Случае 1. Тогда у y_{it} нет плотности, а есть вероятностная масса, которую можно получить из КФР y_{it}^* следующим образом (см. вторую часть (26)):

$$\begin{aligned} \mathbb{P}\{y_{it} = 0 \,|\, \boldsymbol{x}_{it}, \alpha_i\} = \mathbb{P}\{y_{it}^* \le 0 \,|\, \boldsymbol{x}_{it}, \alpha_i\} &= \mathbb{P}\left\{ -\varepsilon_{it} \ \ge \ \frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma} \middle| \boldsymbol{x}_{it}, \alpha_i \right\} \\ &= 1 - \Phi\left(\frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i}{\sigma}\right) \equiv \mathcal{M}_{1it}(\alpha_i) \end{aligned} \tag{32}$$

Часть функции правдоподобия, «относящуюся» к индивиду i, условную на α_i, можно переписать из-за независимости $y_{i1}, \dots, y_{iT}$ ввиду (23) как

$$\mathcal{M}_i(\alpha_i) \equiv \prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i). \tag{33}$$

Эта функция правдоподобия для конкретного индивида, таким образом, является смесью функций плотности и КФР. Здесь $t : y_{it} > 0$ и $t : y_{it} = 0$ внизу оператора взятия произведения $\prod$ означают, для индивида i, взятие произведений по всем t таким, что $y_{it} > 0$, и по всем t таким, что $y_{it} = 0$, соответственно.

Соответствующая часть *маржинальной* функции правдоподобия выводится следующим образом. Предположим, что α_i имеет функцию плотности $g(\alpha_i;\gamma)$, где γ – вектор неизвестных параметров, включающий среднее и стандартное отклонение. Тогда маржинальный аналог (34) выводится интегрированием (27) по области определения α_i:

$$\mathcal{M}_i^* \equiv \int_{-\infty}^{+\infty} \mathcal{M}_i(\alpha_i) g(\alpha_i;\gamma) d\alpha_i \equiv \int_{-\infty}^{+\infty} \left[\prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i) \right] g(\alpha_i;\gamma) d\alpha_i. \tag{34}$$

3.4 Задача максимизации правдоподобия в Случаях 1 и 2

Случай 1: Модель дискретного выбора

Поскольку N индивидов наблюдаются независимо, функция правдоподобия является произведением индивидуальных функций правдоподобия в (29), что дает

$$\mathcal{L}^* \equiv \prod_{i=1}^{N} \mathcal{L}_i^* = \prod_{i=1}^{N} \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\,\alpha_i \right]. \tag{35}$$

Максимизируя $\mathcal{L}^*$, или, что несколько проще, максимизируя

$$\ln(\mathcal{L}^*) = \sum_{i=1}^{N} \ln \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}=1} \mathcal{L}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{L}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\alpha_i \right] \equiv H\left(\frac{\boldsymbol{\beta}}{\sigma}, \gamma \right)$$

по $(\boldsymbol{\beta}/\sigma, \gamma)$, получаем оценки максимального правдоподобия этих параметров. Эту задачу приходится решать численно. Заметим, что поскольку интегрирование – это что-то типа суммирования, нельзя переместить ln-оператор под интеграл!

Решение данной максимизационной задачи подразумевает интегрирование в двух местах. Во-первых, (логарифмическая) функция правдоподобия содержит интегрирование по отношению к случайному индивидуальному эффекту α_i. Во-вторых, КФР $\Phi(\cdot)$ определяется как некий интеграл. Заметим, что *мы не можем оценить уровень вектора коэффициентов $\boldsymbol{\beta}$, а можем лишь отношение этого вектора к стандартному отклонению возмущения в (23).* Интерпретация: можно измерить (идентифицировать) лишь отклик на переменные в $\boldsymbol{x}_{it}$ в «единицах стандартного отклонения». Это следствие того, что наблюдаемы лишь качественные свойства y_{it}^*.

Случай 2: Регрессионная модель с цензурированием

Поскольку N индивидов наблюдаются независимо, функция правдоподобия является произведением индивидуальных функций правдоподобия в (34), что дает

$$\mathcal{M}^* \equiv \prod_{i=1}^{N} \mathcal{M}_i^* = \prod_{i=1}^{N} \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\,\alpha_i \right]. \tag{36}$$

Максимизируя $\mathcal{M}^*$, или, что несколько проще, максимизируя

$$\ln(\mathcal{M}^*) = \sum_{i=1}^{N} \ln \left[\int_{-\infty}^{+\infty} \left(\prod_{t:y_{it}>0} \mathcal{M}_{1it}(\alpha_i) \prod_{t:y_{it}=0} \mathcal{M}_{0it}(\alpha_i) \right) g(\alpha_i;\gamma) d\alpha_i \right] \equiv G(\boldsymbol{\beta}, \sigma, \gamma)$$

по (β, σ, γ), получаем оценки максимального правдоподобия этих параметров.

И вновь решение подразумевает интегрирование в двух местах. Во-первых, (логарифмическая) функция правдоподобия содержит интегрирование по отношению к случайному индивидуальному эффекту α_i. Во-вторых, КФР $\Phi(\cdot)$, реализующегося в Режиме 0, определяется как некий интеграл. *В данном случае мы уже можем оценить абсолютное значение вектора коэффициентов* $\boldsymbol{\beta}$ *вместе с* σ *и* γ. Это происходит благодаря тому, что, в отличие от полностью дискретного Случая 1, y^*_{it} на некотором отрезке наблюдаема как количественная (непрерывная) переменная. Этого достаточно для раздельной идентификации $\boldsymbol{\beta}$ и σ.

Estimation of discrete choice and censoring models

Erik Biørn

University of Oslo, Oslo, Norway

This expository note gives an overview of model specifications, likelihood functions and a structure of maximum likelihood problems for discrete choice and censoring models. One part deals with estimation in a single equation case with unidimensional (cross-sectional) observations. Another part extends the framework to a two-equation case. The last part is concerned with an extension to a panel data situation.

Советы изучающим эконометрику

Где найти данные в сети?*

Станислав Анатольев†

Российская экономическая школа, Москва, Россия

Александр Цыплаков‡

Новосибирский государственный университет, Новосибирск, Россия

Мы приводим тематические списки полезных веб-сайтов, на которых прикладной экономист может найти данные для исследований.

Введение

В данном эссе содержатся списки полезных веб-сайтов, на которых прикладной экономист может найти данные для исследований. Авторы отдают себе отчет, что интернет меняется очень быстро и через пару лет многие из приведенных ссылок окажутся неактуальными. Тем не менее, эссе позволяет получить общее представление о диапазоне имеющихся в сети данных, названиях ресурсов, и т. п. Даже если какой-нибудь ресурс сменит URL-адрес, его можно будет отыскать по ключевым словам с помощью интернет-поиска.

Заметим, что ссылки не сортированы по значимости, и порядок перечисления довольно произволен.[1] Таким образом, то, что база данных Бюро переписи населения США указана прежде данных Мирового банка, ни в коей мере не означает, что данные по переписи населения следует считать более полезными. Читатель должен прибегнуть здесь к собственному здравому смыслу.

Для некоторых сайтов мы привели примеры данных, которые там можно найти, и цепочки надписей, по которым надо щелкать мышью. Это сделано для того, чтобы продемонстрировать потенциальную полезность соответствующей ссылки и помочь в навигации по сайтам.

Данные, связанные с бизнесом и финансами

Здесь мы приводим только ссылки на общие международные данные, связанные с бизнесом и финансами. Ссылки на данные по России см. ниже.

- Служба Google Finance:
  `finance.google.com`
  Набрать в поле ввода «NASDAQ:MYGN» – [Get quotes] → [Historical Prices] ⇒ Данные о курсе акций компании Myriad Genetics на бирже NASDAQ по дням.

- Служба Yahoo Finance:
  `finance.yahoo.com`

*Цитировать как: Анатольев, Станислав и Александр Цыплаков (2009) «Где найти данные в сети?», Квантиль, №6, стр. 59–71. Citation: Anatolyev, Stanislav, and Alexander Tsyplakov (2009) “Where to find data on the Web?” Quantile, No.6, pp. 59–71.

†Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: *sanatoly@nes.ru*

‡Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: *tsy@academ.org*

[1] В некоторых разделах ссылки расположены в порядке доменных имен второго или третьего уровня, в некоторых – по алфавиту.

- История валютных котировок – служба FXHistory:
  www.oanda.com/convert/fxhistory
  Starting Date: 01/01/00, Ending Date: 01/01/01, Swiss Franc to German Mark, Interbank rate ⇒ Курс швейцарского франка к германской марке за 2000 г., межбанковская ставка по дням.

- Поисковик финансовых данных на сайте университета штата Огайо / Ohio State University – Department of Finance – Financial Data Finder:
  fisher.osu.edu/fin/osudata.htm

- Обменные курсы, в том числе исторические, от университета Британской Колумбии / University of British Columbia – PACIFIC – Foreign Exchange Daily and Historical Rates:
  fx.sauder.ubc.ca
  [Database Retrieval] → Base Currency: U.S. Dollars, Target Currencies: British Pounds, Start Date: 1 Jan 1971, End Date: 31 Dec 1974, Data Frequency: daily – [Retrieve Data] ⇒ Курс британского фунта к доллару США за 1971–1974 гг. по дням.

- Центр исследований цен активов бизнес-школы Чикагского университета (по институциональной подписке) / University of Chicago – Graduate School of Business – Center for Research in Security Prices (CRSP):
  www.crsp.com

- Система WRDS – данные от бизнес-школы Вортона Пенсильванского университета (по институциональной подписке) / University of Pennsylvania – The Wharton School – WRDS:
  wrds.wharton.upenn.edu

- Бизнес-ссылки, собранные библиотекой университета Рутгерс / Rutgers University Libraries – Subject Research Guides – Business:
  www.libraries.rutgers.edu/rul/rr_gateway/research_guides/busi/business.shtml

- Глобальные финансовые данные (большая часть данных доступна лишь по платной подписке):
  www.globalfindata.com

Международные данные

Данные по странам мира, в том числе данные различных международных организаций.

- Международная база данных Бюро переписи населения США / U.S. Census Bureau – International Data Base (IDB):
  www.census.gov/ipc/www/idb
  [Tables] → Select one table: 028 Total Fertility Rate, Select output type: Spreadsheet mode [Отправить запрос] → How to aggregate: Treat each region separately, Aggregation output options: Both region and country data, Select one or more regions: World, Select one or more years: all available years – [Go] ⇒ Данные об уровне рождаемости по странам мира с 1996 г. по годам.

- Информационный бюллетень ЦРУ / CIA World Factbook:
  https://www.cia.gov/library/publications/the-world-factbook
  Select a country or location: World → [Communications] → Telephones - mobile cellular: [Иконка с таблицей] ⇒ Количество мобильных телефонов по странам мира.

- Международные данные по энергетике Управления энергетической информации США / Energy Information Administration (EIA) – International Energy Annual – International

Energy Data and Analysis:
www.eia.doe.gov/emeu/iea

- Форум по исследованиям в области эмпирики международной торговли / Forum for Research on Empirical International Trade (F.R.E.I.T.):
www.eiit.org/Resources.html

- Продовольственная и сельскохозяйственная организация ООН (ФАО) / Food and Agriculture Organization of the United Nations (FAO) – Statistics:
www.fao.org/corp/statistics/en

- Индекс экономической свободы / Index of Economic Freedom:
www.heritage.org/index/
[Ranking the Countries] ⇒ Рейтинг стран мира по индексу экономической свободы.

- Интернет-центр исследований в области коррупции (совместная инициатива университета Пассау и организации Transparency International) / Internet Center for Corruption Research:
www.icgg.org

- Энергетическая статистика от Международного энергетического агентства / International Energy Agency (IEA) – Energy Statistics:
www.iea.org/Textbase/stats

- Статистическое бюро Международной организации труда (МОТ) / International Labour Organization (ILO) – Bureau of Statistics:
www.ilo.org/stat
[LABORSTA – database of labour statistics] → [Strikes and Lockouts] → [Main statistics] → Select one or more countries/groups of countries: EUROPE, Select one or more tables: 9A, [Go] → [View] ⇒ Статистика забастовок и локаутов в странах Европы с 1998 г. (с пропусками).

