

Эконометрический ликбез: бутстрап

Основы бутстрапирования*

Станислав Анатольев[†]

Российская экономическая школа, Москва, Россия

Настоящее эссе – введение в принципы и методологию бутстрапа. Даются основы бутстраповской инференции, ресэмплинга, асимптотического рафинирования. Изложение сопровождается поясняющими примерами. Приводятся краткий анонс методологических эссе данного выпуска журнала «Квантиль» и ссылки на неза- тронутый материал.

1 Приближение бутстрапом

Не секрет, что точная инференция¹ в современном эконометрическом анализе практически недостижима, если модель хоть сколь-нибудь замысловатая, а вовлеченные распределения не являются нормальными. Среди приближенных, конечно, лидирует асимптотический подход, т.е. когда при инференции используются приближения, свойственные теории больших выборок. Асимптотическая теория, правда, частенько тоже разочаровывает неточностью предоставляемых приближений или трудностью аналитического вывода асимптотических результатов (см. Анатольев, 2005). В этих случаях используют метод *бутстрапирования*, предоставляющий альтернативные асимптотическим приближения или возможность обходиться без сложных аналитических выводов.

Формально изобретение бутстрапа приписывают Эфрону (Efron, 1979, см. также Эфрон, 1988), хотя и ранее можно было встретить интуитивные применения бутстраповских идей. Например, в Cowles (1934) автор использовал колоду карт для генерации случайной подвыборки активов для последующего сравнения ее доходности с доходностью подвыборки, построенной по определенным правилам. Кстати, интенсивное использование при бутстрапировании симуляций (экспериментов Монте-Карло) является той причиной, по которой многие прикладные исследователи считают бутстрап методом Монте-Карло. На самом деле суть бутстрапа вовсе не в симуляциях.

В основе бутстраповского подхода лежит та идея, что истинное распределение данных можно хорошо приблизить эмпирическим, т.е. тем, как данные легли в выборке. На самом деле, эмпирическое распределение – единственный источник информации об истинном распределении данных, что у исследователя есть помимо модели. Идеологически бутстрап как раз и подразумевает это приближение истинного распределения эмпирическим; техническая же сторона дела – трансформировать приближение для распределения данных в приближенное распределение интересующих исследователя статистик, где симуляции и используются.

Идея того, что распределение самих данные используются вместо их истинного распределения нашла отражение в названии метода. «Бутстрапами» называются ремешки на обуви, ухватясь за которые, барон Мюнхаузен, по слухам, вытащил себя из болота. Не очень ясно,

*Работа основана на лекциях, читаемых автором в РЭШ. Автор благодарит Александра Цыплакова за полезные замечания. Цитировать как: Анатольев, Станислав (2007) «Основы бутстрапирования», Квантиль, №3, стр. 1–12. Citation: Anatolyev, Stanislav (2007) “The basics of bootstrapping,” *Quantile*, No.3, pp. 1–12.

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

¹Инференция включает в себя построение доверительных областей и проверку статистических гипотез.

правда, не за волосы ли себя он тащил, и почему в электронной версии книги не присутствуют ни болото (кроме эпизода охоты на какого-то зверя в трясине), ни бутстрап вообще.²

Итак, пусть из исходной популяции случайной величины z с распределением $F(z)$ получена выборка размера n . Тогда (кумулятивная) эмпирическая функция распределения (ЭФР)

$$F^*(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{[z_i \leq z]},$$

где $\mathbb{I}_{[\cdot]}$ – индикатор-функция, равномерно почти наверное стремится к $F(z)$ при $n \rightarrow \infty$, что составляет суть леммы Гливенко–Кантелли. Это свойство мотивирует использование бутстрапа и его благоприятные свойства.

Чтобы более наглядно пояснить бутстраповский метод, рассмотрим простейший пример. Пусть в выборке всего два наблюдения:

$$(x_1, y_1) = (1, 2); (x_2, y_2) = (1, 2).$$

Допустим, нас интересует коэффициент θ линейной проекции y на x без свободного члена, т.е. $y_i = \theta x_i + \varepsilon_i$, $\mathbb{E}[x_i \varepsilon_i] = 0$. В этом случае МНК-оценка параметра θ равна

$$\hat{\theta} = \frac{x_1 y_1 + x_2 y_2}{x_1^2 + x_2^2} = \frac{1 \times 2 + 2 \times 1}{1^2 + 2^2} = \frac{4}{5}.$$

Эмпирическая функция распределения данных представляет собой

$$(x, y) = \begin{cases} (1, 2) & \text{с вероятностью } 1/2, \\ (2, 1) & \text{с вероятностью } 1/2. \end{cases}$$

По отношению к этому распределению данные выборки из двух наблюдений распределены следующим образом:

$$(x_1^*, y_1^*); (x_2^*, y_2^*) = \begin{cases} (1, 2); (1, 2) & \text{с вероятностью } 1/4, \\ (2, 1); (2, 1) & \text{с вероятностью } 1/4, \\ (1, 2); (2, 1) & \text{с вероятностью } 1/4, \\ (2, 1); (1, 2) & \text{с вероятностью } 1/4. \end{cases}$$

Это распределение является бутстраповским распределением (бутстраповской) выборки. Соответственно, бутстраповская МНК-оценка распределена как

$$\hat{\theta}^* = \begin{cases} 1/2 & \text{с вероятностью } 1/4, \\ 4/5 & \text{с вероятностью } 1/2, \\ 2 & \text{с вероятностью } 1/4. \end{cases}$$

Это и есть бутстраповское распределение $\hat{\theta}$.

2 Приближение симуляциями

Пример, рассмотренный нами в предыдущем разделе, был чрезвычайно прост: размер исходной выборки был равен 2. В общем случае, когда в выборке n наблюдений, количество вариантов значений бутстраповской статистики имеет порядок n^n . Таким образом, в вычислительном плане задача сильно усложняется по мере роста n . К решению этой задачи можно привлечь компьютер, но не стоит ему поручать эту задачу в буквальном смысле. Даже если компьютер справится с этой сложной комбинаторной задачей за разумное время, полученный результат окажется для наших целей точным сверх необходимого. Гораздо выгодней

²Rudolph Erich Raspe “The Surprising Adventures of Baron Munchausen.” Электронная книга от Authorama: <http://www.authorama.com/book/adventures-of-baron-munchausen.html>.

воспользоваться дополнительным приближением, на этот раз с помощью симуляций. Эта идея прекрасна тем, что как раз из эмпирического распределения, присваивающего всем наблюдениям выборки равные веса, вытягивать «наблюдения» (*ресэмплировать*) просто и удобно.

Предположим, мы хотим пробутстрапировать статистику $\widehat{\varphi} = \widehat{\varphi}(\{z_1; \dots; z_n\})$. Выберем количество будущих бутстраповских выборок B (обычно хватает порядка 1000). Для каждого $b = 1, 2, \dots, B$ построим *бутстраповскую выборку* $\{z_1^*; z_2^*; \dots; z_n^*\}_b$, вытягивая ее элементы случайным образом с возвращением из исходной выборки $\{z_1; \dots; z_n\}$, и вычислим *бутстраповскую статистику* $\widehat{\varphi}_b^* = \widehat{\varphi}(\{z_1^*; \dots; z_n^*\}_b)$. Полученный набор бутстраповских статистик $\widehat{\varphi}_1^*, \dots, \widehat{\varphi}_B^*$ с приписанием каждой веса $1/B$ составляет (приближенное) бутстраповское распределение статистики $\widehat{\varphi}$. Чаще всего бутстраповское распределение используют для получения бутстраповских квантилей, для чего нужно лишь отсортировать бутстраповские статистики в порядке возрастания и в качестве квантилей $q_{\alpha_1}^*, q_{1-\alpha_2}^*$ взять значения $\widehat{\varphi}_{[B\alpha_1]}^*, \widehat{\varphi}_{[B(1-\alpha_2)+1]}^*$, где $[\cdot]$ означает взятие целой части.

Сгенерировать дискретную случайную величину, равномерно распределенную на множестве индексов от 1 до n просто: надо сначала сгенерировать случайную величину, распределенную равномерно на $[0, 1]$, затем присвоить значение индекса, соответствующее попаданию в отрезки $[0, 1/n], [1/n, 2/n], \dots, [1 - 1/n, 1]$. Например, в программе GAUSS запись `Z[ceil(n*randu(n,1)),.]` возвращает готовую матрицу с бутстраповской выборкой из строк матрицы Z .

Может возникнуть вопрос – а почему бы предварительно не сгладить «лесенку» ЭФР прежде чем ресэмплировать. Предложение сомнительное, ибо если полезность этого шага неочевидна, то процесс ресэмплинга технически сильно усложняется.

3 Рецентрирование

Вернемся к идеологии бутстрапа. При его реализации существуют некоторые тонкие моменты, игнорируя которые, можно легко все испортить. Один из таких тонких моментов – необходимость *рецентрирования*, когда речь идет о разностях (или расстояниях) между выборочными и популяционными объектами. Предположим, мы бутстрапируем разность $\widehat{\theta} - \theta$, где $\widehat{\theta}$ – аналоговая³ оценка параметра θ . При бутстрапировании $\widehat{\theta}$ мы создаем бутстраповские статистики, подсчитываемые по той же формуле, что и $\widehat{\theta}$, но на бутстраповских выборках. А вот бутстраповским аналогом параметра θ является не он сам, а первоначальная статистика $\widehat{\theta}$. Действительно, если при приближении бутстрапом истинное распределение искривляется в эмпирическое, то и истинный параметр искривляется в свою оценку. Итак, правильным бутстраповским аналогом разности $\widehat{\theta} - \theta$ является, таким образом, $\widehat{\theta}^* - \widehat{\theta}$, но ни в коем случае не $\widehat{\theta}^* - \theta$. Это и есть рецентрирование.

Даже имеющий верное представление о рецентрировании исследователь может ошибиться, действуя механически. Например, бутстрапируя обычную t -статистику

$$t = \frac{\widehat{\theta}}{\text{se}(\widehat{\theta})},$$

легко забыть о наличии в числителе разницы между $\widehat{\theta}$ и нулем, истинным значением параметра θ при нулевой гипотезе

$$H_0 : \theta = 0.$$

³Пусть интересующий нас параметр θ известным образом зависит от функции распределения z , $F(z)$. Согласно принципу аналогий, аналоговая оценка $\widehat{\theta}$ строится с помощью замены истинной функции распределения $F(z)$ на ее эмпирический аналог $F^*(z)$. Например, если нас интересует популяционное среднее $\mathbb{E}[z] = \int_{-\infty}^{+\infty} z dF(z)$, то его аналоговая оценка будет равна $\widehat{\mathbb{E}[z]} = \int_{-\infty}^{+\infty} z dF_n(z) = n^{-1} \sum_{i=1}^n z_i$.

Правильная версия бутстраповского аналога t-статистики есть, таким образом,

$$t^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{\text{se}^*(\widehat{\theta})}.$$

Причем точно так же выглядит бутстраповская t-статистика для гипотезы $H_0 : \theta = 13$, например, хотя сама t-статистика, разумеется, выглядит несколько по-другому.

Приведем другие, более сложные примеры рецентрирования, не акцентируясь на обозначениях, о значении которых читатель сам догадается. Положим, нас интересует нулевая гипотеза

$$H_0 : g(\theta) = 0.$$

Бутстраповский аналог статистики Вальда

$$\mathcal{W} = ng(\widehat{\theta})' \left(\widehat{G} \widehat{V}_\theta \widehat{G}' \right)^{-1} g(\widehat{\theta})$$

есть

$$\mathcal{W}^* = n \left(g(\widehat{\theta}^*) - g(\widehat{\theta}) \right)' \left(\widehat{G}^* \widehat{V}_\theta^* \widehat{G}^{*'} \right)^{-1} \left(g(\widehat{\theta}^*) - g(\widehat{\theta}) \right).$$

Бутстраповский аналог статистики отношения правдоподобия

$$\mathcal{LR} = 2 \left(\max_{\theta} \ell_n(\theta) - \max_{\theta: g(\theta)=0} \ell_n(\theta) \right)$$

есть

$$\mathcal{LR}^* = 2 \left(\max_{\theta} \ell_n^*(\theta) - \max_{\theta: g(\theta)=g(\widehat{\theta})} \ell_n^*(\theta) \right).$$

Бутстраповский аналог статистики разницы оптимумов, основанной на ОММ,

$$\mathcal{DD} = n \left[\min_{\theta: g(\theta)=0} \mathcal{Q}_n(\theta) - \min_{\theta} \mathcal{Q}_n(\theta) \right],$$

где $\mathcal{Q}_n(\theta)$ – целевая функция ОММ:

$$\mathcal{Q}_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n m(z_i, \theta) \right)' \widehat{\Sigma}^{-1} \left(\frac{1}{n} \sum_{i=1}^n m(z_i, \theta) \right),$$

и где $\widehat{\Sigma}$ состоятельно оценивает $\Sigma = \mathbb{E} [m(z, \theta) m(z, \theta)']$, есть

$$\mathcal{DD}^* = n \left[\min_{\theta: g(\theta)=g(\widehat{\theta})} \mathcal{Q}_n^*(\theta) - \min_{\theta} \mathcal{Q}_n^*(\theta) \right],$$

где $\mathcal{Q}_n^*(\theta)$ – бутстраповская целевая функция ОММ:

$$\mathcal{Q}_n^*(\theta) = \left(\frac{1}{n} \sum_{i=1}^n m(z_i^*, \theta) - \frac{1}{n} \sum_{i=1}^n m(z_i, \widehat{\theta}) \right)' \widehat{\Sigma}^{*-1} \left(\frac{1}{n} \sum_{i=1}^n m(z_i^*, \theta) - \frac{1}{n} \sum_{i=1}^n m(z_i, \widehat{\theta}) \right).$$

Заметим, что в последнем примере встретилось сразу два случая рецентрирования.

Отметим еще один момент. Может возникнуть вопрос, почему мы не рецентрируем, например, в знаменателе $\text{se}^*(\widehat{\theta})$ бутстраповской t-статистики или в бутстраповских пивотизирующих матрицах \widehat{V}_θ^* и $\widehat{\Sigma}^*$. Дело в том, что это непринципально, ибо это множители, от которых всего лишь требуется сходимость по вероятности к неслучайным пределам. В случае же, когда речь идет именно о расстояниях (разностях), рецентрирование принципиально.

В следующих двух разделах мы познакомимся с примерами использования бутстрапа.

4 Бутстраповская корректировка смещения

Сначала рассмотрим нетипичное применение бутстрапа, не связанное с инференцией, но помогающее еще глубже понять принципы его работы и использования. Бутстрап позволяет скорректировать смещение, связанное с конечностью выборки. Пусть у нас есть состоятельная оценка $\hat{\theta}$ параметра θ , построенная согласно принципу аналогий, причем $\mathbb{E}[\hat{\theta}] \neq \theta$. Смещение оценки $\hat{\theta}$ равно $\mathbb{B}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta$. Бутстраповским аналогом этого смещения является величина $\mathbb{B}^*[\hat{\theta}] = \mathbb{E}^*[\hat{\theta}^*] - \hat{\theta}$ (вновь видим рецентрирование!), где \mathbb{E}^* обозначает математическое ожидание по отношению к ЭФР. Если это бутстраповское смещение вычислить, то можно будет скорректировать исходную статистику на смещение:

$$\hat{\theta}_{BC} = \hat{\theta} - \mathbb{B}^*[\hat{\theta}] = 2\hat{\theta} - \mathbb{E}^*[\hat{\theta}^*].$$

Рассмотрим пару примеров.

Пусть случайная величина z имеет среднее μ . Исследователь оценивает μ с помощью \bar{z}_n , а μ^2 – с помощью \bar{z}_n^2 . Как выглядит бутстраповская корректировка смещения этих двух оценок? В первом случае, \bar{z}_n^* , бутстраповский аналог \bar{z}_n , имеет среднее \bar{z}_n по отношению к эмпирическому распределению: $\mathbb{E}^*[\bar{z}_n^*] = \bar{z}_n$. Поэтому бутстраповский аналог смещения (которое само по себе – ноль) есть $\mathbb{B}^*[\bar{z}_n] = \mathbb{E}^*[\bar{z}_n^*] - \bar{z}_n = 0$, так что оценка μ , скорректированная на смещение, есть $\bar{z}_n - \mathbb{B}^*[\bar{z}_n] = \bar{z}_n$, что логично. Во втором случае, смещение \bar{z}_n^2 равно

$$\mathbb{B}[\bar{z}_n^2] = \mathbb{E}[\bar{z}_n^2] - \mu^2 = \mathbb{V}[\bar{z}_n] = \frac{1}{n}\mathbb{V}[z],$$

поэтому бутстраповский аналог смещения – выборочный аналог этой величины:

$$\mathbb{B}^*[\bar{z}_n^2] = \frac{1}{n}\mathbb{V}^*[z] = \frac{1}{n} \left(\frac{1}{n} \sum_{i=1}^n z_i^2 - \bar{z}_n^2 \right),$$

так что оценкой μ^2 , скорректированной на смещение, будет

$$\bar{z}_n^2 - \mathbb{B}^*[\bar{z}_n^2] = \frac{n+1}{n}\bar{z}_n^2 - \frac{1}{n^2} \sum_{i=1}^n z_i^2.$$

В этих простых примерах удалось аналитически вывести бутстраповское смещение. В более типичных ситуациях это сделать невозможно, и на практике прибегают к приближению симуляциями, оценивая $\mathbb{E}^*[\hat{\theta}^*]$ с помощью усреднения по бутстраповским выборкам:

$$\hat{\theta}_{BC} = 2\hat{\theta} - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*.$$

Еще раз подчеркнем, что бутстрап способен удалить существенную часть смещения, вызванную конечностью выборки, но не способен справиться с асимптотическим смещением. Так что если оценка несостоятельна, бутстрап бессилён.

5 Бутстраповская инференция

Основное использование бутстрапа – это, конечно же, инференция. При этом, по сути, вся бутстраповская технология используется лишь для получения одного или двух бутстраповских критических значений, используя которые наряду с первоначальной оценкой и, возможно, стандартными ошибками или пивотизирующей матрицей, строят бутстраповские доверительные интервалы или проверяют гипотезы.

Рассмотрим несколько вариантов бутстраповских статистик, используемых для построения доверительных интервалов, и подчеркнем их положительные и отрицательные качества.

Пусть нас интересует построение статистических выводов относительно параметра θ на основе ее состоятельной оценки $\hat{\theta}$.

Пробутстрапируем статистику $\hat{\theta} - \theta$, бутстраповский аналог которой $\hat{\theta}^* - \hat{\theta}$. Получим бутстраповские квантили $q_{\alpha_1/2}^{*\%}$ и $q_{1-\alpha_2/2}^{*\%}$ и построим бутстраповский *процентильный* доверительный интервал с покрытием $1 - \alpha_1 - \alpha_2$

$$CI_{\%} = \left[\hat{\theta} - q_{1-\alpha_2/2}^{*\%}, \hat{\theta} - q_{\alpha_1/2}^{*\%} \right].$$

Заметим, что неправильно бутстрапировать нецентрированную статистику $\hat{\theta}$, ибо использующий квантили ее бутстраповского аналога $\hat{\theta}^*$ так называемый *эфронов* доверительный интервал плох (кроме случая, характеризуемого симметричным распределением), ибо, не используя рецентрирование, только усиливает смещение, присущее исходной выборке.

Обратим внимание, что процентильный доверительный интервал не требует подсчета стандартных ошибок, что удобно, особенно в случаях, когда их трудно вывести или подсчитать. Тем не менее, есть причины (их мы обсудим позже), по которым стоит бутстрапировать t -статистики, если стандартные ошибки возможно качественно подсчитать. Итак, пробутстрапируем статистику $(\hat{\theta} - \theta)/se(\hat{\theta})$, бутстраповский аналог которой $(\hat{\theta}^* - \hat{\theta})/se^*(\hat{\theta})$. Получим бутстраповские квантили $q_{\alpha_1/2}^{*\%t}$ и $q_{1-\alpha_2/2}^{*\%t}$ и построим бутстраповский *t-процентильный* доверительный интервал

$$CI_{\%t} = \left[\hat{\theta} - se(\hat{\theta})q_{1-\alpha_2/2}^{*\%t}, \hat{\theta} - se(\hat{\theta})q_{\alpha_1/2}^{*\%t} \right].$$

Далее, чаще всего в приложениях $\alpha_1 = \alpha_2 = \alpha/2$. В таком случае можно бутстрапировать не центрированную оценку или t -статистику, а их модули. Например, пробутстрапировав статистику $|\hat{\theta} - \theta|$, бутстраповский аналог которой равен $|\hat{\theta}^* - \hat{\theta}|$, получим бутстраповскую квантиль $q_{1-\alpha}^{*|\%|}$ и построим бутстраповский *симметричный процентильный* доверительный интервал с покрытием $1 - \alpha$

$$CI_{|\%|} = \left[\hat{\theta} - q_{1-\alpha}^{*|\%|}, \hat{\theta} + q_{1-\alpha}^{*|\%|} \right].$$

Аналогично, пробутстрапировав статистику $|\hat{\theta} - \theta|/se(\hat{\theta})$, бутстраповский аналог которой $|\hat{\theta}^* - \hat{\theta}|/se^*(\hat{\theta})$, получим бутстраповскую квантиль $q_{1-\alpha}^{*|\%t|}$ и построим бутстраповский *симметричный t-процентильный* доверительный интервал

$$CI_{|\%t|} = \left[\hat{\theta} - se(\hat{\theta})q_{1-\alpha}^{*|\%t|}, \hat{\theta} + se(\hat{\theta})q_{1-\alpha}^{*|\%t|} \right].$$

В следующем разделе мы узнаем о причинах, по которым лучше строить симметричные версии доверительных интервалов (разумеется, если выполнено $\alpha_1 = \alpha_2$), а также почему надо стремиться к бутстрапированию нормализованных разностей (t -статистик), а не просто разностей.

Теперь обсудим тестирование статистических гипотез при помощи бутстрапа. Пусть параметр скалярен, а нулевая гипотеза имеет простейший вид

$$H_0 : \theta = \theta^0,$$

где θ^0 – конкретная величина. Если альтернативная гипотеза двусторонняя, $H_A : \theta \neq \theta^0$, то, пробутстрапировав модуль t -статистики $|\hat{\theta} - \theta|/se(\hat{\theta})$, получим бутстраповское распределение $|\hat{\theta}^* - \hat{\theta}|/se^*(\hat{\theta})$ и ее квантиль $q_{1-\alpha}^{*|\%t|}$. Гипотеза H_0 отвергается, если $|\hat{\theta} - \theta^0|/se(\hat{\theta}) > q_{1-\alpha}^{*|\%t|}$. Если же альтернативная гипотеза односторонняя, например правосторонняя $H_A : \theta > \theta^0$, то приходится бутстрапировать t -процентильную статистику $(\hat{\theta} - \theta^0)/se(\hat{\theta})$. Получив бутстраповское распределение $(\hat{\theta}^* - \hat{\theta})/se^*(\hat{\theta})$ и соответствующую квантиль $q_{1-\alpha}^{*\%t}$, отвергаем гипотезу H_0 , если $(\hat{\theta} - \theta^0)/se(\hat{\theta}) > q_{1-\alpha}^{*\%t}$.

Если нулевая гипотеза имеет вид системы нелинейных ограничений $H_0 : g(\theta) = 0$, бутстрапируется статистика Вальда (см. раздел 3), и ее бутстраповская квантиль $q_{1-\alpha}^{*W}$ используется для принятия решения: если $\mathcal{W} > q_{1-\alpha}^{*W}$, гипотеза H_0 отвергается. Например, если система ограничений линейна,

$$H_0 : R\theta = r,$$

то бутстраповская статистика Вальда принимает вид $(\widehat{\theta}^* - \widehat{\theta})' R' (R\widehat{V}_{\widehat{\theta}}^* R')^{-1} R(\widehat{\theta}^* - \widehat{\theta})$, и гипотеза отвергается, если $(R\widehat{\theta} - r)' (R\widehat{V}_{\widehat{\theta}} R')^{-1} (R\widehat{\theta} - r) > q_{1-\alpha}^{*W}$.

Заметим, что при асимптотическом подходе и доверительные интервалы, и решающие правила при тестировании гипотез выглядят точно так же, как и при бутстрапировании, за единственным исключением – используются другие квантили. Таким образом, вся процедура бутстрапирования, включая симуляции, ресэмплинг, подсчет бутстраповских статистик и т.д., подчинена единственной цели – получить правильные бутстраповские квантили.