- Глобальные, финансовые и другие данные от Международного валютного фонда (МВФ) / International Monetary Fund (IMF) – Data and Statistics:
www.imf.org/external/data.htm

- Люксембургское исследование доходов / Luxembourg Income Study:
www.lisproject.org

- Проект по обследованиям в области демографии и здравоохранения / Demographic and Health Surveys (MEASURE DHS):
www.measuredhs.com

- Международная статистика и межстрановые сопоставления на сайте NationMaster / NationMaster – World Statistics, Country Comparisons:
www.nationmaster.com

- Сайт Организации экономического сотрудничества и развития / Organization for Economic Co-operation and Development (OECD) – OECD Online:
www.oecd.org

- База данных ООН по международной торговле товарами / UN comtrade – International Merchandise Trade Statistics (IMTS):
comtrade.un.org

- Единая система доступа к данным ООН UN data:
  data.un.org

- Система TRAINS Конференции ООН по торговле и развитию (ЮНКТАД) / United Nations Conference on Trade and Development (UNCTAD) – TRAINS (Trade Analysis and Information System):
  r0.unctad.org/trains_new/index.shtm

- Статистика по общественному развитию Программы развития ООН (ПРООН) / United Nations Development Programme (UNDP) – Human Development Reports – Statistics:
  hdr.undp.org/en/statistics

- Статистический институт Организации ООН по вопросам образования, науки и культуры (ЮНЕСКО) / United Nations Educational, Scientific and Cultural Organization (UNESCO) – Institute for Statistics:
  www.uis.unesco.org

- Отдел статистики ООН / United Nations Statistics Division:
  unstats.un.org

- База данных о неравенстве доходов по странам мира от университета ООН / World Institute for Development Economics Research of the United Nations University (UNU-WIDER) – World Income Inequality Database:
  www.wider.unu.edu/research/Database/en_GB/database/

- Данные Центра международных сопоставлений Пенсильванского университета о паритетах покупательной способности и счетах национального дохода по большому количеству стран мира с 1950 года / University of Pennsylvania – Center for International Comparisons of Production, Income and Prices – Penn World Table:
  pwt.econ.upenn.edu

- Статистическая информационная система Всемирной организации здравоохранения (ВОЗ) / World Health Organization (WHO) – Statistical Information System (WHOSIS):
  www.who.int/whosis

- Данные Мирового банка / World Bank Data & Research
  econ.worldbank.org
  [Research] → [Projects & Programs] → [Living Standards Measurement Study (LSMS)] → [LSMS Survey Finder] → Make selection(s): Tajikistan ⇒ Данные обследования жизненного уровня по Таджикистану за 1999 и 2003 гг.

- Международные данные о торговле и тарифах от Всемирной торговой организации (ВТО) / World Trade Organization (WTO) – International trade and tariff data:
  www.wto.org/english/res_e/statis_e/statis_e.htm

Данные по отдельным регионам мира

- Данные Азиатского банка развития / Asian Development Bank – Economics and Statistics:
  www.adb.org/Economics

- Статистическое бюро Европейского сообщества / Statistical Office of the European Communities (Eurostat):
  epp.eurostat.ec.europa.eu

- Экономическая комиссия стран Латинской Америки и Карибского бассейна / Comisión Económica para América Latina (CEPAL) / Economic Commission for Latin America and the Caribbean (ECLAC):
  www.eclac.cl

- Данные Межамериканского банка развития / Inter-American Development Bank – Research & Data:
  www.iadb.org/research

- Статистический отдел Экономической и социальной комиссии ООН по азиатско-тихоокеанскому региону / United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) – Statistics Division:
  www.unescap.org/stat

- Статистический отдел Экономической комиссии ООН по Европе / United Nations Economic Commission for Europe (UNECE) – Statistical Division:
  www.unece.org/stats
  [Get statistics on-line] → [Forest Resources] → [Area of forest...] → Availability for Wood Supply: Forest, Forest Type: Total, Country: Select all, Year: 2005 ⇒ Площадь лесов в европейских странах в 2005 г.

- Европейский центральный банк / European Central Bank:
  www.ecb.int/home/html/index.en.html

Данные по отдельным странам

Мы не стремились перечислить здесь как можно больше ресурсов, относящиеся к статистике отдельных стран. В сети можно найти достаточно много списков ссылок на национальные статистические органы. Достаточно провести интернет-поиск по фразе "statistical agencies" (в кавычках). Ниже приведен пример такого списка (составлен японским Бюро статистики).

Чаще всего заглавная страница статистического агентства страны создана на соответствующем государственном языке. Английский вариант сайта обычно можно найти по надписи «English». Названия на национальных языках мы привели только для тех стран, в которых используется латиница. К сожалению, далеко не для всех стран бесплатно доступные данные такие же богатые, как для США (данные по США и России мы выделили в отдельные разделы, см. ниже).

- Ссылки на национальные статистические агентства от Бюро статистики Японии / Links to Statistical Agencies:
  www.stat.go.jp/english/info/148.htm

- Австралия / Australian Bureau of Statistics:
  www.abs.gov.au
  [National Accounts] → [Australian National Accounts: State Accounts] → [Downloads] → Table 11. Agricultural Income, All States: Current prices [.xls] ⇒ Государственные счета системы национальных счетов Австралии, сельскохозяйственный доход по всем штатам за 1990-2008 гг. по годам.

- Канада / Canadian economy online:
  www.canadianeconomy.gc.ca

- Чехия / Český statistický úřad / Czech Statistical Office:
  www.czso.cz

- Германия / Statistisches Bundesamt Deutschland / Germany – Federal Statistical Office:
  www.destatis.de
- Эстония / Eesti Statistika / Statistics Estonia:
  www.stat.ee
- Финляндия / Tilastokeskus – Statistikcentralen / Statistics Finland:
  www.tilastokeskus.fi
- Израиль / Israel – The Central Bureau of Statistics:
  cbs.gov.il
- Япония / Japan – Statistics Bureau:
  www.stat.go.jp/english
- Норвегия / Statistisk sentralbyrå / Statistics Norway:
  www.ssb.no
- Польша / Główny Urząd Statystyczny / Central Statistical Office of Poland:
  www.stat.gov.pl
- Турция / Türkiye İstatistik Kurumu / Turkish Statistical Institute:
  www.tuik.gov.tr
- Статистика от Банка Англии / Bank of England – Statistics:
  www.bankofengland.co.uk/statistics
- Экономические данные по Великобритании от Статистического агентства этой страны / UK Statistics Authority (Office for National Statistics):
  www.statistics.gov.uk
  www.statistics.gov.uk/default.asp
  Те же данные на сайте Biz/ed / Biz/ed – Office for National Statistics (ONS) Data:
  www.bized.co.uk/dataserv/ons/onshome.htm
- Нидерланды / Centraal Bureau voor de Statistiek (CBS) / Statistics Netherlands:
  www.cbs.nl
  [English] → Themes [Construction and housing] → [Figures] → [Dwellings; permit issued; completed] ⇒ Статистика строительства жилья в целом и по регионам за 2000-2008 гг. по годам.
- Исследовательский центр ИПМ, исследования, прогнозы, мониторинг, Беларусь, Минск – Статистические данные:
  www.research.by/rus/data

Российские данные

Основными поставщиками российских макроэкономических данных являются государственные или окологосударственные структуры. Бесплатная статистика не такая обширная, как в США, но в последнее время Федеральная служба государственной статистики существенно расширила перечень данных, доступных через ее официальный сайт.

- Центральный Банк РФ:
  www.cbr.ru
  [Статистика] → [Денежно-кредитная и финансовая статистика] → Денежная масса: 2000 г. [Получить] ⇒ Денежная масса М0 и М2 в 2000 году помесячно.

- Федеральная служба государственной статистики (бывший Госкомстат):
  www.gks.ru

- Экономическая экспертная группа:
  www.eeg.ru

- Статистическая база данных по российской экономике ИИР ГУ-ВШЭ (Центр Анализа Данных):
  stat.hse.ru

- Макроэкономика России, США, Еврозоны, Великобритании (инвестиции, инфляция, безработица, ВВП, финансы) – Агентство экономической информации ПРАЙМ-ТАСС:
  e3.prime-tass.ru/macro

- Аграрная российская информационная система:
  www.aris.ru

- Панельное обследование российских домохозяйств от университета штата Северная Каролина / University of North Carolina (UNC) – Carolina Population Center – Russia Longitudinal Monitoring Survey:
  www.cpc.unc.edu/rlms

Главный источник российских финансовых данных – база Финам. Полезными могут оказаться также сайты РБК, агенства AK&M и фондовых бирж ММВБ и РТС.

- Финам:
  www.finam.ru

- Портал РБК:
  www.rbc.ru
  www.quote.ru

- Агенство AK&M:
  www.akm.ru

- Московская межбанковская валютная биржа (ММВБ):
  www.micex.ru

- Биржа РТС:
  www.rts.ru

- Котировки ценных бумаг от Cbonds.Ru:
  www.cbonds.info/all/rus/quotes

Данные по США

Мы выделили эти ссылки в специальный раздел, поскольку свободно доступных данных много.

- Сайт FedStats объединяет официальную информацию множества агентств Федерального правительства США:
  www.fedstats.gov

- Информационная система министерства сельского хозяйства США / U.S. Department of Agriculture – Economics, Statistics and Market Information System (ESMIS):
  usda.mannlib.cornell.edu

- Бюро переписи населения США / U.S. Census Bureau:
  www.census.gov
- Бюро трудовой статистики США / U.S. Bureau of Labor Statistics:
  stats.bls.gov
- Бюро транспортной статистики США / Research and Innovative Technology Administration (RITA) – Bureau of Transportation Statistics:
  www.bts.gov
- База данных FRED от ФРС США / Federal Reserve Economic Data (FRED):
  research.stlouisfed.org/fred2
- Американские и международные данные от Бюро экономического анализа США / U.S. Bureau of Economic Analysis (BEA):
  www.bea.gov
- Данные Национального бюро экономических исследований / National Bureau of Economic Research (NBER) – Data:
  www.nber.org/data
- Экономический отчет президента США / Economic Report of the President:
  www.gpoaccess.gov/eop
- Администрация энергетической информации предоставляет официальную энергетическую статистику правительства США / Energy Information Administration (EIA):
  www.eia.doe.gov
- Отчеты о преступности Федерального бюро расследований США (ФБР) / Federal Bureau of Investigation – Uniform Crime Reports:
  www.fbi.gov/ucr/ucr.htm
- Информация от Совета управляющих Федеральной резервной системы США (ФРС) / Federal Reserve Board (FRB) – Economic Reseach & Data:
  www.federalreserve.gov/econresdata/default.htm
- Национальный центр статистики здравоохранения / National Center for Health Statistics (NCHS):
  www.cdc.gov/nchs
- Панельное исследование динамики доходов Мичиганского университета / University of Michigan – Panel Study of Income Dynamics (PSID):
  psidonline.isr.umich.edu
- Статистика по налогам и доходам Налогового управления США / Internal Revenue Service – Tax Statistics:
  www.irs.gov/taxstats
- Служба STAT-USA/Internet Министерства торговли США предоставляет различную экономическую информацию федерального правительства / U.S. Department of Commerce – STAT-USA/Internet:
  www.stat-usa.gov
- База данных о состоянии крупных городов / State of the Cities Data System (SOCDS):
  socds.huduser.org

- Торговая статистика США в базе данных TradeStats Express Министерство торговли / U.S. Department of Commerce – TradeStats Express:
  tse.export.gov

Данные, использованные в учебниках

Во многих учебниках и монографиях по эконометрике имеются учебные эмпирические примеры и упражнения. Если раньше такой учебник обычно сопровождал компьютерный диск с данными и, возможно, программами, сейчас такая информация размещается на веб-сайте учебника. Вот некоторые из них (см. также обзоры самих учебников в Анатольев, 2007, 2008):

- *Econometrics*, 4-е издание, автор Badi H. Baltagi:
  www.springer.com/economics/econometrics/book/978-3-540-76515-8
- *Econometric Analysis of Panel Data*, 3-е издание, автор Badi H. Baltagi:
  www.wiley.com/legacy/wileychi/baltagi3e/data_sets.html
- *The Practice of Econometrics*, автор Ernst R. Berndt:
  www.stanford.edu/~clint/berndt
  shazam.econ.ubc.ca/student/berndt/data.htm
- *Microeconometrics: Methods and Applications*, авторы Colin Cameron & Pravin Trivedi:
  cameron.econ.ucdavis.edu/mmabook/mmadata.html
- *Econometric Theory and Methods*, авторы Russell Davidson & James G. MacKinnon:
  www.econ.queensu.ca/ETM/data
- *Introduction to Econometrics*, 3-е издание, автор Christopher Dougherty:
  www.oup.com/uk/orc/bin/9780199280964/01student/datasets
- *Time Series Models for Business and Economic Forecasting*, автор Philip Hans Franses:
  people.few.eur.nl/franses/research/data/tsa-data-2002.zip
- *Nonlinear Time Series Models in Empirical Finance*, авторы Philip Hans Franses & Dick van Dijk:
  people.few.eur.nl/djvandijk/nltsmef/nltsmef.htm
- *Econometric Analysis*, 6-е издание, автор William Greene:
  pages.stern.nyu.edu/~wgreene/Text/Edition6/tablelist6.htm
- *Learning and Practicing Econometrics*, авторы William E. Griffiths, R. Carter Hill & George G. Judge:
  ftp://ftp.wiley.com/public/college/economics/griffiths
- *Time Series Analysis*, автор James D. Hamilton:
  www.estima.com/Hamilton's\%20Time\%20Series\%20Analysis.shtml
- *Econometrics*, автор Fumio Hayashi:
  fhayashi.fc2web.com/hayashi_econometrics.htm
- *Principles of Econometrics*, 3-е издание, авторы R. Carter Hill, William E. Griffiths & Guay C. Lim:
  bcs.wiley.com/he-bcs/Books?action=resource&bcsId=4137&itemId=0471723606&resourceId=12080