6 Асимптотическое рафинирование

Говорят, что с помощью бутстрапа можно достичь асимптотического рафинирования. В этом разделе мы обсудим, что такое асимптотическое рафинирование и в каких случаях оно имеет место.

Пусть у нас есть некоторая статистика $\widehat{\varphi}$, истинное распределение которой $F_{\widehat{\varphi}}(z)$. Обозначим бутстраповское распределение этой статистики через $F_{\widehat{\varphi}}^*(z)$. Говорят, что с помощью бутстрапа достигается асимптотическое рафинирование, если ошибка аппроксимации истинного распределения $F_{\widehat{\varphi}}(z)$ бутстраповским $F_{\widehat{\varphi}}^*(z)$ большего порядка малости, чем ошибка аппроксимации асимптотическим распределением при стремлении объема выборки к бесконечности.

Инструментарием здесь являются так называемое *разложение Эджворта* (кумулятивной) функции распределения статистики вокруг ее предельного распределения по степеням $1/\sqrt{n}$. Рассмотрим асимптотически pivotalную t-статистику, т.е.

$$\widehat{\varphi} = \frac{\widehat{\theta} - \theta}{\text{se}(\widehat{\theta})} \xrightarrow{d} N(0, 1).$$

Для (кумулятивной) функции стандартного нормального распределения используем обычное обозначение $\Phi(z)$. Разложения Эджворта истинного и бутстраповского распределений вокруг асимптотического выглядят следующим образом:

$$F_{\widehat{\varphi}}(z) = \Phi(z) + \frac{h_1(z, F)}{\sqrt{n}} + \frac{h_2(z, F)}{n} + O\left(\frac{1}{n\sqrt{n}}\right),$$

$$F_{\widehat{\varphi}}^*(z) = \Phi(z) + \frac{h_1(z, F^*)}{\sqrt{n}} + \frac{h_2(z, F^*)}{n} + O\left(\frac{1}{n\sqrt{n}}\right).$$

Здесь $h_1(z, F)$ – четная по z , непрерывная по F функция, $h_2(z, F)$ – нечетная по z , непрерывная по F функция. Ошибки аппроксимации точного распределения асимптотическим и бутстраповским, соответственно, равны

$$\Phi(z) - F_{\widehat{\varphi}}(z) = \frac{h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right),$$

$$F_{\widehat{\varphi}}^*(z) - F_{\widehat{\varphi}}(z) = \frac{h_1(z, F^*) - h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right).$$

Здесь мы воспользовались тем фактом, что скорость сходимости разности $h_1(z, F^*) - h_1(z, F)$ есть \sqrt{n} согласно уже упоминавшейся лемме Гливленко–Кантелли. Сравнивая порядки ошибок аппроксимации, делаем вывод, что в данном случае использование бутстрапа приводит к асимптотическому рафинированию. На практике это означает, что в достаточно больших выборках ошибка бутстраповского приближения, по идее, намного меньше, чем ошибка асимптотического приближения.

Рассмотрим теперь асимптотически непивотальную статистику

$$\widehat{\varphi} = \sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta),$$

где непивотальность проявляется в зависимости предельного распределения от неизвестного параметра V_θ . Как и в предыдущем случае, разложим точное и бутстраповское распределения вокруг асимптотического:

$$F_{\widehat{\varphi}}(z) = \Phi(z, V_\theta) + \frac{h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{n}\right),$$

$$F_{\widehat{\varphi}}^*(z) = \Phi(z, V_\theta^*) + \frac{h_1(z, F^*)}{\sqrt{n}} + O\left(\frac{1}{n}\right).$$

Ошибки аппроксимации для асимптотического и бутстраповского распределений следующие:

$$\Phi(z, V_\theta) - F_{\widehat{\varphi}}(z) = -\frac{h_1(z, F)}{\sqrt{n}} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right),$$

$$F_{\widehat{\varphi}}^*(z) - F_{\widehat{\varphi}}(z) = \Phi(z, V_\theta^*) - \Phi(z, V_\theta) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right),$$

ибо $V_\theta^* - V_\theta$ сходится к нулю со скоростью \sqrt{n} . Видим, что в данном случае использование бутстрапа не приводит к асимптотическому рафинированию. Итак, бутстрапирование асимптотически непивотальных статистик не дает асимптотического рафинирования.

А теперь рассмотрим модуль t-статистики

$$\widehat{\varphi} = \frac{|\widehat{\theta} - \theta|}{\text{se}(\widehat{\theta})} \xrightarrow{d} |N(0, 1)|.$$

Вновь разложим точное и бутстраповское распределения по Эджворту:

$$F_{\widehat{\varphi}}(z) = 2\Phi(z) - 1 + \frac{2h_2(z, F)}{n} + O\left(\frac{1}{n\sqrt{n}}\right),$$

$$F_{\widehat{\varphi}}^*(z) = 2\Phi(z) - 1 + \frac{2h_2(z, F^*)}{n} + O\left(\frac{1}{n\sqrt{n}}\right).$$

Здесь ошибки аппроксимации для асимптотики и бутстрапа имеют порядки соответственно

$$2\Phi(z) - 1 - F_{\widehat{\varphi}}(z) = O\left(\frac{1}{n}\right),$$

$$F_{\widehat{\varphi}}^*(z) - F_{\widehat{\varphi}}(z) = \frac{2}{n}(h_2(z, F^*) - h_2(z, F)) + O\left(\frac{1}{n\sqrt{n}}\right) = O\left(\frac{1}{n\sqrt{n}}\right).$$

Естественно, мы достигаем асимптотического рафинирования, но вдобавок заметим, что бутстрапирование симметричного двустороннего теста (модуля t-статистики) имеет ошибку большего порядка малости, чем бутстрапирование одностороннего (просто t-статистики).

Итак, теперь мы знаем причину, по которой желательно, если это возможно, пивотизировать и симметризовывать бутстрапируемые статистики.

7 Построение бутстраповских выборок (ресэмплинг) в регрессиях

Чаще всего исследователь производит инференцию в регрессионном контексте. Рассмотрим для начала кросс-секционную регрессию

$$y = x'\beta + e, \quad E[e|x] = 0.$$

Пусть имеется набор пар $\{(x_i, y_i)\}_{i=1}^n$, представляющий из себя случайную выборку.

Самый непосредственный способ ресэмплинга, так называемый *непараметрический бутстрап*, – вытягивать n пар (x_i^*, y_i^*) из исходной выборки $\{(x_i, y_i)\}_{i=1}^n$. Ровно того же результата можно достичь ресэмплингом пар (x_i^*, \hat{e}_i^*) из набора $\{(x_i, \hat{e}_i)\}_{i=1}^n$, где $\hat{e}_i = y_i - x_i'\hat{\beta}$ – МНК-остатки, и последующего восстановления переменных левой части с помощью модели: $y_i^* = x_i^{*\prime}\hat{\beta} + \hat{e}_i^*$. Нетрудно убедиться, что этот так называемый *остаточный бутстрап* численно идентичен непараметрическому. Зачем же он нужен? Дело в том, что в модели может присутствовать дополнительная информация в форме взаимоотношения регрессионных ошибок и регрессоров, и необходимо принять эту информацию во внимание при ресэмплинге, что удобно сделать именно через остаточный бутстрап. Например, если известно, что ошибки и регрессоры независимы, то эффективность бутстрапа можно увеличить, извлекая случайно с возвращением x_i^* из $\{x_i\}_{i=1}^n$ и \hat{e}_i^* из $\{\hat{e}_i\}_{i=1}^n$ независимо друг от друга. А если вдобавок ошибки распределены нормально, т.е. $e_i \sim N(0, \sigma^2)$, то эффективность бутстрапа можно еще больше увеличить, извлекая бутстраповские ошибки не из остатков, а из нормального распределения $N(0, \hat{\sigma}^2)$.

Теперь обсудим ресэмплинг при бутстрапе временных рядов. Временной ряд отличается от кросс-секционной выборки тем, что наблюдения (уравнения) здесь зависимы, и перемешивание при непараметрическом бутстрапе разрушает эту зависимость, так что вероятностная структура бутстраповских данных уже не имитирует вероятностную структуру исходных данных. Чтобы избежать этого, чаще всего используется *блочный бутстрап*, в котором бутстраповская выборка строится из блоков исходной выборки. Аналогично случаю независимых наблюдений, во временных рядах возможен остаточный бутстрап, однако такой способ применим только в тех редких случаях, когда регрессионные ошибки (инновации) серийно независимы.

Рассмотрим несколько альтернативных способов блочного ресэмплинга. Пусть $\{(x_t, y_t)\}_{t=1}^T$ – исходная выборка, а l – длина блока. Блоки могут быть перекрывающимися, тогда в первый блок войдут пары $(x_1, y_1), \dots, (x_l, y_l)$, во второй – $(x_2, y_2), \dots, (x_{l+1}, y_{l+1})$, в третий – $(x_3, y_3), \dots, (x_{l+2}, y_{l+2})$, и наконец в $(T-l+1)$ -ый – наблюдения $(x_{T-l+1}, y_{T-l+1}), \dots, (x_T, y_T)$. Блоки могут быть неперекрывающимися, тогда в первый блок войдут $(x_1, y_1), \dots, (x_l, y_l)$, во второй – $(x_{l+1}, y_{l+1}), \dots, (x_{2l}, y_{2l})$, и наконец, в последний $\lfloor \frac{T}{l} \rfloor$ -ый блок – пары $y_{l\lfloor \frac{T}{l} \rfloor - l + 1}, \dots, y_{l\lfloor \frac{T}{l} \rfloor}$. При построении бутстраповской выборки блоки извлекаются случайно с возвращением, как отдельные пары извлекаются из выборки при непараметрическом бутстрапе в кросс-секциях. Длина блока выбирается исследователем и отражает компромисс между сохранением временной структуры ряда и разнообразием бутстраповских выборок.

Преыдушие два варианта блочного ресэмплинга нарушают стационарность ряда, т.е. из стационарной исходной выборки получаются нестационарные бутстраповские выборки из-за разрывов зависимости между блоками. Чтобы бутстраповские выборки получались стационарными, был предложен *стационарный бутстрап*, основанный на нефиксированной длине блоков. А именно, задается вероятность окончания блока p . Первый элемент бутстраповской выборки выбирается случайно; затем с вероятностью $1-p$ в текущий блок включается следующий элемент исходной выборки, а с вероятностью p начинается новый блок, первый элемент которого снова выбирается случайно из исходной выборки. Так продолжается, пока в бутстраповскую выборку не будет набрано нужное количество элементов.

8 Что вошло в «Квантиль» и что осталось за кадром

Помимо данной работы, содержащей в основном вводный материал, настоящий выпуск журнала содержит еще три статьи по бутстрапу.

- Эссе Расселла Дэвидсона посвящено ресэмплинговым схемам в различных эконометрических моделях. Автор формулирует «золотые правила», которых следует придерживаться ради получения наиболее точной бутстраповской инференции. Несколько под другим углом обосновываются рецентрирование и пивотизация. Кроме того, автор обсуждает вычислительные аспекты, актуальные при проведении симуляций.
- Статья Питера Бюльмана обсуждает существующие ресэмплинговые схемы при бутстрапировании временных рядов: блочная, решетчатая и локальная бутстрап-схемы. Помимо теоретических свойств этих схем, автор приводит результаты экспериментов Монте-Карло, описывающие их поведение на практике.
- Наконец, работа Валентины Корради посвящена бутстрапу в контексте (обобщенного) метода моментов. Особое внимание уделяется временным рядам и асимптотическому рафинированию.

Некоторые специфические, хоть и важные, методы бутстрапирования недостаточно полно освещены в данном выпуске. Ниже приводятся ссылки на некоторые опубликованные работы.