- *Theory and Practice of Econometrics*, 2-е издание, авторы George G. Judge, William E. Griffiths, R. Carter Hill, Helmut Lütkepohl & Tsoung-Chao Lee:
  ftp://ftp.wiley.com/public/college/economics/hill
- *State-Space Models with Regime-Switching*, авторы Chang-Jin Kim & Charles R. Nelson:
  www.econ.washington.edu/user/cnelson/markov/prgmlist.htm
- *Applied Time Series Econometrics*, редакторы Helmut Lütkepohl & Markus Krätzig:
  www.jmulti.de/data_atse.html
- *Introduction to Econometrics*, 3-е издание, автор G. S. Maddala:
  www.wiley.co.uk/maddala/datasets.html
- *The Econometric Modelling of Financial Time Series*, 3-е издание, авторы Terence C. Mills & Raphael N. Markellos:
  www.lboro.ac.uk/departments/ec/cup
- *Basic Econometrics*, 4-е издание, автор Damodar Gujarati:
  gretl.sourceforge.net/gujarati/index.html
  3-е издание:
  shazam.econ.ubc.ca/student/gujarati
- *Econometrics: A Modern Introduction*, автор Michael P. Murray:
  wps.aw.com/aw_murray_economtrcs_1/37/9551/2445250.cw/index.html
- *An Introduction to Classical Econometric Theory*, автор Paul Ruud:
  elsa.berkeley.edu/~ruud/cet/data.html
- *Introduction to Econometrics*, авторы James H. Stock & Mark W. Watson:
  wps.aw.com/aw_stock_ie_2/50/13016/3332253.cw/index.html
- *Using Econometrics: A Practical Guide*, 5-е издание, автор A. H. Studenmund:
  wps.aw.com/aw_studenmund_useecon_5/34/8912/2281679.cw/index.html
- *Analysis of Financial Time Series*, автор Ruey S. Tsay:
  faculty.chicagobooth.edu/ruey.tsay/teaching/fts
- *A Guide to Modern Econometrics*, автор Marno Verbeek:
  www.econ.kuleuven.be/gme/data.htm
- *Econometric Analysis of Cross Section and Panel Data*, автор Jeffrey Wooldridge:
  www.msu.edu/~ec/faculty/wooldridge/book2.htm
- *Introductory Econometrics: A Modern Approach*, 4-е издание, автор Jeffrey Wooldridge:
  websites.swlearning.com/cgi-wadsworth/course_products_wp.pl?fid=M20bI&product_isbn_issn=9780324581621&discipline_number=413

Данные, использованные в исследованиях

Некоторые известные эконометристы размещают на своих сайтах данные, которые использовали в своих статьях, иногда вместе со своими статьями. Вот ссылки на некоторых из них:

- Данные из статей Брюса Хансена (Университет Висконсина в Мэдисоне):
  www.ssc.wisc.edu/~bhansen/progs/progs.htm

- Данные из статей Марка Уотсона (Принстонский университет):
  www.princeton.edu/~mwatson/publi.html
- Данные из статей Джеймса Гамильтона (Университет Калифорнии в Сан-Диего):
  weber.ucsd.edu/~jhamilto/software.htm
- Данные по доходностям и структурным сдвигам от Кеннета Френча (Дартмаутский Колледж):
  mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
- Данные по доходностям акций от Уильяма Шверта (Рочестерский университет):
  schwert.ssb.rochester.edu/data.htm
- Макроэкономические данные в реальном времени от Нормана Свансона (Университет Рутгерс):
  econweb.rutgers.edu/nswanson/realtime.htm
- Макроэкономические временные ряды от Мэнфреда Гартнера:
  www.fgn.unisg.ch/eumacro/macrodata/dmtrxneu.htm
- Сверхвысокочастотные финансовые временные ряды от Роберта Энгеля:
  weber.ucsd.edu/~mbacci/engle/index_data.html
- Данные из статей и монографий Дэвида Хэндри:
  www.nuff.ox.ac.uk/users/hendry/hendry.html\#data

Некоторые журналы с сильной эмпирической направленностью требуют от авторов доступности данных (а иногда и программ). Благодаря этому появляется возможность повторить расчеты авторов и проверить их корректность. Таких журналов, впрочем, немного.

- Архив данных в *Journal of Applied Econometrics*:
  qed.econ.queensu.ca/jae
- Архив данных (FTP) в *Journal of Business and Economic Statistics*:
  ftp://www.amstat.org/JBES_View
- Архив данных в *Journal of Money, Credit, and Banking*:
  web.econ.ohio-state.edu/jmcb/IndexDataArchive.php

Онлайновые данные из различных источников

- Экономические временные ряды от Economagic.com:
  www.economagic.com
- Бесплатные данные FreeLunch от Moody’s Analytics:
  www.economy.com/freelunch/default.asp
- Образовательный сайт Biz/ed:
  www.bized.co.uk/dataserv/datahome.htm
- Микро-данные и временные ряды от Economics Web Institute:
  www.economicswebinstitute.org/ecdata.htm
- Библиотека временных рядов от Роба Хайндмана / Time Series Data Library:
  www.robjhyndman.com/TSDL
- Данные по M-competition (соревнование моделей прогнозирования):
  www.neural-forecasting.com/m-competitions.htm

Другие подборки данных и ссылок на данные

- Региональные экономические данные от EconData.Net:
econdata.net
- Статистические данные по общественным наукам от Калифорнийского университета в Сан-Диего:
3stages.org/idata
- Перечень ссылок портала Econometric Links:
www.feweb.vu.nl/econometriclinks/\#data
- Перечень ссылок портала Quantitative Macroeconomics and Real Business Cycle:
dge.repec.org/data.html
- Проект межотраслевого прогнозирования INFORUM от Мэрилендского университета:
inforumweb.umd.edu
- Ссылки на данные по переходным экономикам от CERGE-EI:
home.cerge-ei.cz/hanousek/transition_data
- Ссылки на статистические ресурсы по России, Восточной Европе и Центральной Азии от университета штата Мичиган / Online Statistical Resources for Russia, Eastern Europe and Central Asia – Michigan State University:
staff.lib.msu.edu/ticklet/staff/statistics.htm
- Ссылки от Эконометрической лаборатории Калифорнийского университета в Беркли / Econometrics Laboratory at the University of California, Berkeley:
elsa.berkeley.edu/eml/emldata.shtml
- Поисковик экономических ресурсов от Федерального резервного банка г. Сент-Луис / St. Louis Fed – Economic Resource Finder:
research.stlouisfed.org/research-links.html
- Ссылки на данные на сайте Resources for Economists on the Internet (RFE):
rfe.org/showCat.php?cat_id=2
- Статистические данные в Сети от библиотеки Мичиганского университета / University of Michigan Library – Statistical Resources on the Web:
www.lib.umich.edu/govdocs/statsnew.html
- Ссылки на экономические данные на сайте WebEc:
www.helsinki.fi/WebEc/framec8d.html
- Ссылки на данные по международным отношениям от Пола Хенсела / International Relations Data Site:
www.paulhensel.org/data.html

Список литературы

Анатольев, С. (2007). Обзор англоязычных учебников по эконометрике. *Квантиль* 3, 73–82.

Анатольев, С. (2008). Обзор англоязычных учебников по анализу временных рядов. *Квантиль* 5, 49–55.

Where to find data on the Web?

Stanislav Anatolyev

New Economic School, Moscow, Russia

Alexander Tsyplakov

Novosibirsk State University, Novosibirsk, Russia

We give thematically arranged lists of useful websites where an empirical economist can find data for research.

Задачи и решения

Задачи

Задача 6.1

Рассмотрим линейную модель, где среди регрессоров присутствует линейный тренд:

$$y_t = \alpha + \beta t + \varepsilon_t,$$

и где ε_t независимы и одинаково распределены согласно некоторому распределению $\mathcal{D}$ с нулевым средним и дисперсией σ^2. Объектом интереса является β.

1. Выпишите МНК-оценку β (назовем ее $\hat{\beta}$) в форме отклонений и найдите ее асимптотическое распределение.
2. Исследоваетель предлагает избавиться от тренда в регрессорах с помощью взятия первых разностей:
$$y_t - y_{t-1} = \beta + \varepsilon_t - \varepsilon_{t-1}$$
с последующим оцениванием β с помощью МНК. Выпишите эту оценку (назовем ее $\check{\beta}$) и найдите ее асимптотическое распределение.
3. Сравните оценки $\hat{\beta}$ и $\check{\beta}$ по асимптотической эффективности.

Задача 6.2

Пусть асимптотическое смещение второго порядка некоторой состоятельной асимптотически нормальной (с асимптотической дисперсией $V_{\hat{\theta}}$) оценки $\hat{\theta}$ скалярного параметра θ равно $B_{\hat{\theta}}$. Выведите асимптотическое смещение второго порядка для $g(\hat{\theta})$ как оценки $g(\theta)$, где $g(\cdot)$ – гладкая нелинейная функция.

Задача 6.3

Рассмотрим регрессию в матричной форме

$$\mathcal{Y} = \mathcal{X}\beta + \mathcal{E},$$

где регрессоры $\mathcal{X}$ коррелируют с ошибками $\mathcal{E}$, но эта корреляция слаба. Рассмотрим разложение $\mathcal{E}$ на проекцию на $\mathcal{X}$ и ей ортогональную компоненту $\mathcal{U}$:

$$\mathcal{E} = \mathcal{X}\pi + \mathcal{U}.$$

Предположим, что $\left(n^{-1}\mathcal{X}'\mathcal{X}, n^{-1/2}\mathcal{X}'\mathcal{U}\right) \xrightarrow{p} (Q, \xi)$, где $\xi \sim \mathcal{N}\left(0, \sigma_u^2 Q\right)$, и матрица Q полного ранга. Покажите, что в предположении о дрейфующем параметре $\pi = c/\sqrt{n}$, где n – размер выборки, а c фиксировано, МНК-оценка для β состоятельна и асимптотически нецентрировано нормальна, и выведите асимптотическое распределение статистика Вальда для тестирования системы линейных ограничений $R\beta = r$, где R имеет полный ранг q.

Решения

Решение 5.1

Могут ли две случайные величины быть некоррелированными безусловно, но коррелироваными условно на третьей? Могут ли две случайные величины быть коррелированными безусловно, но некоррелироваными условно на третьей?

Ответ: могут в обоих случаях.

Для демонстрации первого явления положим $x = zu$ и $y = v$, где z, u и v – случайные величины с нулевым средним, и z независима от (u, v), в то время как ковариация между u и v ненулевая. Тогда $\mathbb{C}[x, y|z] = z\mathbb{C}[u, v] \neq 0$, в то время как $\mathbb{C}[x, y] = \mathbb{E}[z]\mathbb{C}[u, v] = 0$.

Для демонстрации второго явления положим $x = z + u$ и $y = z + v$, где z, u и v – независимые случайные величины с нулевым средним. Тогда $\mathbb{C}[x, y|z] = \mathbb{C}[u, v] = 0$, в то время как $\mathbb{C}[x, y] = \mathbb{V}[z] \neq 0$.

Решение 5.2

Известно, что для простейшей авторегрессии с независимыми и одинаково распределенными инновациями

$$y_t = \rho y_{t-1} + \varepsilon_t$$

в случае единичного корня $\rho = 1$ МНК-оценка $\hat{\rho}$ для ρ состоятельна и имеет распределение Дики–Фуллера

$$T(\hat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 B(r)dB(r)}{\int_0^1 B(r)^2 dr},$$

где $B(r)$ – стандартный Винеровский процесс на $[0, 1]$. Пусть «по глупости», вместо того, чтобы регрессировать y_t на y_{t-1}, мы регрессируем y_{t-1} на y_t и тестируем гипотезу о единичном корне. Выведите асимптотическое распределение такой оценки при наличии единичного корня. Состоятельна ли она для единицы?

В обратной регрессии МНК-оценка $\tilde{\rho}$ равна

$$\tilde{\rho} = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=2}^T y_t^2} = \frac{\sum_{t=2}^T y_{t-1}^2 + \sum_{t=2}^T y_{t-1}\varepsilon_t}{\sum_{t=3}^{T+1} y_{t-1}^2},$$

так что

$$T(\tilde{\rho} - 1) = \frac{T^{-1}\sum_{t=2}^T y_{t-1}\varepsilon_t + T^{-1}y_1^2 - T^{-1}y_T^2}{T^{-2}\sum_{t=3}^{T+1} y_{t-1}^2} \xrightarrow{d} \frac{\int_0^1 B(r)dB(r) - B(1)^2}{\int_0^1 B(r)^2 dr} = -\frac{B(1)^2 + 1}{2\int_0^1 B(r)^2 dr},$$

поскольку

$$T^{-1}y_T^2 \xrightarrow{d} E[\varepsilon_t^2]B(1)^2.$$

Получили распределение, несколько отличающееся от распределения Дики–Фуллера. Ясно, что $\tilde{\rho}$ состоятельна для единицы.