- Бутстрапирование в моделях с порогами (Hansen, 1997) и структурными сдвигами (Hansen, 2000b) являются частными случаями общего подхода бутстрапа с фиксированными регрессорами для тестирования гипотезы линейности (Hansen, 2000a).
- Бутстрапирование условных моделей временных рядов включает в себя марковский бутстрап (Horowitz, 2003) и его упрощенную версию – бутстрапирование марковской цепью (Anatolyev & Vasnev, 2002).
- Бутстрапирование моделей VAR и функций импульсного отклика обсуждается в Kilian (1998) и Kilian (1999), а применение бутстрапа при прогнозировании – в Kim (1999).
- Бутстрапирование моделей с единичным корнем и соответствующих тестов обсуждается в Inoue & Kilian (2002) и Park (2003), а сеточный бутстрап для моделей с корнем, близким к единичному – в Hansen (1999).
- Использование бутстрапа для выбора модели из парных альтернатив продемонстрировано в Маккракен (2006).
- Бутстрапирование с целью борьбы с «добычей данных» разработано в White (2000), а применение к трейдинговым стратегиям – в Sullivan, Timmermann & White (1999).
- Бутстрапирование в моделях бинарного выбора исследуется в Wang, Wang & Carroll (1997), а в моделях с цензурированием – в Efron (1981).
- Различные варианты бутстрапирования моделей панельных данных (точнее, моделей компонент ошибки) описываются в Andersson & Karlsson (2001).
- Эффективное бутстрапирование в контексте условий на моменты, оцениваемых с помощью метода эмпирического правдоподобия или ОММ, привлекающее неравномерные веса при ресэмплинге, разработано в Brown & Newey (2002).

- Бутстрапирование непараметрических моделей и ядерных оценок обсуждается в Härdle & Bowman (1988).
- Бутстрапирование негладких функций статистик и применение, в частности, к медианной и квантильной регрессиям разработано в Hahn (1995) и Horowitz (1998).
- Ближнему по духу к бутстрапу методу подвыборок посвящена книга Politis, Romano & Wolf (1999).

Наконец, существует довольно много обзорного материала по бутстрапу:

- Книги по бутстрапу в основном написаны статистиками: Efron & Tibshirani (1993), Davison & Hinkley (1997), Shao & Tu (1995).
- Эконометрическая специфика лучше отражена в обзорных статьях эконометристов, например, в Hall (1994) и Horowitz (2001).
- Отдельная монография Hall (1994) посвящена разложениям Эджворта. Хорошим пособием по подобным асимптотическим разложениям является учебник Барндорф-Нильсен & Кокс (1999).
- В сборнике статей Эфрон (1988) имеются переведенные на русский язык статьи Эфрона, включая классическую работу Efron (1979).

Список литературы

- Анатольев, С.А. (2005). Асимптотические приближения в современной эконометрике. *Экономика и математические методы* 41, 84–94.
- Барндорф-Нильсен О. & Д. Кокс (1999). *Асимптотические методы в математической статистике*. Москва: Мир.
- Маккракен, М. (2006). Парные тесты на одинаковую точность прогнозов. *Квантиль* 1, 53–62.
- Эфрон, Б. (1988). Нетрадиционные методы многомерного статистического анализа (сборник статей). Москва: Финансы и статистика.
- Anatolyev, S. & A. Vasnev (2002). Markov chain approximation in bootstrapping autoregressions. *Economics Bulletin* 3, 1–8
- Andersson, M. & S. Karlsson (2001). Bootstrapping error component models. *Computational Statistics* 16, 221–231.
- Brown, B.W. & W.K. Newey (2002). Generalized method of moments, efficient bootstrapping, and improved inference. *Journal of Economic & Business Statistics* 20, 507–517.
- Cowles, A. (1934). Can stock market forecasters forecast? *Econometrica* 1, 309–324.
- Davison, A.C. & D.V. Hinkley (1997). *Bootstrap Methods and their Applications*. Cambridge University Press.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1–26.
- Efron, B. (1981). Censored data and the bootstrap. *Journal of American Statistical Association* 76, 312–319.
- Efron, B. & R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
- Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory* 11, 105–121.
- Hall, P. (1994). Methodology and theory for the bootstrap. Глава 39 в *Handbook of Econometrics* (под редакцией Engle, R. & D. McFadden), том 4. Elsevier Science.
- Hansen, B.E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics & Statistics* 81, 594–607.
- Hansen, B.E. (2000a). Sample splitting and threshold estimation. *Econometrica* 68, 575–604.
- Hansen, B.E. (2000b). Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.

- Hansen, B.E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics & Econometrics* 2, 1–14.
- Härdle, W. & A. Bowman (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of American Statistical Association* 83, 102–110.
- Horowitz, J.L. (2003). Bootstrap methods for Markov processes. *Econometrica* 71, 1049–1082.
- Horowitz, J.L. (1998). Bootstrap methods for median regression models. *Econometrica* 66, 1327–1352.
- Horowitz, J.L. (2001). The bootstrap. Глава 52 в *Handbook of Econometrics* (под редакцией Heckman, J.J. & E.E. Leamer), том 5. Elsevier Science.
- Inoue, A. & L. Kilian (2002). Bootstrapping autoregressive processes with possible unit roots. *Econometrica* 70, 377–391.
- Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics & Statistics* 80, 218–230.
- Kilian, L. (1999). Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. *Review of Economics & Statistics* 81, 652–660.
- Kim J.H. (1999). Asymptotic and bootstrap prediction regions for vector autoregression. *International Journal of Forecasting* 15, 393–403.
- Park, J.Y. (2003). Bootstrap unit root tests. *Econometrica* 71, 1845–1895.
- Politis, D.N., J.P. Romano & M. Wolf (1999). *Subsampling*. Springer.
- Shao, J. & D. Tu (1995). *The Jackknife and Bootstrap*. Springer.
- Sullivan, R., A. Timmermann & H. White (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance* 54, 1647–1691.
- Wang, C.Y., S. Wang & R.J. Carroll (1997). Estimation in choice-based sampling with measurement error and bootstrap analysis. *Journal of Econometrics* 77, 65–86.
- White, H. (2000). A reality check for data snooping. *Econometrica* 68, 1097–1126.

The basics of bootstrapping

Stanislav Anatolyev

New Economic School, Moscow, Russia

This essay is an introduction to principles and methodology of the bootstrap. The basics of bootstrap inference, resampling and asymptotic refinement are given. The narration is accompanied with clarifying examples. There is also a brief description of other methodological essays of the current issue of *Quantile* and references to non-included material.

Бутстрапирование эконометрических моделей*

Расселл Дэвидсон[†]

Университет МакГилл и CIRÉQ, Монреаль, Канада
GREQAM, Марсель, Франция

Бутстрап является статистическим методом, все более широко применяемым в эконометрике. Хотя он способен обеспечить очень надежную инференцию, следует соблюдать некоторые меры предосторожности, чтобы ее добиться. Формулируются два «золотых правила», соблюдение которых позволяет получить наилучшие результаты, которые может дать бутстрап. Бутстрапирование всегда включает формирование бутстраповского процесса, порождающего данные (DGP). Обсуждаются основные типы используемых в настоящее время бутстраповских DGP с примерами их применения в эконометрике. Способы использования бутстрапа для построения доверительных множеств несколько отличаются от методов тестирования гипотез. Рассматривается взаимосвязь между этими двумя типами задач.

Ключевые слова: бутстрап, тестирование гипотез, доверительное множество
Классификация JEL: C10, C12, C15

1 Введение

Бутстрап – это статистический метод, наиболее часто реализуемый с помощью симуляций. Симуляция не является неотъемлемым элементом бутстрапа, хотя на практике только тривиальные случаи не требуют ее применения. Основная идея бутстраповского тестирования заключается в том, что когда интересующая исследователя тестовая статистика имеет неизвестное распределение при нулевой гипотезе, это распределение можно охарактеризовать, используя информацию, содержащуюся в анализируемых данных.

Все просто, если статистика является *пивотальной* при нулевой гипотезе. Это означает, что распределение статистики одинаково при любом порождающем ее DGP, при условии, что этот DGP удовлетворяет нулевой гипотезе. Если обозначить множество всех DGP, удовлетворяющих нулевой гипотезе, за \mathbb{M} , то если статистика пивотальна, каждую процедуру, дающую ее распределение при любом DGP из \mathbb{M} , можно использовать для получения информации о распределении. Можно рассматривать множество \mathbb{M} как *модель*, а нулевую гипотезу – как утверждение о *правильной спецификации* этой модели, что означает принадлежность истинного неизвестного DGP, породившего анализируемые данные, множеству \mathbb{M} .

Процедура, наиболее полезная для поиска распределения статистики при нулевой гипотезе, – это симуляции. Исследователь генерирует много искусственных выборок из произвольного DGP, принадлежащего \mathbb{M} и наиболее простого для симулирования, и для каждой из этих выборок, обычно называемых *бутстраповскими выборками*, подсчитывает реализацию статистики. Эмпирическая функция распределения (ЭФР) этих *бутстраповских статистик* затем используется как основанная на симуляциях оценка неизвестного распределения.

*Перевод Б. Гершмана и С. Анатольева. Работа осуществлена при поддержке Канадской программы обеспечения исследовательских кафедр (кафедра экономики, Университет МакГилл) и грантов Канадского совета по исследованиям в области социальных и гуманитарных наук и Квебекского фонда исследований в области науки и культуры. Автор выражает благодарность Джеймсу Маккиннону за полезные замечания. Цитировать как: Дэвидсон, Расселл (2007) «Бутстрапирование эконометрических моделей», Квантиль, №3, стр. 13–36. Citation: Davidson, Russell (2007) “Bootstrapping econometric models,” Quantile, No.3, pp. 13–36.

[†]Адрес: Department of Economics, McGill University, Montréal, Québec, H3A 2T7; GREQAM, Centre de la Vieille Charité, 2 Rue de la Charité, 13236 Marseille cedex 02, France. Электронная почта: russell.davidson@mcgill.ca

Если распределение статистики при нулевой гипотезе известно, становится возможной различного рода инференция. Наиболее часто применяемые типы инференции основаны на *критических значениях* или на *P-значениях*. Первые представляют собой квантили распределения статистики при нулевой гипотезе, определенные как функции от желаемого уровня значимости теста. Последние являются предельными уровнями значимости, то есть уровнями, при которых тест находится на грани между отвержением и неотвержением нулевой гипотезы.

В частности, если нулевая гипотеза отвергается, когда реализуемая статистика слишком большая, то для теста при уровне значимости α критическое значение – это $(1 - \alpha)$ -квантиль распределения статистики при нулевой гипотезе. Для реализации τ статистики соответствующее *P-значение* равно $1 - F(\tau)$, где F – кумулятивная функция распределения (КФР) тестовой статистики при нулевой гипотезе. Для тестов, отвергающих нулевую гипотезу при маленьких значениях тестовой статистики, критическим значением является α -квантиль, а *P-значение* равно $F(\tau)$. Для двусторонних тестов требуются два критических значения, нижнее и верхнее. В качестве первого обычно берут $(\alpha/2)$ -квантиль, а в качестве второго – $(1 - \alpha/2)$ -квантиль. *P-значение* для реализации τ равно $2 \min(F(\tau), 1 - F(\tau))$.

Есть другие способы построения критических значений для двусторонних тестов. Если использовать β - и γ -квантили в качестве нижнего и верхнего критических значений соответственно, достаточно выполнения равенства $1 - \gamma + \beta = \alpha$, чтобы уровень значимости был равен α . Можно также выбирать β и γ , минимизируя расстояние между двумя критическими значениями при наличии этого ограничения.

Поскольку распределение статистики, пивотальной при нулевой гипотезе, можно оценить с помощью симуляций, инференция может быть основана на квантилях КФР оцененного распределения. В пределе, при бесконечном числе бутстраповских выборок, ошибка симуляции исчезает и достигается точная инференция, в том смысле, что вероятность отвержения тестом нулевой гипотезы при уровне значимости α в точности равна α , когда нулевая гипотеза истинна. Если инференция основана на *P-значении*, то для любого α между 0 и 1 вероятность получения *P-значения*, меньшего, чем α , при истинности нулевой гипотезы, в точности равна α .

Для получения точной инференции необязательно стремиться к недостижимому пределу бесконечного числа бутстраповских выборок, если исследователь готов ограничиться конкретным уровнем значимости. Если обозначить конечное число используемых бутстраповских выборок за B , инференция будет точной, если уровень α таков, что $\alpha(B + 1)$ является целым числом. Чтобы убедиться в этом, заметим, что бутстраповские статистики, обозначаемые τ_j^* , $j = 1, \dots, B$, как и статистика τ , полученная из исходной выборки, составляют множество из $B + 1$ статистик, которые при нулевой гипотезе являются независимыми одинаково распределенными (IID) случайными величинами. Следовательно, число r бутстраповских статистик, критических относительно τ , в соответствии с любым правилом определения критических областей равномерно распределено на множестве целых чисел $0, 1, \dots, B$, причем каждое возможное значение r имеет вероятность $1/(B + 1)$. Бутстраповское *P-значение* – это вероятностная масса бутстраповского распределения (то есть эмпирического распределения B бутстраповских статистик) в области, критической для τ , а эта вероятностная масса есть просто r/B .

Тогда вероятность получения бутстраповского *P-значения*, меньшего, чем α , равна $\mathbb{P}(r < \alpha B)$. Пусть $\lceil \alpha B \rceil$ – наименьшее целое число, не меньшее, чем αB . Тогда число возможных значений r , (строго) меньших αB , равно $\lceil \alpha B \rceil$. Следовательно, $\mathbb{P}(r < \alpha B) = \lceil \alpha B \rceil / (B + 1)$. Эта вероятность равна α тогда и только тогда, когда $\alpha(B + 1) = \lceil \alpha B \rceil$. Требование, чтобы $\alpha(B + 1)$ было целым числом, естественно, является необходимым. В обратную сторону, предположим, что $\alpha(B + 1) = k$, k целое. Тогда $\alpha B = k - \alpha$, а значит, $\lceil \alpha B \rceil = k$, так как $0 < \alpha < 1$. Поэтому вероятность того, что $r < \alpha B$, равна $k / (B + 1) = \alpha(B + 1) / (B + 1) = \alpha$.

Данное свойство является причиной того, что во многих исследованиях, число бутстраповских выборок полагается равным, например, 99, 199, 399 или 999. Десятичная система привела к стандартной привычке выбирать в качестве уровней значимости целое число процентов, а такие числа при добавлении единицы без остатка делятся на 100. В современную компьютерную эру, возможно, было бы более рационально устанавливать B , кратное 16 или 256 (в десятичной системе!) и вычитать единицу.

Тестирование на основе симуляций с использованием пивотальной статистики на самом деле гораздо старше бутстрапирования. Подобные процедуры называются *тестами Монте-Карло* и были впервые применены в 1950-х годах; см. Dwass (1957), а также Dufour & Khalaf (2001) с более современным обзором. В то время не было чем-то неслыханным использовать тест Монте-Карло на основе только 19 симулированных выборок, так как он позволяет проводить точную инференцию на уровнях значимости 5% и 10%.

Полностью пивотальные статистики редко встречаются в эконометрической практике, хотя они не совершенно безызвестны. Гораздо чаще встречаются приблизительно пивотальные статистики, распределения которых зависят, но не очень сильно, от конкретного DGP из \mathbb{M} , который их порождает. Чтобы придать определенный смысл этому размытому определению, обычно строят *асимптотическую теорию* для модели \mathbb{M} . Это означает, что строится математическая конструкция, которая позволяет каждому DGP из \mathbb{M} генерировать выборки произвольно большого размера. Часто довольно очевидно, как это сделать, как, например, в случае, если наблюдения в выборке являются IID, но в других случаях может быть проблематично найти подходящую асимптотику для имеющейся задачи. Когда проблема решена, необходимо, чтобы предельное распределение статистики, когда размер выборки стремится к бесконечности, было одинаковым для всех DGP из \mathbb{M} . Статистика, для которой можно получить асимптотику, удовлетворяющую этому требованию, называется *асимптотически пивотальной*.

Бутстраповское тестирование осуществляется аналогично описанному выше тестированию Монте-Карло. Однако возникает новая проблема. Поскольку распределение непивотальной статистики в конечных выборках зависит от конкретного DGP из \mathbb{M} , уже нельзя произвольно выбирать DGP для генерации симулированных выборок. Как поступать с выбором *бутстраповского DGP*, обсуждается в следующем разделе.

Основное внимание в настоящем эссе уделяется бутстраповскому тестированию гипотез, но бутстрап может применяться и более широко. Основной принцип состоит в том, что в рамках некоторого множества DGP, или модели, DGP, который на самом деле породил имеющиеся данные, можно оценить по этим данным. Тогда любую величину, будь то скаляр, вектор или матрица, которая может быть представлена в виде функции или функционала от DGP, можно оценить как эту функцию или функционал от оцененного DGP. Таким образом, бутстрап можно использовать для оценивания смещения и дисперсии, квантилей, моментов и многих других величин. Бутстрап может давать хорошие оценки подобных величин не в любых обстоятельствах, но, как будет видно, в случае тестирования он может обеспечить более надежную инференцию, чем другие стандартные методы.

Настоящее эссе дополняет обзор бутстраповских методов в Davidson & MacKinnon (2006a). Здесь рассматривается бутстрапирование независимых данных и не обсуждаются сложные проблемы, которые могут возникнуть, когда наблюдения в выборке взаимозависимы. Многие из этих проблем рассмотрены в Politis (2003). Другие полезные обзоры о бутстрапе содержатся в Horowitz (2001) и Horowitz (2003).

2 Золотые правила бутстрапирования

Если тестовая статистика τ асимптотически пивотальна для данной модели \mathbb{M} , ее распределение не должно слишком сильно меняться как функция конкретного DGP, скажем, μ , в

рамках этой модели. Обычно можно показать, что расстояние между распределением τ для DGP μ при размере выборки n и при n , стремящемся к бесконечности, стремится к нулю как некоторая отрицательная степень n , обычно $n^{-1/2}$. Концепцию «расстояния» между распределениями можно определить различными способами, некоторые из которых более, чем другие, подходят для бутстраповского тестирования.

Асимптотическое рафинирование

Говоря эвристически, если расстояние между распределением в конечной выборке для любого DGP $\mu \in \mathbb{M}$ и предельным распределением имеет порядок $n^{-\delta}$ для некоторого $\delta > 0$, то, поскольку предельное распределение одинаково для всех $\mu \in \mathbb{M}$, расстояние между распределениями в конечной выборке для двух DGP μ_1 и μ_2 также имеет порядок $n^{-\delta}$. Если расстояние между μ_1 и μ_2 также мало в некотором смысле, скажем, имеет порядок $n^{-\epsilon}$, то расстояние между распределениями τ при μ_1 и при μ_2 должно быть порядка $n^{-(\delta+\epsilon)}$.

Рассуждения такого рода используются, чтобы показать, что бутстрап при благоприятных обстоятельствах может давать выгоду в виде *асимптотического рафинирования*. Очертание доказательства было дано в известной статье Beran (1988). Несомненно мудро Беран ограничивается в этой статье лишь наброском доказательства без обсуждения формальных условий регулярности. На сегодняшний день так и не существует по-настоящему общей теории бутстраповского тестирования, которая бы формально воплощала простую идею, предложенную Бераном. Вместо этого имеется множество частных результатов, доказывающих существование рафинирования в частных случаях, наряду с теми результатами, которые показывают, что бутстрап не работает в других особых случаях. Возможно, наиболее важный пример негативного результата подобного рода, часто называемый *провалом бутстрапа*, касается бутстрапирования в ситуации, когда истинный DGP порождает данные с распределением, имеющим тяжелые хвосты; см. Athreya (1987) для случая бесконечной дисперсии. Дела обстоят значительно лучше при *параметрическом бутстрапе*, который рассматривается в следующем разделе.

Техника, широко применяемая в работах по асимптотическому рафинированию при бутстрапировании, – это *разложения Эджворта* для распределений, обычно таких, которые стремятся к стандартному нормальному, когда размер выборки стремится к бесконечности. Стандартная ссылка для этого направления исследований – работа Hall (1992), хотя нет недостатка в более свежих работах, основанных на разложениях Эджворта. Несмотря на то, что эта техника может приводить к важным теоретическим результатам, она не очень полезна для количественного объяснения свойств бутстраповских тестов. В частных случаях истинное распределение бутстраповского P -значения в конечной выборке, оцененное с помощью симуляций, можно в дальнейшем легко удалить из аппроксимации Эджворта его распределения, а не асимптотического предельного распределения.

Правила бутстрапирования

Несмотря на все теоретические предостережения, долгий опыт показал, что бутстраповские тесты во многих важных для прикладной эконометрики обстоятельствах намного более надежны, чем тесты, основанные на асимптотической теории разного рода. В остальной части данного раздела изложены некоторые правила, которые следует соблюдать для получения надежных бутстраповских P -значений.

DGP, используемый для формирования бутстраповских выборок, по которым вычисляются бутстраповские статистики, называется *бутстраповским DGP* и будет обозначаться μ^* . Поскольку при тестировании бутстрап используется для оценивания распределения тестовой статистики при нулевой гипотезе, первое золотое правило бутстрапирования заключается в следующем:

Золотое правило 1: Бутстраповский DGP μ^* должен принадлежать модели \mathbb{M} , которая представляет нулевую гипотезу.

Этому правилу не всегда возможно следовать, или, даже если возможно, это может быть в некоторых случаях сложно. Это станет яснее при рассмотрении доверительных множеств. В таких случаях обычной техникой является изменение нулевой гипотезы таким образом, чтобы используемый бутстраповский DGP ей удовлетворял.

Если в нарушение этого правила бутстраповский DGP не удовлетворяет нулевой гипотезе, тестируемой с помощью бутстраповской статистики, бутстраповский тест может полностью потерять мощность. Мощность теста происходит из того факта, что статистика имеет разные распределения при нулевой и альтернативной гипотезах. Бутстрапирование при альтернативной гипотезе смешивает эти разные распределения и ведет к совершенно ненадежной inferенции, даже в асимптотическом пределе.

Нарушение Золотого правила 1 в настоящее время встречается в эконометрических работах крайне редко, хотя оно встречалось в эконометрической литературе на первых порах применения бутстрапа. Одно из следствий этого правила состоит в том, что модель при нулевой гипотезе \mathbb{M} должна быть четко определена до выбора бутстраповского DGP. Как будет объясняться в следующем разделе, тестовые статистики, основанные на оценивании методом максимального правдоподобия, следует бутстрапировать, используя параметрический бутстрап, чтобы выполнялось Золотое правило 1. Ресэмплинг будет адекватным только в том случае, если модель при нулевой гипотезе допускает DGP, основанные на дискретных распределениях.

Если Золотое правило 1 должно выполняться, чтобы иметь асимптотически оправданный тест, то Золотое правило 2 заботится о том, чтобы сделать вероятность отвержения истинной нулевой гипотезы в результате бутстраповского теста как можно ближе к уровню значимости. Оно мотивировано приведенной выше идеей Берана.

Золотое правило 2: Если тестовая статистика не является пивотальной для модели при нулевой гипотезе \mathbb{M} , бутстраповский DGP должен быть наилучшей возможной оценкой истинного DGP, при предположении, что истинный DGP принадлежит \mathbb{M} .

Каким образом можно соблюдать это правило, сильно зависит от конкретного проводимого теста, но в целом оно означает, что хотелось бы, чтобы бутстраповский DGP был основан на оценках, *эффективных* при нулевой гипотезе.

Как только тип бутстраповского DGP выбран, процедура бутстраповского тестирования, основанного на симулированных бутстраповских выборках, осуществляется по следующей схеме.

- (i) Рассчитать тестовую статистику по исходной выборке, обозначив эту величину $\hat{\tau}$.
- (ii) Найти реализации всех остальных зависящих от данных компонент, необходимых для формирования бутстраповского DGP μ^* .
- (iii) Сгенерировать B бутстраповских выборок, используя μ^* , и для каждой из них рассчитать реализацию бутстраповской статистики τ_j^* , $j = 1, \dots, B$. Разумно выбирать B так, чтобы $\alpha(B + 1)$ было целым числом для всех интересующих уровней значимости α , обычно 1%, 5% и 10%.
- (iv) Рассчитать симулированное бутстраповское P -значение как долю бутстраповских статистик τ_j^* , критических относительно $\hat{\tau}$. Для статистики, отвергающей нулевую гипотезу при больших значениях, например, имеем:

$$P_{\text{bs}} = \frac{1}{B} \sum_{j=1}^B \mathbb{I}_{[\tau_j^* > \hat{\tau}]},$$

где $\mathbb{I}_{[\cdot]}$ – индикатор-функция, равная 1, если ее булевский аргумент истинен, и 0, если ложен.

Бутстраповский тест отвергает нулевую гипотезу на уровне значимости α , если $P_{bs} < \alpha$.