Заметим, что для аналогичной регрессии со стационарным переменными

$$y_t = \rho x_t + u_t, \quad \mathbb{E}[u_t|x_t] = 0,$$

в то время как «прямая» МНК-оценка ρ состоятельна для единицы при $\rho = 1$, «обратная» МНК-оценка несостоятельна для единицы:

$$\tilde{\rho} = \frac{\sum_{t=2}^T y_t x_t}{\sum_{t=2}^T y_t^2} = \frac{T^{-1}\sum_{t=2}^T x_t(x_t + u_t)}{T^{-1}\sum_{t=2}^T (x_t + u_t)^2} \xrightarrow{p} \frac{\mathbb{E}[x_t^2]}{\mathbb{E}[x_t^2] + \mathbb{E}[u_t^2]} \neq 1.$$

Объясняется такое расхождение результатов следующим образом. В то время как в стационарном случае источником состоятельности МНК-оценки является некоррелированность ошибки с регрессором, в нестационарном случае эта коррелированность не имеет значения: источником состоятельности является разная скорость роста слагаемых в числителе и знаменателе МНК-оценки за вычетом истинного параметра.

Решение 5.3

Пусть скалярные случайные величины x и y имеют одно и то же математическое ожидание μ. Покажите, что тест Хаусмана на верность условия на моменты $\mathbb{E}[y]=\mu$ при верности условия на моменты $\mathbb{E}[x]=\mu$ асимптотически эквивалентен J-тесту на верность модели, состоящей из обоих условий на моменты. Каково интуитивное объяснение этого результата?

Для теста Хаусмана возьмем в качестве эффективной ОММ-оценку на основе системы из двух условий на моменты $\mathbb{E}[x-\mu]=\mathbb{E}[y-\mu]=0$, которая, как нетрудно вывести, равна $\hat{\mu}_0=\hat{\alpha}\bar{x}+(1-\hat{\alpha})\bar{y}$, где $\hat{\alpha}$ – оценка некоторой константы. В качестве второй возьмем ММ-оценку на основе одного условия на моменты $\mathbb{E}[x-\mu]=0$, то есть $\hat{\mu}_1=\bar{x}$. Заметим, что $\hat{\mu}_1-\hat{\mu}_0=\bar{x}-\hat{\alpha}\bar{x}-(1-\hat{\alpha})\bar{y}=(1-\hat{\alpha})(\bar{x}-\bar{y})$. Таким образом, тест Хаусмана основан на разнице $(\bar{x}-\bar{y})^2$, но и J-тест основан на ней же. Нормализирующие коэффициенты, конечно же, должны иметь один и тот же предел по вероятности.

Интуитивно, одно из условий на моменты (неважно какое) является лишь определением для μ, так что и J-тест, и тест Хаусмана проверяют соответствие второго условия на моменты этому определению.

Статьи: эконометрическая теория

Смещение второго порядка в статистике, оценивающей качество прогнозов*

Виктор Китов†

Московский государственный университет, Москва, Россия

В статье выводится асимптотическое смещение второго порядка для статистики, оценивающей качество вневыборочных прогнозов параметрических моделей. В предыдущих исследованиях признавалось, что качество асимптотической аппроксимации первого порядка может оказаться неудовлетворительным. Найденное в данной статье асимптотическое смещение второго порядка позволяет во многом объяснить причину недостаточной точности асимптотической аппроксимации. Симуляции подтверждают полученные аналитические результаты.
Ключевые слова: критерии качества прогнозов, сравнение моделей, оценивание параметров, асимптотическое смещение второго порядка, проверка гипотез
Классификация JEL: C22, C52, C53

1 Введение

В данной работе рассматривается задача оценивания качества вневыборочных прогнозов. Пусть имеется выборка из наблюдений $z_1, z_2, \ldots, z_T$, где $z_t = (x_t, y_t) \in \mathbb{R}^{\mathtt{1x(n+1)}}$, $x_t \in \mathbb{R}^{\mathtt{1xn}}$, $y_t \in \mathbb{R}$, $t = 1, 2, \ldots, T$, которая подчиняется модели $y_t = g(x_t, \theta) + \varepsilon_t$, где θ – вектор неизвестных параметров, оцениваемый по располагаемой выборке. Качество прогнозов модели определяется математическим ожиданием $\mathbb{E}[f_t]$ некоторой функции $f_t = f(x_t, y_t, \theta)$. Типичными примерами функции f_t являются $\varepsilon_t, \varepsilon_t^2$ и ε_t^2/y_t^2, когда оценивается среднее ошибок прогнозов, среднее квадратов ошибок прогнозов и относительное среднее квадратов ошибок прогнозов. Другими примерами критерия качества прогнозов могут быть серийная ковариация ошибок прогнозов или ковариация ошибок прогнозов с прогнозом, полученным при помощи другой модели. Точечные оценки качества прогнозов обычно строятся с использованием выборочного аналога математического ожидания

$$\frac{1}{P}\sum_{t=R}^{T-1} f\left(x_{t+1}, y_{t+1}, \hat{\theta}_t\right) = \frac{1}{P}\sum_{t=R}^{T-1} \hat{f}_{t+1},$$

где параметр R выбирается исследователем, $P = T - R$, а $\hat{\theta}_t$ обозначает оценку неизвестного параметра θ. Для тестирования гипотез требуется знать асимптотическое распределение оценки качества прогнозов, которое будет зависеть от методики оценки вектора неизвестных параметров. В работах West (1996) и West & McCracken (1998) выводится (см. также Clements, 2005 и Маккракен, 2006) асимптотическое распределение первого порядка величины

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\hat{f}_{t+1} - \mathbb{E}[f_t]\right),$$

*Цитировать как: Китов, Виктор (2009). «Смещение второго порядка в статистике, оценивающей качество прогнозов», Квантиль, №6, стр. 77–91. Citation: Kitov, Victor (2009). "Second order bias in a forecast evaluation statistic," Quantile, No.6, pp. 77–91.
†Адрес: 119992, Россия, г. Москва, Воробьевы горы. Электронная почта: *vkitov@mail.ru*

которое является нормальным с нулевым средним.

Однако статистические испытания, приводимые авторами, свидетельствуют о неточности полученного асимптотического распределения, что выражается в отклонении фактической значимости тестов от теоретической. Величина данного отклонения зависит от метода оценивания вектора неизвестных параметров. В случае, когда все предыдущие наблюдения участвуют в оценивании неизвестных параметров (метод расширяющегося окна), отклонения в значимости тестов небольшие. Однако когда для оценивания параметров используется фиксированное число предшествующих моменту прогноза наблюдений (метод скользящего окна), отклонения в значимости тестов являются существенными, и они тем больше, чем больше отношение P/R.

В качестве решения проблемы неточной аппроксимации асимптотического распределения первого порядка в данной работе выводится асимптотическое разложение второго порядка для статистики качества прогнозов. В качестве метода оценивания параметров модели используется экстремальное оценивание общего вида, которое включает нелинейный метод наименьший квадратов и метод максимального правдоподобия как частные случаи. Учет асимптотического смещения второго порядка позволяет существенно повысить точность асимптотической аппроксимации распределения статистики, оценивающей качество прогнозов, особенно в случае оценивания θ методом скользящего окна. В работе приводятся статистические испытания, иллюстрирующие полученные аналитические результаты на численных примерах.

2 Постановка задачи

Рассматривается выборка наблюдений $z_1, z_2, \ldots, z_T$, где $z_t \in \mathbb{R}^{\mathtt{1x(n+1)}}$, $t = 1, 2, \ldots, T$, полученная с помощью некоторого строго стационарного и эргодического случайного процесса, и порождаемое данным случайным процессом вероятностное пространство с борелевской сигма-алгеброй. Каждое наблюдение делится на вектор регрессоров $x_t \in \mathbb{R}^{\mathtt{1xn}}$ и зависимую скалярную переменную y_t, так что $z_t = (x_t, y_t)$. Зависимость в данных описывается моделью

$$y_t = g(x_t, \theta) + \varepsilon_t,$$

где $\theta = \left(\theta^1, \theta^2, \ldots, \theta^m\right)'$ – вектор неизвестных параметров, оцениваемый по располагаемой выборке.

Целью исследования является построение статистических выводов о математическом ожидании $\mathbb{E}[f_t] \in \mathbb{R}$, где $f_t = f(x_t, y_t, \theta)$ – это некоторая функция, определяющая качество прогнозов. Для этого выбирается параметр R. Затем последовательно строятся прогнозы зависимой переменной y_t для моментов времени $t = R+1, R+2, \ldots, T$ с использованием модели $y_{t+1} = g(x_{t+1}, \hat{\theta}_t)$, где $\hat{\theta}_t$ оценивается двумя возможными способами: методом расширяющегося окна, в котором учитываются все предшествующие наблюдения $z_1, z_2, \ldots, z_t$, либо методом скользящего окна, где используются R последних наблюдений $z_{t-R+1}, z_{t-R+2}, \ldots, z_t$. Предполагается, что параметр θ удовлетворяет условию экстремального оценивания общего вида:

$$\theta = \arg\max_{q\in\Theta} \mathbb{E}\left[\psi(z, q)\right]$$

с некоторой функцией ψ. Оценка находится из равенства

$$\hat{\theta}_t = \arg\max_{q\in\Theta} \hat{\mathbb{E}}_t\left[\psi(z, q)\right],$$

где $\hat{\mathbb{E}}_t$ – оператор выборочного среднего. Экстремальное оценивание включает в себя нелинейный метод наименьших квадратов при $\psi(z, q) = -\left(y - g(x, q)\right)^2$ и метод максимального правдоподобия при $\psi(z,q) = \ln p(y|x,q)$, где $p(y|x,q)$ – функция плотности условного распределения.

Введем обозначения:

$$J = \mathbb{E}\left[\frac{\partial^2\psi(z_t,\theta)}{\partial q \partial q'}\right],\ K = \mathbb{E}\left[\frac{\partial^3\psi(z_t,\theta)}{\partial q^3}\right],\ h_\tau = \frac{\partial\psi(z_\tau,\theta)}{\partial q},\ m_\tau = \frac{\partial^2\psi(z_\tau,\theta)}{\partial q \partial q'} - J$$

и $B = -J^{-1}$. В лемме 1, приведенной в приложении, выводится разложение второго порядка для оценки, полученной с помощью метода экстремального оценивания. В случае расширяющегося окна разложение имеет вид

$$\hat{\theta}_t - \theta = B\frac{1}{\sqrt{t}}H_t + \frac{1}{t}BM_tBH_t + \frac{1}{2}\frac{1}{t}BH_t'BKBH_t + O_p\left(\frac{1}{t\sqrt{t}}\right), \tag{1}$$

где

$$H_t = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t} h_\tau, \quad M_t = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t} m_\tau.$$

В случае скользящего окна

$$\hat{\theta}_t - \theta = B\frac{1}{\sqrt{R}}H_t + \frac{1}{R}BM_tBH_t + \frac{1}{2}\frac{1}{R}BH_t'BKBH_t + O_p\left(\frac{1}{R\sqrt{R}}\right), \tag{2}$$

где

$$H_t = \frac{1}{\sqrt{R}}\sum_{\tau=t-R+1}^{t} h_\tau, \quad M_t = \frac{1}{\sqrt{R}}\sum_{\tau=t-R+1}^{t} m_\tau.$$

Для построения статистических выводов относительно величины $\mathbb{E}[f_t]$ строится статистика качества прогнозов

$$\hat{V} = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(f_{t+1}(x_{t+1}, y_{t+1}, \hat{\theta}_t) - \mathbb{E}[f_t]\right).$$

Рассматривается асимптотический переход, в котором при $T \to \infty$ значения $R \to \infty$ и $P \to \infty$, причем $P/R = \pi + o(1/\sqrt{R}), \quad 0 < \pi < \infty$.

Для краткости будут использоваться следующие обозначения:

- $\sup_t$ обозначает $\sup_{R \le t \le T-1}$: $f_{t+1} = f(x_{t+1}, y_{t+1}, \theta)$, $\hat{f}_{t+1} = f(x_{t+1}, y_{t+1}, \hat{\theta}_t)$.
- Модуль матрицы $|U|$ обозначает матрицу A, такую что $\{A\}_{i,j} = |\{U\}_{i,j}|$.
- Для матрицы $K \in \mathbb{R}^{\mathtt{mxmxm}}$, имеющей три измерения, и вектора $b \in \mathbb{R}^{\mathtt{mx1}}$ матричная операция $b'Kb$ обозначает вектор размерности $m \times 1$ с элементами

$$\{b'Kb\}_i = \sum_{j=1}^{m}\sum_{k=1}^{m} b_j b_k K_{i,j,k}.$$

- Оператор $\operatorname{vec}[\cdot]$ обозначает операцию векторизации. Для матрицы $A \in \mathbb{R}^{\mathtt{mxn}}$, где $a_{ij} = A_{i,j}$, $\operatorname{vec}[A] = [a_{11}, a_{21}, \ldots a_{m1}, a_{12}, a_{22}, \ldots a_{m2}, \cdots, a_{1n}, a_{2n}, \ldots a_{mn}]$.
- Операция $\otimes$ обозначает кронекерово произведение.

Далее,

$$\frac{\partial f_{t+1}}{\partial \theta}=\left(\frac{\partial f_{t+1}}{\partial \theta^1},\frac{\partial f_{t+1}}{\partial \theta^2},\dots\frac{\partial f_{t+1}}{\partial \theta^m}\right)\in\mathbb{R}^{\mathtt{1xm}},\quad \frac{\partial^2 f_{t+1}}{\partial \theta^2}=\left\{\frac{\partial^2 f_{t+1}}{\partial \theta^i\partial\theta^j}\right\}_{i,j}\in\mathbb{R}^{\mathtt{mxm}},$$

$$F=\mathbb{E}[f_t]\in\mathbb{R},\quad F_1=\mathbb{E}\left[\frac{\partial f_t}{\partial \theta}\right]\in\mathbb{R}^{\mathtt{1xm}},\quad F_2=\mathbb{E}\left[\frac{\partial^2 f_t}{\partial \theta^2}\right]\in\mathbb{R}^{\mathtt{mxm}},$$

$$G_t=\mathrm{vec}\left[\frac{\partial\psi\left(z_t,\theta\right)}{\partial q};\frac{\partial^2\psi\left(z_t,\theta\right)}{\partial q\partial q'};\ f(z_t;\theta)-F;\ \frac{\partial f}{\partial\theta}\left(z_t;\theta\right)-F_1;\ \frac{\partial^2 f}{\partial\theta^2}\left(z_t;\theta\right)-F_2\right].$$

Предполагается, что для всех $u\in\Theta$ (Θ – компакт, содержащий значение θ) выполнены условия:
(1) Функция $f_t\left(u\right)$ измерима и трижды непрерывно дифференцируема по u.
(2) Для некоторого $\delta>0\quad\mathbb{E}\left[\left|\mathrm{vec}\left[\otimes_{i=1}^3 G_t\right]\right|^{2+\delta}\right]<\infty$.
(3) Процесс G_t является строго перемешивающим, т.е. для сигма-алгебр $S_a^b=\sigma\left\{G_a,\dots,G_b\right\}$ выполнено условие

$$\sup_{A\in S_{-\infty}^t,\ B\in S_{t+\tau}^{+\infty}}\left|\mathbb{P}\left(AB\right)-\mathbb{P}\left(A\right)\mathbb{P}\left(B\right)\right|=\alpha\left(\tau\right)\to0 \text{ при } \tau\to\infty,$$

Дополнительно предполагается, что $\exists\tilde{C}:\alpha\left(\tau\right)\le\tilde{C}\tau^{-\omega},\ \frac{\omega\delta}{2+\delta}>2$.
(4) $\exists C_0<\infty, t_0>0:\forall t\ge t_0$

$$\sup_{\substack{\theta\in\Theta\\1\le i,j,k\le m}}\left|\frac{\partial^3 f_{t+1}(u)}{\partial\theta^i\partial\theta^j\partial\theta^k}\right|<v_t,$$

где v_t – некоторая измеримая функция, для которой выполнено условие $\mathbb{E}[v_t]<C_0$.

В работе предполагаются выполненными также технические предположения, обеспечивающие состоятельность оценок вектора θ, приведенные в приложении.

3 Смещение второго порядка статистики качества прогнозов

Введем обозначения:

$$\Phi_{t+1}^1=\left(\frac{\partial f_{t+1}}{\partial\theta}-F_1\right),\quad\Phi_{t+1}^2=\left(\frac{\partial^2 f_{t+1}}{\partial\theta^2}-F_2\right),\quad\Gamma_j=\mathbb{E}\left[\Phi_t^1 Bh_{t-j}\right],$$

$$\Omega_j=\mathbb{E}\left[m_t' Bh_{t-j}\right],\quad\Lambda_j=\mathbb{E}\left[h_t' BKBh_{t-j}\right],\quad\Psi_j=\mathbb{E}\left[h_t BF_2 Bh_{t-j}\right].$$

Указанные математические ожидания существуют согласно предположению 2 и неравенству Гельдера.

Теорема 1: При указанных выше предположениях смещение второго порядка для статистики, оценивающей качество прогнозов, равно

$$b_{st}=\frac{1}{\sqrt{P}}\lambda_B\left(\sum_{j=1}^{+\infty}\Gamma_j+F_1B\sum_{j=-\infty}^{+\infty}\Omega_j+\frac{1}{2}F_1B\sum_{j=-\infty}^{+\infty}\Lambda_j+\frac{1}{2}\sum_{j=-\infty}^{+\infty}\Psi_j\right)+o\left(\frac{1}{\sqrt{R}}\right),$$

где параметр λ_B равен $\ln\left(1+\pi\right)$ в случае расширяющегося окна и π в случае скользящего окна.

Доказательство: Рассмотрим случай расширяющегося окна. Применяя разложение Тэйлора второго порядка к каждой функции $\hat{f}_{t+1} = f(x_{t+1}, y_{t+1}, \hat{\theta}_t)$ и суммируя для моментов времени $R \le t \le T-1$, получим

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\hat{f}_{t+1} = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} f_{t+1} + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{\partial f_{t+1}}{\partial\theta}(\hat{\theta}_t-\theta) + \frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\hat{\theta}_t-\theta)'\frac{\partial^2 f_{t+1}}{\partial\theta^2}(\hat{\theta}_t-\theta)$$

плюс остаток

$$\frac{1}{6}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\sum_{i,j,k=1}^{m}\frac{\partial^3 f(z_{t+1},\tilde{\theta}_t)}{\partial\theta^i\partial\theta^j\partial\theta^k}\left(\hat{\theta}_t^i-\theta^i\right)\left(\hat{\theta}_t^j-\theta^j\right)\left(\hat{\theta}_t^k-\theta^k\right).$$

Согласно лемме 2, остаток равен $o_p\left(1/\sqrt{R}\right)$, значит

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\hat{f}_{t+1}-\mathbb{E}[f_{t+1}]\right) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(f_{t+1}-\mathbb{E}[f_{t+1}]) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{\partial f_{t+1}}{\partial\theta}-F_1\right)(\hat{\theta}_t-\theta)$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F_1(\hat{\theta}_t-\theta) + \frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\hat{\theta}_t-\theta)'\left(\frac{\partial^2 f_{t+1}}{\partial\theta^2}-F_2\right)(\hat{\theta}_t-\theta)$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\hat{\theta}_t-\theta)'F_2(\hat{\theta}_t-\theta) + o_p\left(\frac{1}{\sqrt{R}}\right).$$

Используя разложение (1), получим

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\hat{f}_{t+1}-\mathbb{E}[f_{t+1}]\right) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(f_{t+1}-\mathbb{E}[f_{t+1}]) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\Phi_{t+1}^1(\hat{\theta}_t-\theta)$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F_1(\hat{\theta}_t-\theta) + \frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\hat{\theta}_t-\theta)'\Phi_{t+1}^2(\hat{\theta}_t-\theta)$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\hat{\theta}_t-\theta)'F_2(\hat{\theta}_t-\theta) + o_p\left(\frac{1}{\sqrt{R}}\right) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(f_{t+1}-\mathbb{E}[f_{t+1}])$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\Phi_{t+1}^1\left(B\frac{1}{\sqrt{t}}H_t + \frac{1}{t}BM_tBH_t + \frac{1}{2}\frac{1}{t}BH_t'BKBH_t + O_p\left(\frac{1}{t\sqrt{t}}\right)\right)$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F_1\left(B\frac{1}{\sqrt{t}}H_t + \frac{1}{t}BM_tBH_t + \frac{1}{2}\frac{1}{t}BH_t'BKBH_t + O_p\left(\frac{1}{t\sqrt{t}}\right)\right)$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(B\frac{1}{\sqrt{t}}H_t + O_p\left(\frac{1}{t}\right)\right)'\Phi_{t+1}^2\left(B\frac{1}{\sqrt{t}}H_t + O_p\left(\frac{1}{t}\right)\right)$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(B\frac{1}{\sqrt{t}}H_t + O_p\left(\frac{1}{t}\right)\right)'F_2\left(B\frac{1}{\sqrt{t}}H_t + O_p\left(\frac{1}{t}\right)\right) + o_p\left(\frac{1}{\sqrt{R}}\right).$$

Исключая слагаемые порядка $o_p(1/\sqrt{R})$ в соответствии с леммами 4 и 6 и неравенством Чебышева, имеем:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\hat{f}_{t+1}-\mathbb{E}[f_{t+1}]\right)=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(f_{t+1}-\mathbb{E}[f_{t+1}])+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\Phi^1_{t+1}B\frac{1}{\sqrt{t}}H_t$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F_1B\frac{1}{\sqrt{t}}H_t+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}F_1BM_tBH_t+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}F_1BH_t'BKBH_t$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{\sqrt{t}}H_t\right)'BF_2B\left(\frac{1}{\sqrt{t}}H_t\right)+o_p\left(\frac{1}{\sqrt{R}}\right)=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(f_{t+1}-\mathbb{E}[f_{t+1}])$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{\sqrt{t}}\Phi^1_{t+1}BH_t-\mathbb{E}\left[\frac{1}{\sqrt{t}}\Phi^1_{t+1}BH_t\right]\right)+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F_1B\frac{1}{\sqrt{t}}H_t$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}F_1BM_tBH_t-\mathbb{E}\left[\frac{1}{t}F_1BM_tBH_t\right]\right)$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}F_1BH_t'BKBH_t-\mathbb{E}\left[\frac{1}{t}F_1BH_t'BKBH_t\right]\right)$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\left(\frac{1}{\sqrt{t}}H_t\right)'BF_2B\left(\frac{1}{\sqrt{t}}H_t\right)-\mathbb{E}\left[\left(\frac{1}{\sqrt{t}}H_t\right)'BF_2B\left(\frac{1}{\sqrt{t}}H_t\right)\right]\right)$$

$$+b_{st}+o_p\left(\frac{1}{\sqrt{R}}\right),$$

где

$$b_{st}=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\mathbb{E}\left[\Phi^1_{t+1}B\sqrt{t}H_t\right]+\frac{1}{\sqrt{P}}F_1B\sum_{t=R}^{T-1}\frac{1}{t}\mathbb{E}\left[M_tBH_t\right]$$

$$+\frac{1}{2}\frac{1}{\sqrt{P}}F_1B\sum_{t=R}^{T-1}\frac{1}{t}\mathbb{E}\left[H_t'BKBH_t\right]+\frac{1}{2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\mathbb{E}\left[H_t'BF_2BH_t\right].$$

Используя леммы 3 и 5, выражение для смещения можно переписать в требуемом виде:

$$b_{st}=\frac{\ln(1+\pi)}{\sqrt{P}}\left(\sum_{j=1}^{+\infty}\Gamma_j+F_1B\sum_{j=-\infty}^{+\infty}\Omega_j+\frac{1}{2}F_1B\sum_{j=-\infty}^{+\infty}\Lambda_j+\frac{1}{2}\sum_{j=-\infty}^{+\infty}\Psi_j\right)+o\left(\frac{1}{\sqrt{R}}\right),\qquad(3)$$

где

$$\Gamma_j=\mathbb{E}\left[\Phi^1_tBh_{t-j}\right],\ \Omega_j=\mathbb{E}\left[m_t'Bh_{t-j}\right],\ \Lambda_j=\mathbb{E}\left[h_t'BKBh_{t-j}\right],\ \Psi_j=\mathbb{E}\left[h_tBF_2Bh_{t-j}\right].$$

Доказательство утверждения для случая скользящего окна полностью повторяет приведенное доказательство для расширяющегося окна со следующими очевидными изменениями:

все суммы вида $\sum_{t=R}^{T-1}\frac{1}{\sqrt{t}}$ и $\sum_{t=R}^{T-1}\frac{1}{t}$ заменяются суммами $\sum_{t=R}^{T-1}\frac{1}{\sqrt{R}}$ и $\sum_{t=R}^{T-1}\frac{1}{R}$. Различия в мультипликативном сомножителе λ_B обусловлены тем, что в случае расширяющегося окна

$$\sum_{t=R}^{T-1}\frac{1}{t}=\ln(1+\pi)+o\left(\frac{1}{\sqrt{R}}\right),$$

а в случае скользящего окна

$$\sum_{t=R}^{T-1}\frac{1}{R}=\frac{P}{R}=\pi+o\left(\frac{1}{\sqrt{R}}\right).$$

При практическом тестировании гипотез используется t-статистика

$$\Sigma(\hat{\theta}_T)^{-\frac{1}{2}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\hat{f}_{t+1}-\mathbb{E}[f_{t+1}]\right),$$

где дисперсия $\Sigma(\theta)$ приближается величиной $\Sigma(\hat{\theta}_T)$. В следующей теореме приводится выражение для смещения второго порядка для t-статистики.