3 Параметрический бутстрап

Если модель \mathbb{M} , представляющую нулевую гипотезу, можно оценить методом максимального правдоподобия (ММП), существует взаимно однозначное соответствие между пространством параметров модели и DGP, которые ей принадлежат. Для любого допустимого набора параметров функция правдоподобия, рассчитанная для этих параметров, является функцией плотности. Таким образом, существует единственный DGP, ассоциированный с набором параметров. Следовательно, в \mathbb{M} входят только те DGP, которые полностью характеризуются множеством параметров.

Если модель \mathbb{M} оценивается методом максимального правдоподобия, то ММП-оценки являются эффективными оценками не только самих параметров, но и истинного DGP. Тогда оба золотых правила выполняются, если в качестве бутстраповского DGP выбирается DGP из \mathbb{M} , характеризующийся ММП-оценками параметров. В этом случае говорят о *параметрическом бутстрапе*.

В микроэконометрике такие модели, как пробит и логит, обычно оцениваются с помощью ММП. Это, конечно, лишь наиболее простые микроэконометрические модели, но они являются репрезентативными для всех других моделей, о которых разумно предположить, что данные можно описать чисто параметрической моделью. Возьмем в качестве примера модель бинарного выбора для иллюстрации параметрического бутстрапа.

Модель бинарного выбора

Предположим, что бинарная зависимая переменная y_t , $t = 1, \dots, n$, принимает только значения 0 и 1, и вероятность того, что $y_t = 1$, равна $F(\mathbf{X}_t\boldsymbol{\beta})$, где \mathbf{X}_t – вектор экзогенных переменных $1 \times k$, $\boldsymbol{\beta}$ – вектор параметров $k \times 1$, а F – функция, отображающая действительные числа в интервал $(0, 1)$. Для пробит-модели F – функция стандартного нормального распределения; для логит-модели – функция логистического распределения.

Вклад наблюдения t в логарифмическую функцию правдоподобия для всей выборки равен

$$\mathbb{I}_{[y_t=1]} \log F(\mathbf{X}_t\boldsymbol{\beta}) + \mathbb{I}_{[y_t=0]} \log(1 - F(\mathbf{X}_t\boldsymbol{\beta})),$$

где вновь $\mathbb{I}_{[\cdot]}$ – индикатор-функция. Теперь предположим, что вектор параметров $\boldsymbol{\beta}$ можно разделить на два подвектора, $\boldsymbol{\beta}_1$ и $\boldsymbol{\beta}_2$, и что при нулевой гипотезе $\boldsymbol{\beta}_2 = \mathbf{0}$. Тогда ММП-оценка *при ограничении*, то есть оценка только подвектора $\boldsymbol{\beta}_1$ при равном нулю $\boldsymbol{\beta}_2$, является асимптотически эффективной только для тех параметров, которые присутствуют при нулевой гипотезе. (Предполагается, что есть асимптотическая конструкция, допускающая произвольно большое число векторов \mathbf{X}_t объясняющих переменных с достаточно похожими свойствами, чтобы можно было применить обычную для ММП асимптотическую теорию.)

Хотя асимптотическая теория применяется для убеждения в желательности ММП-оценки, сам по себе бутстрап является процедурой, предназначенной исключительно для конечной выборки. Если обозначить ММП-оценку при ограничении за $\tilde{\boldsymbol{\beta}} \equiv [\tilde{\boldsymbol{\beta}}_1; \mathbf{0}]$, бутстраповский DGP можно представить следующим образом:

$$y_t^* = \begin{cases} 1 & \text{с вероятностью } F(\mathbf{X}_t\tilde{\boldsymbol{\beta}}), \\ 0 & \text{с вероятностью } 1 - F(\mathbf{X}_t\tilde{\boldsymbol{\beta}}), \end{cases} \quad t = 1, \dots, n. \quad (1)$$

Здесь используются обычные обозначения, в соответствии с которыми переменные, порожденные бутстраповским DGP, помечены звездочками. Заметим, что объясняющие переменные \mathbf{X}_t не помечены звездочками. Поскольку они предполагаются экзогенными, в бутстрапировании DGP не входит их повторное генерирование; напротив, они воспринимаются как фиксированные характеристики бутстраповского DGP и используются без изменения в каждой бутстраповской выборке. Поскольку бутстраповские выборки имеют тот же размер n , что и исходная выборка, не требуется генерировать объясняющие переменные для каких-либо других наблюдений, кроме тех, что имеются в действительности.

Формулу (1) легко реализовать для формирования бутстраповских выборок. С помощью генератора случайных чисел из равномерного распределения $U[0, 1]$ выбирается случайное число m_t . Далее генерируется y_t^* как $\mathbb{I}_{[m_t \leq F(\mathbf{X}_t \tilde{\beta})]}$. Большинство матричных или эконометрических компьютерных пакетов могут осуществить это действие как векторную команду, так что после подсчета n -мерного вектора с типичным элементом $F(\mathbf{X}_t \tilde{\beta})$ вектор \mathbf{y}^* с типичным элементом y_t^* можно сгенерировать с помощью одной команды.

Рекурсивная симуляция

В динамических моделях формирование бутстраповского DGP может потребовать *рекурсивной симуляции*. Возьмем в качестве примера очень простую авторегрессионную модель:

$$y_t = \alpha + \rho y_{t-1} + u_t, \quad u_t \sim \text{NID}(0, \sigma^2), \quad t = 2, \dots, n. \quad (2)$$

Согласно обозначениям, u_t независимо и одинаково распределены как $N(0, \sigma^2)$. То есть теперь зависимая переменная y_t непрерывна, в отличие от рассмотренной выше бинарной зависимой переменной. Параметрами модели являются α , ρ и σ^2 . Однако, даже если значения этих параметров определены, (2) не дает полную характеристику DGP. Поскольку (2) – рекурсивное соотношение, прежде чем оно даст единственное решение, необходимо начальное значение, или инициализация. Таким образом, хотя оно не является параметром в обычном смысле, первое наблюдение y_1 также должно быть специфицировано для полного описания модели.

ММП-оценивание модели (2) аналогично оцениванию обычным методом наименьших квадратов (МНК) с исключением первого наблюдения. Если (2) представляет нулевую гипотезу, то с помощью МНК и правда находятся оценки α , ρ и σ . Если нулевая гипотеза определяет значение какого-либо из этих параметров, требуя, например, чтобы $\rho = \rho_0$, то МНК используется для оценки модели с ограничением:

$$y_t - \rho_0 y_{t-1} = \alpha + u_t$$

при той же спецификации ошибок u_t , как в (2).

Теперь бутстраповский DGP – это DGP, содержащийся в нулевой гипотезе, который характеризуется оценками параметров и некоторым подходящим выбором начального значения y_1^* . Один из способов выбрать y_1^* – просто взять y_1 , значение в исходной выборке. В большинстве случаев это наилучший выбор. Он ограничивает модель (2), фиксируя начальное значение. Теперь бутстраповскую выборку можно сгенерировать рекурсивно, начиная с y_2^* . Для всех $t = 2, \dots, n$, имеем

$$y_t^* = \tilde{\alpha} + \tilde{\rho} y_{t-1}^* + \tilde{\sigma} v_t^*, \quad v_t^* \sim \text{NID}(0, 1). \quad (3)$$

Часто возникает потребность ограничить возможные значения ρ в пределах от -1 до 1 . Это ограничение делает ряд y_t асимптотически стационарным, под чем подразумевается, что, если сгенерировать очень длинную выборку из рекурсивного соотношения (2), то по мере приближения к концу выборки распределение y_t становится независимым от t , как и совместное распределение любой пары наблюдений, скажем, y_t и y_{t+s} . Иногда имеет смысл

потребуется стационарность ряда y_t , а не просто асимптотическую стационарность, чтобы распределение каждого наблюдения y_t , включая первое, всегда было одним и тем же. Часто возможно включить информацию о первом наблюдении в процедуру ММП-оценивания и получить, таким образом, более эффективную оценку, использующую дополнительную информацию. Теперь для бутстраповского DGP наблюдение y_1^* должно случайно выбираться из стационарного распределения.

Бутстраповское расхождение

В отличие от теста Монте-Карло, основанного на полностью пивотальной статистике, бутстраповский тест, вообще говоря, не обеспечивает точную инференцию. Это означает, что существует разница между действительной вероятностью отвержения нулевой гипотезы и номинальным уровнем значимости теста. Можно определить *бутстраповское расхождение* как эту разницу, являющуюся функцией от истинного DGP и номинального уровня значимости. Чтобы изучить бутстраповское расхождение, предположим без ограничения общности, что тестовая статистика, обозначаемая τ , уже имеет форму асимптотического P -значения. Тогда отвержение на уровне значимости α соответствует событию $\tau < \alpha$.

Введем две функции номинального уровня значимости теста α и DGP μ . Первая из них – это *функция вероятности отвержения*, или ФВО. Значение этой функции – это истинная вероятность отвержения нулевой гипотезы при μ с помощью теста на уровне значимости α для некоторого фиксированного конечного размера выборки n . Она определяется так:

$$R(\alpha, \mu) \equiv \mathbb{P}_\mu[\tau < \alpha]. \quad (4)$$

Здесь и далее предполагается, что для всех $\mu \in \mathbb{M}$ распределение τ имеет носитель $[0, 1]$ и абсолютно непрерывно по отношению к равномерному распределению на этом интервале.

Для конкретного μ функция $R(\alpha, \mu)$ – это просто функция распределения τ , подсчитанная в α . Обратная функция для ФВО – это *функция критического значения*, или ФКЗ, которая неявно задается уравнением

$$\mathbb{P}_\mu[\tau < Q(\alpha, \mu)] = \alpha. \quad (5)$$

Из (5) ясно, что $Q(\alpha, \mu)$ – это α -квантиль распределения τ при μ . Кроме того, из определений (4) и (5) следует, что

$$R(Q(\alpha, \mu), \mu) = Q(R(\alpha, \mu), \mu) = \alpha \quad (6)$$

для всех α и μ .

Далее абстрагируемся от случайности симуляции и предположим, что распределение τ при бутстраповском DGP точно известно. Бутстраповское критическое значение для τ при уровне значимости α равно $Q(\alpha, \mu^*)$; напомним, что μ^* обозначает бутстраповский DGP. Это случайная величина, которая была бы неслучайной и равной α , если бы τ была полностью пивотальной. Если τ приблизительно (например, асимптотически) пивотальна, реализации $Q(\alpha, \mu^*)$ должны быть близки к α . Это верно независимо от того, соответствует ли истинный DGP нулевой гипотезе, поскольку бутстраповский DGP μ^* ей соответствует, согласно первому Золотому правилу. Бутстраповское расхождение при DGP $\mu \in \mathbb{M}$ возникает из-за возможности того, что в конечной выборке $Q(\alpha, \mu^*) \neq Q(\alpha, \mu)$.

Отвержение нулевой гипотезы бутстраповским тестом – это событие $\tau < Q(\alpha, \mu^*)$. Применяя возрастающее преобразование $R(\cdot, \mu^*)$ к обеим частям и используя (6), становится понятно, что бутстраповский тест отвергает нулевую гипотезу, если

$$R(\tau, \mu^*) < R(Q(\alpha, \mu^*), \mu^*) = \alpha.$$

Таким образом, бутстраповское P -значение – это просто $R(\tau, \mu^*)$. Его можно интерпретировать как бутстраповскую тестовую статистику. Вероятность при μ того, что бутстраповский тест отвергает нулевую гипотезу при номинальном уровне значимости α , равна

$$\mathbb{P}_\mu[\tau < Q(\alpha, \mu^*)] = \mathbb{P}_\mu[R(\tau, \mu^*) < \alpha].$$

Определим две случайные переменные, которые являются детерминистскими функциями от двух случайных элементов, τ и μ^* , необходимых для расчета бутстраповского P -значения $R(\tau, \mu^*)$. Первая из этих случайных переменных распределена как $U[0, 1]$ при μ ; это

$$p \equiv R(\tau, \mu). \quad (7)$$

Равномерность распределения p следует из того, что $R(\cdot, \mu)$ – функция распределения τ при μ , и из предположения о том, что распределение τ абсолютно непрерывно на единичном интервале при любых $\mu \in \mathbb{M}$. Вторая случайная переменная – это

$$r \equiv R(Q(\alpha, \mu^*), \mu). \quad (8)$$

Можно переписать событие, ведущее к отвержению нулевой гипотезы бутстраповским тестом на уровне значимости α , как $R(\tau, \mu) < R(Q(\alpha, \mu^*), \mu)$, воздействуя на обе части неравенства $\tau < Q(\alpha, \mu^*)$ возрастающей функцией $R(\cdot, \mu)$. Учитывая определения (7) и (8), это событие формулируется просто как $p < r$. Обозначим функцию распределения r при μ условно на случайной величине p за $F(r|p)$. Тогда вероятность при μ отвергнуть нулевую гипотезу бутстраповским тестом на уровне значимости α равна

$$\mathbb{E}[\mathbb{I}_{[p < r]}] = \mathbb{E}[\mathbb{E}[\mathbb{I}_{[p < r]}|p]] = \mathbb{E}[\mathbb{E}[\mathbb{I}_{[r > p]}|p]] = \mathbb{E}[1 - F(p|p)] = 1 - \int_0^1 F(p|p) dp, \quad (9)$$

поскольку маргинальное распределение p – это $U[0, 1]$.