Теорема 2: При предположениях, указанных в постановке задачи, смещение второго порядка для t-статистики равно

$$b_{t.st}=\frac{1}{\Sigma^{1/2}}b_{st}-\frac{1}{2\Sigma^{3/2}}\frac{1}{\sqrt{P}}\frac{\pi}{1+\pi}\left(\sum_{j=-\infty}^{+\infty}\phi_j+\sum_{j=-\infty}^{+\infty}\varphi_j\right)+o\left(\frac{1}{\sqrt{R}}\right), \tag{4}$$

где b_{st} – асимптотическое смещение второго порядка для знаменателя t-статистики, найденной в теореме 1,

$$\phi_j=\mathbb{E}\left[(f_t-\mathbb{E}[f_t])\,h'_{t-j}\right]B\frac{\partial\Sigma'}{\partial\theta},\quad \varphi_j=F_1B\mathbb{E}\left[h_th'_{t-j}\right]B\frac{\partial\Sigma'}{\partial\theta},$$

$$\Sigma=\Sigma(\theta)=S_{ff}+\lambda_{fh}\left[F_1BS'_{fh}+S_{fh}BF'_1\right]+\lambda_{hh}F_1BS_{hh}BF'_1+o_p\left(\frac{1}{\sqrt{R}}\right),$$

$$S_{ff}=\sum_{j=-\infty}^{+\infty}\mathbb{E}\left[(f_t-\mathbb{E}[f_t])\,(f_{t-j}-\mathbb{E}[f_t])'\right],\ S_{fh}=\sum_{j=-\infty}^{+\infty}\mathbb{E}\left[(f_t-\mathbb{E}[f_t])\,h'_{t-j}\right],$$

$$S_{hh}=\sum_{j=-\infty}^{+\infty}\mathbb{E}\left[h_th'_{t-j}\right],\ \frac{\partial\Sigma}{\partial\theta}\in\mathbb{R}^{1\mathrm{xm}}.$$

Параметры λ_{fh} и λ_{hh} зависят от типа использованного окна при оценивании параметров: для расширяющегося окна $\lambda_{fh}=1-\frac{1}{\pi}\ln(1+\pi)$ и $\lambda_{hh}=2\left[1-\frac{1}{\pi}\ln(1+\pi)\right]$, для скользящего окна при $\pi\le 1$, $\lambda_{fh}=\frac{\pi}{2}$ и $\lambda_{hh}=\pi-\frac{\pi^2}{3}$, при $\pi>1$, $\lambda_{fh}=1-\frac{1}{2\pi}$ и $\lambda_{hh}=1-\frac{1}{3\pi}$.

Замечание: Поскольку $b_{st}=O(1/\sqrt{R})$, формулу (4) можно применить, когда истинные значения параметров в ней аппроксимируются их оценками, поскольку ошибка аппроксимации будет иметь порядок $o(1/\sqrt{R})$.

Доказательство: Доказательство приведено лишь для случая расширяющегося окна, поскольку для случая скользящего окна доказательство аналогично. Формула для ковариационной матрицы непосредственно следует из леммы 7. Обозначим

$$W_1 = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(f_{t+1} - \mathbb{E}[f_{t+1}]\right), \ W_2 = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} F_1 B \frac{1}{\sqrt{t}} H_t,$$

$$W_3 = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\frac{1}{\sqrt{t}} \Phi^1_{t+1} B H_t - \mathbb{E}\left[\frac{1}{\sqrt{t}} \Phi^1_{t+1} B H_t \right] \right),$$

$$W_4 = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\frac{1}{t} F_1 B M_t B H_t - \mathbb{E}\left[\frac{1}{t} F_1 B M_t B H_t \right] \right),$$

$$W_5 = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\frac{1}{t} F_1 B H_t' B K B H_t - \mathbb{E}\left[\frac{1}{t} F_1 B H_t' B K B H_t \right] \right),$$

$$W_6 = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(\frac{1}{\sqrt{t}} H_t \right)' B F_2 B \left(\frac{1}{\sqrt{t}} H_t \right) - \mathbb{E}\left[\left(\frac{1}{\sqrt{t}} H_t \right)' B F_2 B \left(\frac{1}{\sqrt{t}} H_t \right) \right] \right).$$

$$\xi = W_1 + W_2, \quad \nu = \sqrt{R} \left(W_3 + W_4 + \frac{1}{2} W_5 + \frac{1}{2} W_6 \right), \quad b = \sqrt{R} b_{st}.$$

Заметим, что $b/\sqrt{R} = O(1/\sqrt{R})$ из формулы (3), а $\nu/\sqrt{R} = o(1)$, поскольку данное слагаемое не участвует в асимптотическом разложении порядка $O(1)$, полученном в работе West & McCracken (1998). Выражение для t-статистики можно переписать в виде

$$\Sigma(\hat{\theta})^{-\frac{1}{2}} \left(\xi + \frac{\nu}{\sqrt{R}} + \frac{b}{\sqrt{R}} + o_p\left(\frac{1}{\sqrt{R}} \right) \right) = \left(\Sigma^{1/2} + \frac{1}{2\Sigma^{1/2}} \frac{\partial \Sigma}{\partial \theta} + o_p\left(\frac{1}{\sqrt{R}} \right) \right)^{-1}$$

$$\times \left(\xi + \frac{\nu}{\sqrt{R}} + \frac{b}{\sqrt{R}} + o_p\left(\frac{1}{\sqrt{R}} \right) \right) = \frac{1}{\Sigma^{1/2}} \left(1 + \frac{1}{2\Sigma} \frac{\partial \Sigma}{\partial \theta} + o_p\left(\frac{1}{\sqrt{R}} \right) \right)^{-1}$$

$$\times \left(\xi + \frac{\nu}{\sqrt{R}} + \frac{b}{\sqrt{R}} + o_p\left(\frac{1}{\sqrt{R}} \right) \right) = \frac{\xi}{\Sigma^{1/2}} + \frac{1}{\sqrt{R}} \frac{\nu}{\Sigma^{1/2}} + \frac{1}{\sqrt{R}} \frac{b}{\Sigma^{1/2}}$$

$$- \left(\frac{1}{2\Sigma^{3/2}} \xi \frac{\partial \Sigma}{\partial \theta} (\hat{\theta}_T - \theta) - \mathbb{E}\left[\frac{1}{2\Sigma^{3/2}} \xi \frac{\partial \Sigma}{\partial \theta} (\hat{\theta}_T - \theta) \right] \right) - \frac{1}{2\Sigma^{3/2}} \mathbb{E}\left[\xi \frac{\partial \Sigma}{\partial \theta} (\hat{\theta}_T - \theta) \right] + o\left(\frac{1}{\sqrt{R}} \right).$$

По лемме 8

$$\mathbb{E}\left[\xi \left(\frac{\partial \Sigma}{\partial \theta} (\hat{\theta}_T - \theta) \right)' \right] = \frac{1}{\sqrt{P}} \frac{\pi}{1+\pi} \left(\sum_{j=-\infty}^{+\infty} \phi_j + \sum_{j=-\infty}^{+\infty} \varphi_j \right) + o\left(\frac{1}{\sqrt{R}} \right),$$

где $\phi_j = \mathbb{E}[(f_t - \mathbb{E}[f_t])\, h_{t-j}'] B (\partial\Sigma/\partial\theta)'$ и $\varphi_j = F_1 B \mathbb{E}[h_t h_{t-j}'] B (\partial\Sigma/\partial\theta)'$, $\partial\Sigma/\partial\theta \in \mathbb{R}^{1\text{xm}}$.

Из полученной формулы следует, что смещение второго порядка для разных типов окон, использующихся при оценивании параметров, различается только коэффициентами λ_B, λ_{fh} и λ_{hh}. Поскольку λ_{fh} и λ_{hh} принадлежат интервалу $(0, 2)$, существенная разница в точности асимпотической аппроксимации первого порядка различными методами может произойти лишь вследствие разницы в параметре λ_B. Данный параметр больше, и с ростом π он растет быстрее, в случае скользящего окна, что свидетельствует о том, что при больших значениях отношения P/R асимптотическая аппроксимация первого порядка статистики, использующей скользящее окно, будет неточной. Данный вывод согласуется с результатами статистических испытаний, приведенных в работе West & McCracken (1998), а также с нижеприведенными статистическими испытаниями.

4 Статистические испытания

Рассмотрим простейшую модель авторегрессии первого порядка

$$y_t = \beta y_{t-1} + \varepsilon_t,$$

в которой ошибки предполагаются независимыми и одинаково распределенными нормальными величинами $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$. Параметр $\sigma_\varepsilon^2 = 1$ считается известным, а $\beta = 1/2$ оценивается методом наименьших квадратов. Данная модель повторяет модель численного эксперимента, приведенного в работе West & McCracken (1998). Рассмотрим три теста, в которых будет проверяться гипотеза о том, что математическое ожидание некоторой функции качества прогнозов $\mathbb{E}[f_t]$ равно своему теоретическому значению D:

$$H_0: \ \mathbb{E}[f_t] = D.$$

При этом будут различаться три разновидности каждого теста в зависимости от альтернативной гипотезы:

Тест 1 $H_A: \ \mathbb{E}[f_t] \neq D.$
Тест 2 $H_A: \ \mathbb{E}[f_t] < D.$
Тест 3 $H_A: \ \mathbb{E}[f_t] > D.$

Качество асимптотической аппроксимации будет определяться путем сравнения фактической значимости тестов, полученной при помощи 50000 испытаний Монте–Карло, с теоретическим значением, равным 5%. В приведенных примерах оказалось, что $\sum_{j=-\infty}^{+\infty} \left(\phi_j + \varphi_j\right) \equiv 0$, поэтому лишь первое слагаемое в формуле (4) влияет на смещение второго порядка. Ниже приведены тестируемые гипотезы и отвечающие им функции качества прогнозов, теоретические значения их математических ожиданий, а также аналитические значения для смещения и дисперсии соответственно.

Тест A: серийная ковариация ошибок: $H_0 : \mathbb{E}[\varepsilon_t \varepsilon_{t-1}] = 0,\ f_t = \varepsilon_t \varepsilon_{t-1},\ D = 0,\ b_{t.st} = 2\frac{1}{\Sigma^{1/2}}\frac{1}{\sqrt{P}}\lambda_B \sigma_\varepsilon^2 \hat{\beta},\ \Sigma = \sigma_\varepsilon^4 + -2\sigma_\varepsilon^4(1-\hat{\beta}^2)\lambda_{fh} + \sigma_\varepsilon^4(1-\hat{\beta}^2)\lambda_{hh}.$

Тест B: ожидаемый квадрат ошибки прогноза: $H_0 : \mathbb{E}[\varepsilon_t^2] = \sigma_e^2,\ f_t = \varepsilon_t^2,\ D = \sigma_e^2,\ b_{t.st} = 2\frac{1}{\Sigma^{1/2}}\frac{1}{\sqrt{P}}\lambda_B \sigma_\varepsilon^2,\ \Sigma = 2\sigma_\varepsilon^4.$

Тест C: связь ошибок прогноза с другими величинами: $H_0 : \mathbb{E}[\varepsilon_t y_{t-2}] = 0,\ f_t = \varepsilon_t y_{t-2},\ D = 0,\ b_{t.st} = -\frac{1}{\Sigma^{1/2}}\frac{1}{\sqrt{P}}\lambda_B \sigma_\varepsilon^2,\ \Sigma = \sigma_\varepsilon^4(1+\hat{\beta}^2(\lambda_{hh} - 2\lambda_{fh}))/(1-\hat{\beta}^2).$

На Рис. 1–6 показаны фактические значимости рассматриваемых тестов при использовании скользящего и расширяющегося окна. Оси абсцисс соответствуют значения параметра R при $P = 200 - R$, а оси ординат соответствует вероятность ошибки первого рода в процентах. Каждая гипотеза при каждой альтернативе тестируется с использованием статистик $\hat{V}/\hat{S}_{ff}^{1/2}$, $\hat{V}/\hat{\Sigma}^{1/2}$, $\hat{V}/\hat{\Sigma}^{1/2} - \hat{b}_{t.st}$ и их асимптотического (нормального) распределения. Переменные пунктиры соответствуют статистике $\hat{V}/\hat{S}_{ff}^{1/2}$, длинные пунктиры – статистике $\hat{V}/\hat{\Sigma}^{1/2}$, сплошные линии – статистике $\hat{V}/\hat{\Sigma}^{1/2} - \hat{b}_{t.st}$. В первой t-статистике не учитывается влияние оценивания неизвестных параметров θ на асимптотическое распределение, вторая t-статистика строится в соответствии с асимптотическими результатами из статей West (1996) и West & McCracken (1998), а третья статистика представляет собой вторую t-статистику, скорректированную на оценку ее асимптотического смещения второго порядка.

В тестах B1, B2 и B3 при скользящем окне и во всех тестах при расширяющемся окне значимости для статистик $\hat{V}/\hat{S}_{ff}^{1/2}$ и $\hat{V}/\hat{\Sigma}^{1/2}$ совпадают, поскольку в этих случаях $S_{ff} \equiv \Sigma$,

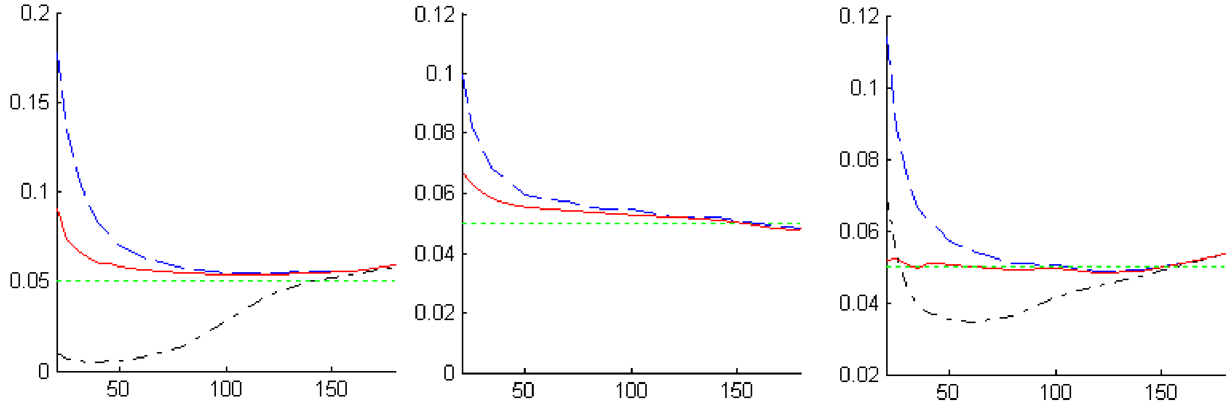

Рис. 1: Уровень значимости при скользящем окне наблюдений для тестов A1, B1 и C1.

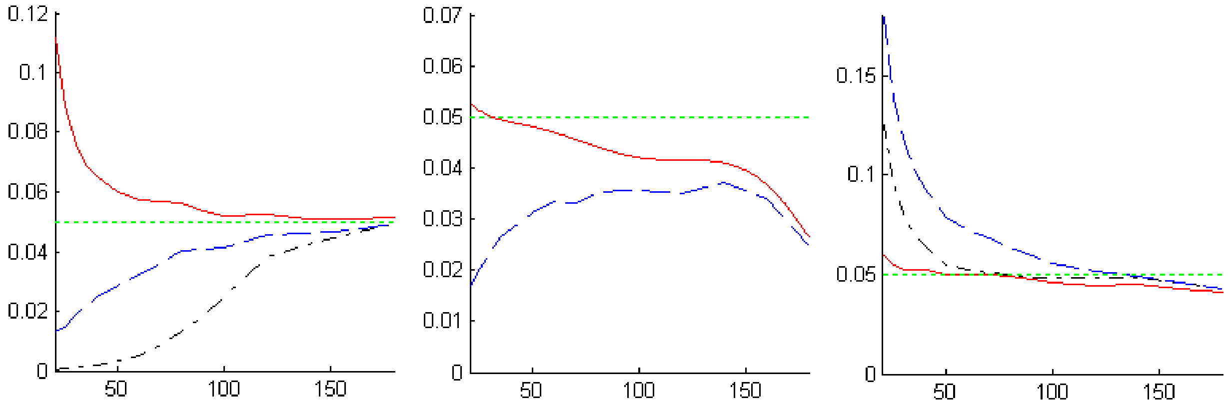

Рис. 2: Уровень значимости при скользящем окне наблюдений для тестов A2, B2 и C2.

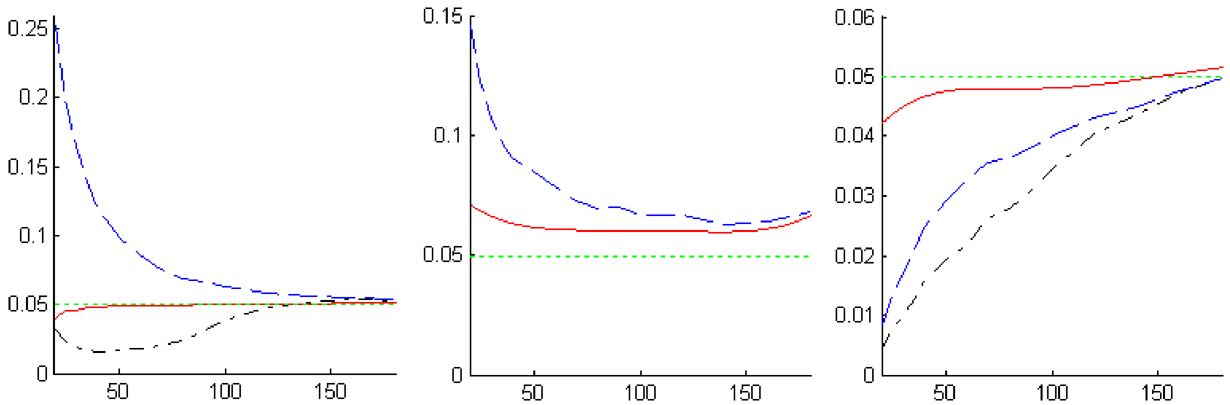

Рис. 3: Уровень значимости при скользящем окне наблюдений для тестов A3, B3 и C3.

и t-статистики оказываются эквивалентными. Из рисунков 1–6 видно, что за исключением ситуаций, в которых $S_{ff} \equiv \Sigma$, значимость двустороннего теста, полученного с помощью обычной t-статистики $\hat{V}/\hat{S}_{ff}^{1/2}$, ниже теоретической значимости, что отражает несостоятельность данного вида тестирования. Фактическая значимость приближается к теоретической лишь когда R достаточно велико, поскольку в этом случае разница между θ и $\hat{\theta}_t$ становится несущественной. Значимость тестов, полученных с помощью t-статистики $\hat{V}/\hat{\Sigma}^{1/2}$, в большей степени соответствует теоретической значимости, равной 5%. Отклонение от 5% тем больше, чем выше отношение P/R, что объясняется большим асимптотическим смеще-

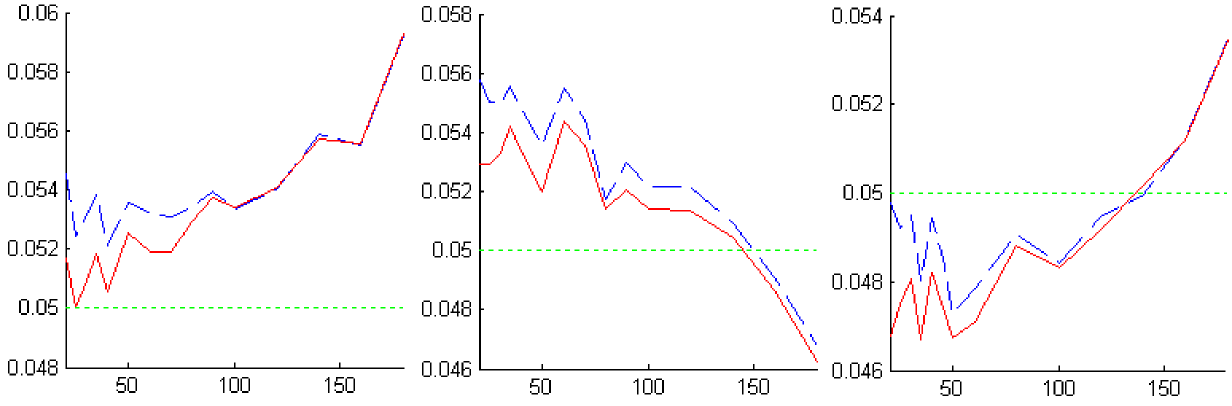

Рис. 4: Уровень значимости при расширяющемся окне наблюдений для тестов A1, B1 и C1.

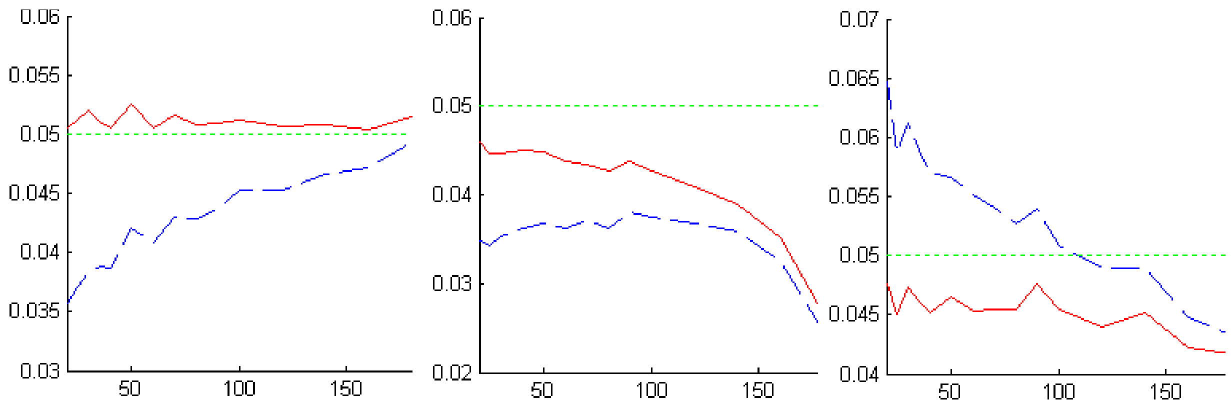

Рис. 5: Уровень значимости при расширяющемся окне наблюдений для тестов A2, B2 и C2.

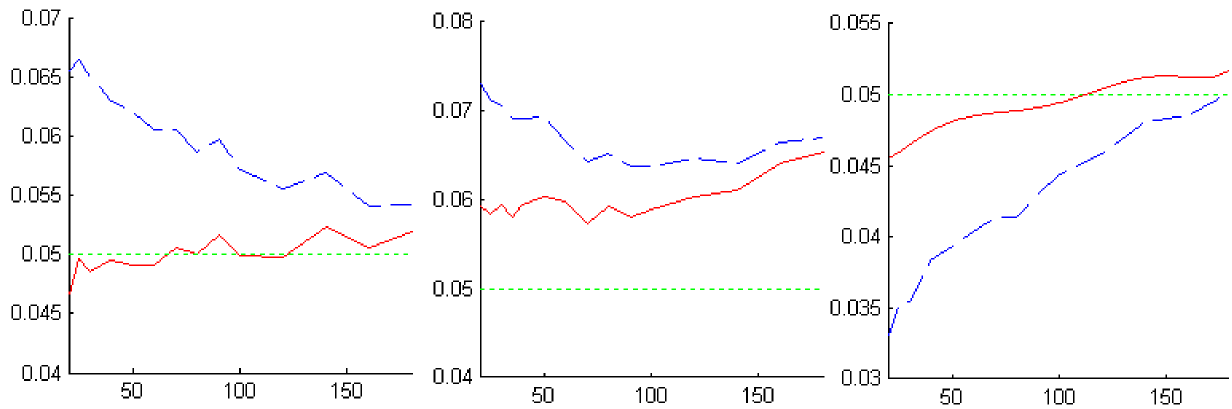

Рис. 6: Уровень значимости при расширяющемся окне наблюдений для тестов A3, B3 и C3.

нием статистики ввиду формулы (4), которое не учитывается в данном виде тестирования. В случае использования расширяющегося окна учет смещения второго порядка не оказывает существенного влияния на соответствие фактической и теоретической значимости для двухстороннего теста 1. Однако учет смещения позволяет существенно улучшить соответствие теоретической и фактической значимости для односторонних тестов 2 и 3, что объясняется большей чувствительностью односторонних тестов к смещенности распределения статистики. В случае использования скользящего окна корректировкой смещения удается существенно улучшить соответствие между теоретической и фактической значимостью как для односторонних, так и для двухсторонних тестов, особенно для случаев, когда отношение P/R велико, что подтверждает аналитические результаты, полученные в статье.

5 Заключение

В работе исследовалась проблема неточной аппроксимации распределения статистики, оценивающей качество прогнозов, посредством его асимптотического распределения первого порядка, полученного в работах West (1996) и West & McCracken (1998). Было получено асимптотическое разложение второго порядка указанной статистики, из которого следует, что статистика обладает ненулевым смещением второго порядка. От использования расширяющегося или же скользящего окна при оценивании неизвестных параметров в формуле смещения второго порядка зависят лишь три коэффициента. Из вида данных коэффициентов следует, что смещение в случае расширяющегося окна принимает существенно меньшие значения, чем в случае скользящего окна, особенно когда отношение P/R велико. Полученные аналитические результаты подтверждаются статистическими испытаниями в работе West & McCracken (1998) и численными примерами данной работы. Учет полученного смещения позволяет значительно повысить точность асимптотической аппроксимации распределения статистики, оценивающей качество прогнозов, в случае использования скользящего окна.

Благодарности

Автор выражает благодарность профессору Российской экономической школы Анатольеву Станиславу за важные комментарии и советы в ходе исследования.

Список литературы

Ибрагимов, И.А. & Ю.В. Линник (1965). *Независимые и стационарно связанные величины.* Ленинград: Наука.

Маккракен, М. (2006). Парные тесты на одинаковую точность прогнозов. *Квантиль* 1, 53–62.

Clements, M.P. (2005). *Evaluating Econometric Forecasts of Economic and Financial Variables.* Palgrave Macmillan.

Manski, C.F. (1988). *Analog Estimation Methods in Econometrics.* Chapman & Hall.

West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

West, K.D. & M. McCracken (1998). Regression-based tests of predictive ability. *International Economic Review* 39, 817–40.

Приложение

В работе используется метод экстремального оценивания, согласно которому истинное значение неизвестного параметра удовлетворяет условию

$$\theta = \arg\max_{q\in\Theta} \mathbb{E}\left[\psi\left(z,q\right)\right].$$

Поскольку вероятностная мера, по которой производится усреднение, неизвестна, для приближенного нахождения параметра θ используется метод аналогий, при котором операция математического ожидания заменяется усреднением по наблюдениям выборки. Существуют различные варианты дополнительных предположений, гарантирующих состоятельность оценки $\hat{\theta}_t$, полученной методом аналогий. Возможными дополнительными ограничениями, согласно теореме 2’, приведенной в главе 7.2.2 книги Manski (1988), являются:

1. Множество Θ является компактом.
2. Решение задачи $\max_{q\in\Theta}\mathbb{E}\left[\psi\left(z,q\right)\right]$ единственно.
3. Существует такая функция $D(z)$, что $\mathbb{E}[D(z)]<\infty$ и $|\psi(z,q)|<D(z)\ \forall(z,q)\in(Z,\Theta)$, где Z – область значений переменной z.