Полезное выражение для бутстраповского расхождения получается, если ввести случайную величину $q \equiv r - \alpha$. Тогда функция распределения q условно на p $F(\alpha + q|p) \equiv G(q|p)$. Вычитание α из вероятности отвержения (9) дает

$$1 - \alpha - \int_0^1 G(p - \alpha|p) dp.$$

Замена переменной интегрирования с p на $x = p - \alpha$ приводит к следующему выражению для бутстраповского расхождения:

$$\begin{aligned} 1 - \alpha - \int_{-\alpha}^{1-\alpha} G(x|\alpha + x) dx &= 1 - \alpha - \left[x G(x|\alpha + x) \right]_{-\alpha}^{1-\alpha} + \int_{-\alpha}^{1-\alpha} x dG(x|\alpha + x) \\ &= \int_{-\alpha}^{1-\alpha} x dG(x|\alpha + x), \end{aligned} \quad (10)$$

так как $G(-\alpha|0) = F(0|0) = 0$ и $G(1 - \alpha|1) = F(1|1) = 1$.

С очень высокой степенью точности (10) часто можно заменить на

$$\int_{-\infty}^{\infty} x dG(x|\alpha), \quad (11)$$

то есть матожидание q условно на p , находящемся на грани отвержения при уровне значимости α . В случаях, когда p и q независимы или почти независимы, хорошей аппроксимацией может быть даже замена (11) безусловным матожиданием q .

Случайная величина r – это вероятность того, что статистика, порожденная DGP μ , меньше α -квантиля бутстраповского распределения условно на этом распределении. Матожидание r за вычетом α тогда можно интерпретировать как смещение вероятности отвержения,

когда она оценивается с помощью бутстрапа. Действительное бутстраповское расхождение, являющееся неслучайной величиной, – это матожидание $q = r - \alpha$ при условии нахождения на грани отвержения. Аппроксимация (11) устанавливает грань на уровне α -квантиля τ при μ , тогда как точное выражение (10) учитывает тот факт, что грань на самом деле определяется бутстраповским DGP.

Если статистика τ асимптотически пивотальна, случайная величина q стремится к нулю при нулевой гипотезе по мере того, как размер выборки n стремится к бесконечности. Это так, поскольку для асимптотически пивотальной статистики предельное значение $R(\alpha, \mu)$ при данном α одно и то же для всех $\mu \in \mathbb{M}$, аналогично и для $Q(\alpha, \mu)$. Обозначим предельные функции от α (только) за $R^\infty(\alpha)$ и $Q^\infty(\alpha)$. При предположении об абсолютной непрерывности распределения функции R^∞ и Q^∞ взаимно обратные (вспомним (6)), а значит, когда $n \rightarrow \infty$, $r = R(Q(\alpha, \mu^*), \mu)$ стремится к $R^\infty(Q^\infty(\alpha)) = \alpha$, и, следовательно, $q = r - \alpha$ стремится к нулю по распределению, а значит, и по вероятности.

Предположим теперь, что случайные величины q и p независимы. Тогда функция условного распределения $G(\cdot|\cdot)$ – это просто функция безусловного распределения q , а бутстраповское расхождение (10) – безусловное матожидание q . Безусловное матожидание случайной величины, стремящейся к нулю, может стремиться к нулю быстрее, чем сама величина, и быстрее, чем матожидание, условное на другой, коррелированной с ней, случайной величине. Независимость q и p нечасто возникает на практике, но приблизительная (асимптотическая) независимость встречается регулярно, когда параметрический бутстрап используется вместе с ММП-оцениванием нулевой гипотезы. Стандартным результатом асимптотической теории ММП-оценивания является то, что ММП-оценки параметров модели асимптотически независимы от классических тестовых статистик, используемых для тестирования нулевой гипотезы о том, что модель правильно специфицирована, против некоторой параметрической альтернативы. В таких случаях бутстраповское расхождение стремится к нулю быстрее, чем если бы для формирования бутстраповского DGP использовались неэффективные параметры. Этот аргумент, дающий поддержку Золотому правилу 2, развивается в Davidson & MacKinnon (1999).

4 Ресэмплинг

Анализ в предыдущем разделе основан на абсолютной непрерывности распределения тестовой статистики для всех $\mu \in \mathbb{M}$. Даже при использовании параметрического бутстрапа абсолютная непрерывность не всегда имеет место. Например, зависимая переменная в модели бинарного выбора является дискретной, а значит, таковы и любые тестовые статистики, являющиеся функцией от нее, если только непрерывность не возникает по некоторой другой причине, чего не случается с тестовыми статистиками, обычно применяемыми в моделях бинарного выбора. Однако, поскольку дискретное множество значений, которые может принимать тестовая статистика, быстро становится очень большим по мере увеличения размера выборки, разумно предположить, что теория предыдущего раздела остается хорошим приближением при реалистичных размерах выборки.

Базовый ресэмплинг

Другое важное обстоятельство, при котором абсолютная непрерывность теряется, – когда бутстраповский DGP использует *ресэмплинг*. Ресэмплинг был ключевым аспектом исходной концепции бутстрапа, предложенной в основополагающей статье Efron (1979). Ресэмплинг ценен, когда нежелательно ограничивать модель так сильно, что все ее возможности охватываются варьированием конечного множества параметров. Классическим примером является модель регрессии, в которой не предполагается нормальность ошибок. В качестве конкретного примера вновь рассмотрим модель авторегрессии (2), ослабив условие на ошибки, то

есть требуя только независимость и одинаковую распределенность с нулевым матожиданием и дисперсией σ^2 .

Бутстраповский DGP (3) удовлетворяет Золотому правилу 1, так как нормальное распределение, очевидно, является допустимым, когда специфицируются всего лишь первые два момента. Но Золотое правило 2 побуждает искать наилучшую возможную оценку неизвестного распределения ошибок. Если бы ошибки были наблюдаемы, наилучшей непараметрической оценкой их распределения была бы эмпирическая функция распределения. Ненаблюдаемые ошибки могут быть оценены, или приближены, остатками от оценивания модели при нулевой гипотезе. Если обозначить эмпирическую функцию распределения этих остатков за \hat{F} , то (3) можно заменить на

$$y_t^* = \tilde{\alpha} + \tilde{\rho}y_{t-1}^* + u_t^*, \quad u_t^* \sim \text{IID}(\hat{F}), \quad t = 2, \dots, n.$$

где из обозначений видно, что бутстраповские ошибки u_t^* являются IID-реализациями из эмпирического распределения, характеризуемого ЭФР \hat{F} .

Термин *ресэмплинг* происходит от того факта, что простейший способ сгенерировать u_t^* – набрать их из остатков случайно с возвращением. Остатки воспринимаются как выборка из истинного DGP, так что эта операция называется «ресэмплинг» (повторное формирование выборки)¹. Для любого $t = 2, \dots, n$ «вытягиваем» случайное число m_t из равномерного $U(0, 1)$ распределения, а затем получаем u_t^* следующим образом:

$$s = \lfloor 2 + (n-1)m_t \rfloor, \quad u_t^* = \tilde{u}_s,$$

где $\lfloor x \rfloor$ обозначает наибольшее целое число, не большее x . Для m_t , близкого к нулю, $s = 2$; для m_t , близкого к 1, $s = n$, и можно видеть, что s равномерно распределено на множестве целых чисел $2, \dots, n$. Полагая u_t^* равным остатку (для модели при ограничении) \tilde{u}_s , таким образом осуществляем требуемую процедуру ресэмплинга.

Более изощренный ресэмплинг

Но является ли эмпирическое распределение остатков действительно наилучшей возможной оценкой распределения ошибок? Не всегда. Рассмотрим даже еще более простую модель, чем (2), без константы:

$$y_t = \rho y_{t-1} + u_t, \quad u_t \sim \text{IID}(0, \sigma^2). \quad (12)$$

Когда она оценивается с помощью МНК, или если нулевая гипотеза фиксирует значение ρ , когда «остатки» – это просто наблюдаемые значения $y_t - \rho_0 y_{t-1}$, сумма остатков, вообще говоря, не равна нулю именно из-за отсутствия константы. Но модель (12) требует, чтобы матожидание распределения ошибок было равным нулю, тогда как матожидание эмпирического распределения остатков – это их среднее. Таким образом, использование этого эмпирического распределения нарушает Золотое правило 1.

Это легко исправить, заменив остатки их отклонениями от среднего, а затем осуществляя ресэмплинг этих центрированных остатков. Но как насчет Золотого правила 2? Дисперсия остатков – это сумма их центрированных квадратов, деленная на n :

$$V = \frac{1}{n} \sum (\tilde{u}_t^2 - \bar{u})^2,$$

где \bar{u} – среднее нецентрированных остатков. Но несмещенная оценка дисперсии ошибок – это

$$s^2 = \frac{1}{n-1} \sum (\tilde{u}_t^2 - \bar{u})^2.$$

¹См. «Словарь» в настоящем номере «Квантиля», стр. 69 – Редактор.

Более общо, в любой регрессионной модели, использующей k степеней свободы при оценивании регрессионных параметров, несмещенная оценка дисперсии – это сумма квадратов остатков, деленная на $n - k$. Это означает, что хотелось бы осуществлять ресэмплинг множества *нормированных* остатков, в данном случае $\sqrt{n/(n-k)}\tilde{u}_t$. Дисперсия эмпирического распределения этих масштабированных остатков тогда равна несмещенной оценке дисперсии.

Конечно, некоторые проблемы не зависят от масштаба. Действительно, тестовые статистики, являющиеся отношениями, не зависят от масштаба для обеих моделей (2) и (12) при предположении о стационарности. Следовательно, для таких моделей не имеет смысла осуществлять нормировку, поскольку бутстраповские статистики, рассчитанные по тому же множеству случайных чисел, не меняются при нормировке. Это свойство сродни пивотальности в том, что изменение некоторых, но не всех, параметров модели при нулевой гипотезе, оставляет распределение тестовой статистики неизменным. В таких случаях не стоит беспокоиться об оценивании параметров, не влияющих на распределение статистики τ .

Индекс бедности

Центрирование и нормировка – простые операции, меняющие первые два момента распределения. В некоторых обстоятельствах может возникнуть желание повлиять на более сложные функционалы распределения. Предположим, например, что возникло желание провести инференцию об индексе бедности. Доступна IID-выборка индивидуальных доходов, извлеченная случайным образом из изучаемой популяции, и нулевая гипотеза состоит в том, что индекс бедности имеет конкретное заданное значение. Для определенности рассмотрим один из FGT-индексов, определяемый следующим образом; см. Foster, Greer, & Thorbecke (1984):

$$\Delta^\alpha(z) = \int_0^z (z - y)^{\alpha-1} dF(y).$$

Здесь z интерпретируется как черта бедности, а F – функция распределения доходов. По мере увеличения параметра α индекс прогрессивно увеличивает вес больших значений *глубины бедности*, то есть разницы $z - y$ между чертой бедности и доходом y бедного индивида. Предположим, что черта бедности z и параметр α фиксированы на некоторых предопределенных уровнях. Очевидная оценка $\Delta^\alpha(z)$ – это просто

$$\hat{\Delta}^\alpha(z) = \int_0^z (z - y)^{\alpha-1} d\hat{F}(y),$$

где \hat{F} – эмпирическая функция распределения доходов в выборке. Для размера выборки n в явном виде имеем

$$\hat{\Delta}^\alpha(z) = \frac{1}{n} \sum_{i=1}^n (z - y_i)_+^{\alpha-1}, \quad (13)$$

где y_i – доход индивида i , а $(x)_+$ обозначает $\max(0, x)$.