Лемма 1: Предположим, функция $\psi\left(z,q\right)$ измерима и три раза непрерывно дифференцируема в области Θ. а также предположим, что выполнено условие равномерной интегрируемости: $\mathbb{E}[|\psi(z,q)|\,|\,|\psi(z,q)|>K]\to 0$ равномерно по $q\in\Theta$. Тогда при условии состоятельности оценки $\hat{\theta}_t$ (ограничения 1–3) имеют место следующие результаты:
(a): При использовании расширяющегося окна наблюдений

$$\hat{\theta}_t=\arg\max_{q\in\Theta}\frac{1}{t}\sum_{\tau=1}^{t}\psi\left(z_\tau,q\right),$$

выполнено

$$\hat{\theta}_t-\theta=B\frac{1}{\sqrt{t}}H_t+\frac{1}{t}BM_tBH_t+\frac{1}{2}\frac{1}{t}BH_t'BKBH_t+O_p\left(\frac{1}{t\sqrt{t}}\right),$$

где

$$H_t=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}h_\tau,\quad M_t=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}m_\tau.$$

(b): При использовании скользящего окна наблюдений

$$\hat{\theta}_t=\arg\max_{q\in\Theta}\frac{1}{R}\sum_{\tau=t-R+1}^{t}\psi\left(z_\tau,q\right),$$

выполнено

$$\hat{\theta}_t-\theta=B\frac{1}{\sqrt{R}}H_t+\frac{1}{R}BM_tBH_t+\frac{1}{2}\frac{1}{R}BH_t'BKBH_t+O_p\left(\frac{1}{R\sqrt{R}}\right),$$

где

$$H_t=\frac{1}{\sqrt{R}}\sum_{\tau=t-R+1}^{t}h_\tau,\quad M_t=\frac{1}{\sqrt{R}}\sum_{\tau=t-R+1}^{t}m_\tau.$$

Доказательство: Утверждения (a) и (b) похожи, поэтому будет доказано лишь утверждение (a). Так как θ доставляет максимум функционалу $\mathbb{E}\left[\psi\left(z,q\right)\right]$, то, в силу предположений о дифференцируемости и равномерной интегрируемости, θ является решением уравнения

$$\mathbb{E}\left[\frac{\partial\psi\left(z,\theta\right)}{\partial q}\right]=0.$$

Выборочным аналогом данного уравнения является

$$\frac{1}{t}\sum_{\tau=1}^{t}\frac{\partial\psi(z_\tau,\hat{\theta})}{\partial q}=0.$$

Выборочный аналог состоятельно оценивает θ согласно предположению.

Разложение Тэйлора первого порядка вокруг истинного параметра θ принимает вид

$$0=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}+\frac{1}{t}\sum_{\tau=1}^{t}\frac{\partial^2\psi(z_\tau,\tilde{\theta})}{\partial q\partial q'}\sqrt{t}\left(\hat{\theta}_t-\theta\right),$$

где $\tilde{\theta}$ лежит между θ и $\hat{\theta}_t$. Отсюда

$$\sqrt{t}\left(\hat{\theta}_t-\theta\right)=-\left(\frac{1}{t}\sum_{\tau=1}^{t}\frac{\partial^2\psi(z_\tau,\tilde{\theta})}{\partial q\partial q'}\right)^{-1}\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}$$

По теореме Биркоффа–Хинчина

$$\sqrt{t}\left(\hat{\theta}_t-\theta\right)=-J^{-1}\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}+O_p\left(\frac{1}{\sqrt{t}}\right).$$

Осуществляя разложение Тэйлора второго порядка относительно θ, получим:

$$0=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}+\frac{1}{t}\sum_{\tau=1}^{t}\frac{\partial^2\psi\left(z_\tau,\theta\right)}{\partial q\partial q'}\sqrt{t}\left(\hat{\theta}_t-\theta\right)$$

$$+\frac{1}{2}\frac{1}{t\sqrt{t}}\sum_{\tau=1}^{t}\left(\sqrt{t}(\hat{\theta}_t-\theta)\right)'\frac{\partial^3\psi\left(z_\tau,\theta\right)}{\partial q^3}\left(\sqrt{t}(\hat{\theta}_t-\theta)\right)+O_p\left(\frac{1}{t}\right)$$

$$=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}+\frac{1}{\sqrt{t}}\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\left(\frac{\partial^2\psi\left(z_\tau,\theta\right)}{\partial q\partial q'}-J\right)\sqrt{t}(\hat{\theta}_t-\theta)+J\sqrt{t}(\hat{\theta}_t-\theta)$$

$$+\frac{1}{2}\frac{1}{t\sqrt{t}}\sum_{\tau=1}^{t}\left(\sqrt{t}(\hat{\theta}_t-\theta)\right)'\frac{\partial^3\psi\left(z_\tau,\theta\right)}{\partial q^3}\left(\sqrt{t}(\hat{\theta}_t-\theta)\right)+O_p\left(\frac{1}{t}\right).$$

Подставляя разложение Тэйлора первого порядка для $\sqrt{t}\left(\hat{\theta}_t-\theta\right)$, имеем

$$\hat{\theta}_t-\theta=-J^{-1}\frac{1}{\sqrt{t}}\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}+\frac{1}{t}J^{-1}\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\left(\frac{\partial^2\psi\left(z_\tau,\theta\right)}{\partial q\partial q'}-J\right)J^{-1}\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}$$

$$-\frac{1}{2}\frac{1}{t}J^{-1}\left(\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}\right)'J^{-1}KJ^{-1}\left(\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q}\right)+O_p\left(\frac{1}{t\sqrt{t}}\right).$$

Таким образом,

$$\hat{\theta}_t-\theta=-J^{-1}\frac{1}{\sqrt{t}}H_t+\frac{1}{t}J^{-1}M_tJ^{-1}H_t-\frac{1}{2}\frac{1}{t}J^{-1}H_t'J^{-1}KJ^{-1}H_t+O_p\left(\frac{1}{t\sqrt{t}}\right),$$

где

$$H_t=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\frac{\partial\psi\left(z_\tau,\theta\right)}{\partial q},\quad m_\tau=\frac{\partial^2\psi\left(z_\tau,\theta\right)}{\partial q\partial q'}-J,\quad M_t=\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\left(\frac{\partial^2\psi\left(z_\tau,\theta\right)}{\partial q\partial q'}-J\right).$$

Остальные леммы приводятся без доказательств. Доказательства используют монографию Ибрагимов & Линник (1965).

Лемма 2: Справедлив результат

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\sum_{i,j,k=1}^{m}\frac{\partial^3 f(z_{t+1},\tilde{\theta}_t)}{\partial\theta^i\partial\theta^j\partial\theta^k}\left(\hat{\theta}_t^i-\theta^i\right)\left(\hat{\theta}_t^j-\theta^j\right)\left(\hat{\theta}_t^k-\theta^k\right)=o_p\left(\frac{1}{\sqrt{R}}\right).$$

Лемма 3: $\mathbb{E}\left[\Phi_{t+1}^1B\sqrt{t}H_t\right]=\sum_{j=1}^{+\infty}\Gamma_j+o\left(1\right)$, $\mathbb{E}\left[M_tBH_t\right]=\sum_{j=-\infty}^{+\infty}\Omega_j+o\left(1\right)$, $\mathbb{E}\left[H_t'BKBH_t\right]=\sum_{j=-\infty}^{+\infty}\Lambda_j+o\left(1\right)$, $\mathbb{E}\left[H_t'BF_2BH_t\right]=\sum_{j=-\infty}^{+\infty}\Psi_j+o\left(1\right)$.

Лемма 4: Пусть существует такая константа K, что $\mathbb{E}[|\eta_t|] < K$ при $t \geq 1$. Тогда $\forall a > 1$

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \frac{1}{t^a} \eta_t = o_p\left(\frac{1}{\sqrt{R}}\right).$$

Лемма 5: Справедлив результат

$$a_K = \frac{1}{K} + \frac{1}{K+1} + \ldots + \frac{1}{R+P-1} = \ln\left(\frac{T}{K}\right) + O\left(\frac{1}{K}\right).$$

Лемма 6: Пусть случайные величины ξ_t, μ_t, η_t и a_t, b_t, c_t являются элементами вектора G_t (указанного в постановке задачи), и пусть $d_{i,j} = \mathbb{E}\left[\xi_t \mu_{t+i} \eta_{t+j}\right]$. Тогда

$$\mathbb{E}\left[\sum_{j=1}^{t} a_j \sum_{j=1}^{t} b_j\right], \mathbb{E}\left[\sum_{j=1}^{t} a_j \sum_{j=1}^{t} b_j \sum_{j=1}^{t} c_j\right] = O(t),\ \mathbb{E}\left[a_1 \sum_{j=1}^{t} b_j \sum_{j=1}^{t} c_j\right] = O(1).$$

Лемма 7: Пусть

$$\Sigma(\theta) = \mathbb{V}\left[\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\hat{f}_{t+1} - \mathbb{E} f_{t+1}\right)\right]$$

обозначает асимптотическую дисперсию статистики, оценивающей качество прогнозов. Тогда

$$\Sigma(\theta) = S_{ff} + \lambda_{fh}\left(F_1 B S'_{fh} + S_{fh} B F'_1\right) + \lambda_{hh} F_1 B S_{hh} B F'_1 + o\left(\frac{1}{\sqrt{R}}\right),$$

где $S_{ff} = \sum_{j=-\infty}^{+\infty} \mathbb{E}[(f_t - \mathbb{E}[f_t])(f_{t-j} - \mathbb{E}[f_{t-j}])']$, $S_{fh} = \sum_{j=-\infty}^{+\infty} \mathbb{E}[(f_t - \mathbb{E}[f_t]) h'_{t-j}]$, $S_{hh} = \sum_{j=-\infty}^{+\infty} \mathbb{E}[h_t h'_{t-j}]$, а параметры λ_{fh}, λ_{hh} принимают следующие значения для различных типов окон: для расширяющегося окна $\lambda_{fh} = 1 - \frac{1}{\pi}\ln(1+\pi)$, $\lambda_{hh} = 2\left(1 - \frac{1}{\pi}\ln(1+\pi)\right)$, для скользящего окна при $\pi \leq 1$, $\lambda_{fh} = \frac{\pi}{2}$ и $\lambda_{hh} = \pi - \frac{\pi^2}{3}$, при $\pi > 1$, $\lambda_{fh} = 1 - \frac{1}{2\pi}$ и $\lambda_{hh} = 1 - \frac{1}{3\pi}$.

Лемма 8: Пусть $\xi = W_1 + W_2$ (используя обозначения теоремы 2), $\partial\Sigma/\partial\theta \in R^{1\mathrm{xm}}$ – вектор производных дисперсии статистики, оценивающей качество прогнозов. Тогда

$$\mathbb{E}\left[\xi\left(\frac{\partial\Sigma}{\partial\theta}(\hat{\theta}_T - \theta)\right)'\right] = \frac{1}{\sqrt{P}} \frac{\pi}{1+\pi}\left(\sum_{j=-\infty}^{+\infty} \phi_j + \sum_{j=-\infty}^{+\infty} \varphi_j\right) + o\left(\frac{1}{\sqrt{R}}\right),$$

где $\phi_j = \mathbb{E}[(f_t - \mathbb{E}[f_t]) h'_{t-j}] B (\partial\Sigma/\partial\theta)'$ и $\varphi_j = F_1 B \mathbb{E}[h_t h'_{t-j}] B (\partial\Sigma/\partial\theta)'$.

Second order bias in a forecast evaluation statistic

Victor Kitov

Moscow State University, Moscow, Russia

We derive the second order asymptotic bias for a statistic evaluating the quality of out-of-sample forecasts by parametric models. The existing literature admits that the quality of the first order asymptotic approximation may be unsatisfactory. We find that the second order asymptotic bias allows one to explain insufficient precision of the asymptotic approximation. Simulations confirm the obtained analytical results.

Keywords: forecast evaluation, model comparison, parameter estimation, second order asymptotic bias, hypothesis testing

JEL Classification: C22, C52, C53

Quantile

#6, March 2009

English page in the world wide web: http://quantile.ru/eng
Electronic mail address: quantile@quantile.ru
Access to the journal is free and unlimited

GUIDE TO AUTHORS

Manuscripts for publication in the “Articles” section should be submitted by electronic mail to the address *submit@quantile.ru*. Submitted work may be drawn from any applied field within the economics profession. The main requirement is correct usage of adequate econometric methodology. The manuscript should be written in Russian (for Russian-speaking persons) or in English (for all others) in the *Microsoft Word* or (preferably) *LaTeX* formats, and not exceed 30 double-spaced A4 pages. All submissions are subject to quality control by the editorial counsel and independent referees. A promising manuscript may be returned to the author(s) for polishing or rewriting. The editor also invites econometrics experts worldwide to contribute to the methodological sections of the journal.

Quantile

international econometric journal in Russian language

#6
March 2009

IN THIS ISSUE