Поскольку в силу (13), $\hat{\Delta}^\alpha(z)$ – это просто среднее множества IID-переменных, его дисперсию можно оценить как

$$\hat{V} = \frac{1}{n} \sum_{i=1}^n (z - y_i)_+^{2\alpha-2} - \left(\frac{1}{n} \sum_{i=1}^n (z - y_i)_+^{\alpha-1} \right)^2. \quad (14)$$

Тогда подходящей тестовой статистикой для гипотезы $\Delta^\alpha(z) = \Delta_0$ будет

$$t = \frac{\hat{\Delta}^\alpha(z) - \Delta_0}{\hat{V}^{1/2}}.$$

При нулевой гипотезе статистика t асимптотически распределена как $N(0, 1)$, и асимптотическое P -значение для двустороннего теста равно $\tau = 2\Phi(-|t|)$, где $\Phi(\cdot)$ – функция стандартного нормального распределения.

С вероятностью 1 оценка $\hat{\Delta}^\alpha(z)$ не равна Δ_0 . Если статистику t бутстрапировать, используя обычный ресэмплинг данных из исходной выборки, этот факт будет означать, что нарушается Золотое правило 1. Простейший выход из этой ситуации, как замечено после формулировки Золотого правила 1, – заменить нулевую гипотезу, проверяемую с помощью бутстраповской статистики, тестируя то, что является истиной для ресэмплированного DGP, а именно $\Delta^\alpha(z) = \hat{\Delta}^\alpha(z)$. Таким образом, каждая бутстраповская статистика принимает вид

$$t^* = \frac{(\Delta^\alpha(z))^* - \hat{\Delta}^\alpha(z)}{(V^*)^{1/2}}.$$

Здесь $(\Delta^\alpha(z))^*$ – это оценка (13), рассчитанная по бутстраповской выборке, а V^* – рассчитанная по ней оценка дисперсии (14). Золотое правило 1 соблюдается благодаря замене нулевой гипотезы для бутстраповских выборок, но Золотое правило 2 выполнялось бы лучше, если бы как-то можно было наложить настоящую нулевую гипотезу на бутстраповский DGP.

Взвешенный ресэмплинг

Один из способов наложить нулевую гипотезу при бутстрапе с ресэмплингом – ресэмплировать с неравными весами. Обыкновенный ресэмплинг присваивает вес n^{-1} каждому наблюдению, но если различным наблюдениям присваивать различные веса, возможно накладывать разного рода ограничения. Этот подход предложен в Brown & Newey (2002).

Непараметрическая техника, разделяющая многие черты с параметрическим методом максимального правдоподобия – это *метод эмпирического правдоподобия*; см. Owen (2001). В случае IID-выборки эмпирическое правдоподобие – это функция множества неотрицательных вероятностей p_i , $i = 1, \dots, n$, такая что $\sum_{i=1}^n p_i = 1$. Логарифмическая функция эмпирического правдоподобия, с которой проще работать, чем с собственно функцией эмпирического правдоподобия, задается как

$$\ell(\mathbf{p}) = \sum_{i=1}^n \log p_i. \quad (15)$$

Здесь \mathbf{p} обозначает n -мерный вектор вероятностей p_i . Идея заключается в том, чтобы максимизировать (15) при ограничении, что FGT-индекс для выборки с измененными весами равен Δ_0 . А именно, $\ell(\mathbf{p})$ максимизируется при ограничении

$$\sum_{i=1}^n p_i (z - y_i)_+^{\alpha-1} = \Delta_0. \quad (16)$$

При очень маленьком размере выборки возможно, что эта задача на условный максимум не имеет решения с неотрицательными вероятностями. В таком случае статистика *отношения эмпирического правдоподобия* полагается равной ∞ , и нулевая гипотеза отвергается сразу без необходимости бутстрапирования.

В более общем случае, когда задача разрешима, бутстраповский DGP ресэмплирует исходную выборку, подвергая ресэмплингу наблюдение i с вероятностью p_i , а не n^{-1} . Использование метода эмпирического правдоподобия для определения p_i означает, что эти вероятности обладают некоторыми свойствами оптимальности по сравнению с любым другим множеством, удовлетворяющим (16). Золотое правило 2 соблюдается.

Наилучший алгоритм взвешенного ресэмплинга малоизвестен в эконометрическом сообществе. Он описан в Knuth (1998). Вкратце, для набора вероятностей p_i , $i = 1, \dots, n$, строятся

две таблицы по n элементов каждая, содержащие значения q_i , $0 < q_i \leq 1$, и y_i , где y_i – целое число из множества $1, \dots, n$. Чтобы получить индекс j наблюдения, подвергаемого ресэмплингу, случайное число m_i из $U(0, 1)$ используется следующим образом:

$$k_i = \lceil nm_i \rceil, \quad r_i = k_i - nm_i, \quad j = \begin{cases} k_i & \text{если } r_i \leq q_i, \\ y_i & \text{иначе.} \end{cases}$$

За подробностями читатель может обратиться к трактату Кнута.

5 Другие бутстраповские методы

Все бутстраповские DGP, рассматривавшиеся до сих пор, основаны на моделях, в которых либо наблюдения являются IID, либо некоторые величины, которые могут быть оценены по данным, например, ошибки в модели регрессии, являются IID. Но если ошибки в регрессионной модели гетероскедастичны с неизвестной формой гетероскедастичности, нет ничего даже близкого к IID. Конечно, существуют статистики, робастные к гетероскедастичности неизвестной формы, основанные на одном из многочисленных вариантов оценки Эйкера–Уайта ковариационной матрицы, состоятельной при гетероскедастичности (НССМЕ); см. Eicker (1963) и White (1980). Использование НССМЕ приводит к статистикам, которые приблизительно пивотальны для моделей, допускающих гетероскедастичность неизвестной формы.

При бутстрапировании легко соблюдать Золотое правило 1, поскольку как параметрический бутстрап, так и бутстрап с ресэмплингом, обсуждавшийся выше, соответствуют нулевой гипотезе, которая, раз уж допускает гетероскедастичность, обязана допускать и частный случай гомоскедастичности. Но Золотое правило 2 ставит более трудную задачу.

Парный бутстрап

Первое предложение по бутстрапированию моделей с гетероскедастичностью носит различные названия, среди которых (y, X) -бутстрап и парный бутстрап. Этот подход был предложен в Freedman (1981). Вместо ресэмплинга зависимой переменной или остатков, возможно центрированных и нормированных, можно бутстрапировать пары, состоящие из зависимой переменной и набора объясняющих переменных, соответствующие одному наблюдению. Индекс s выбирается случайно из множества $1, \dots, n$, и тогда наблюдение из бутстраповской выборки – это пара (y_s, \mathbf{X}_s) , где \mathbf{X}_s – вектор-строка всех объясняющих переменных для наблюдения s .

Такой бутстрап неявно предполагает, что пары (y_t, \mathbf{X}_t) являются IID при нулевой гипотезе. Хотя это все еще ограничительное предположение, отвергающее любую форму зависимости между наблюдениями, оно допускает любой вид гетероскедастичности y_t условно на \mathbf{X}_t . Объекты ресэмплинга – IID-вытягивания из *совместного* распределения y_t и \mathbf{X}_t .

Предположим, что регрессионная модель имеет вид

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad t = 1, \dots, n, \quad (17)$$

где \mathbf{X}_t – вектор $1 \times k$, а $\boldsymbol{\beta}$ – вектор параметров $k \times 1$. Ошибки u_t могут быть гетероскедастичными, но обязаны иметь нулевое матожидание условно на объясняющих переменных. Таким образом, $\mathbb{E}[y_t | \mathbf{X}_t] = \mathbf{X}_t \boldsymbol{\beta}_0$, если $\boldsymbol{\beta}_0$ – вектор параметров для истинного DGP. Рассмотрим нулевую гипотезу, согласно которой подвектор $\boldsymbol{\beta}$, скажем, $\boldsymbol{\beta}_2$, равен нулю. DGP для парного бутстрапа не соответствует нулевой гипотезе. Следовательно, чтобы выполнялось Золотое правило 1, необходимо модифицировать либо нулевую гипотезу, тестируемую по бутстраповским выборкам, либо сам бутстраповский DGP.

Для эмпирического совместного распределения пар (y_t, \mathbf{X}_t) матожидание первого элемента y условно на втором элементе \mathbf{X} определено только если $\mathbf{X} = \mathbf{X}_t$ для некоторого $t = 1, \dots, n$.

Тогда $\mathbb{E}[y|\mathbf{X} = \mathbf{X}_t] = y_t$. Этот результат не помогает определить, каким может быть истинное значение β или β_2 для бутстраповского DGP. Учитывая это, в качестве истины для бутстраповского DGP обычно берут МНК-оценку $\hat{\beta}_2$ и тестируют гипотезу $\beta_2 = \hat{\beta}_2$ при подсчете бутстраповских статистик.

В работе Flachaire (1999) меняется бутстраповский DGP. Ресэмплируются пары $(\hat{u}_t, \mathbf{X}_t)$, где \hat{u}_t – остатки при МНК-оценивании модели *без ограничений*, возможно, нормированные различными способами. Тогда, если s – целое число, случайно выбранное из множества $1, \dots, n$, y_t^* формируется как

$$y_t^* = \mathbf{X}_{s1}\tilde{\beta}_1 + \hat{u}_s, \quad (18)$$

где β_1 содержит элементы β , не принадлежащие β_2 , а $\tilde{\beta}_1$ – МНК-оценка модели *с ограничением*. Аналогично, \mathbf{X}_{s1} содержит элементы \mathbf{X}_s , коэффициенты при которых являются элементами β_1 . По построению, вектор из \hat{u}_t ортогонален всем векторам, содержащим наблюдения объясняющих переменных. Таким образом, в эмпирическом совместном распределении пар $(\hat{u}_t, \mathbf{X}_t)$ первый элемент, \hat{u} , не коррелирует со вторым элементом, \mathbf{X} . Однако всякое соотношение между дисперсией \hat{u} и объясняющими переменными сохраняется, как при фридмановском парном бутстрапе. Кроме того, бутстраповский DGP (18) теперь удовлетворяет нулевой гипотезе в исходной постановке.

Дикий бутстрап

Модель при нулевой гипотезе, на которой основана любая форма парного бутстрапа, постулирует совместное распределение зависимой переменной y и объясняющих переменных. Если предполагается, что объясняющие переменные экзогенны, обычной практикой является расчет статистик и их распределений условно на регрессорах. Один из способов это сделать – использовать так называемый *дикий бутстрап*; см. Wu (1986), Liu (1988), Mammen (1993) и Davidson & Flachaire (2001).

Для регрессионной модели (17) DGP для дикого бутстрапа имеет вид

$$y_t^* = \mathbf{X}_t\tilde{\beta} + s_t^*\tilde{u}_t \quad (19)$$

где $\tilde{\beta}$, как обычно, – МНК-оценка регрессионных параметров при ограничении, а \tilde{u}_t – остатки от МНК-оценивания при ограничении. Заметим, что в данном случае ресэмплинг отсутствует; как объясняющие переменные, так и остаток для бутстраповского наблюдения t поступают от наблюдения t в исходной выборке. Вводятся новые случайные элементы s_t^* , которые являются IID-вытягиваниями из распределения с матожиданием 0 и дисперсией 1.

Бутстраповский DGP легко удовлетворяет Золотому правилу 1: поскольку s_t^* и \tilde{u}_t независимы, так как последние порождаются реальным DGP, а первые – генератором случайных чисел, матожидание бутстраповских ошибок $s_t^*\tilde{u}_t$ равно 0. Условно на остатке \tilde{u}_t дисперсия $s_t^*\tilde{u}_t$ равна \tilde{u}_t^2 . Если остаток принимается в качестве аппроксимации для ненаблюдаемой ошибки u_t , матожидание \tilde{u}_t^2 является истинной дисперсией u_t , и этот факт сильно способствует выполнению Золотого правила 2.

Долгое время наиболее часто применяемым распределением для s_t^* было следующее двухточечное распределение:

$$s_t^* = \begin{cases} -(\sqrt{5}-1)/2 & \text{с вероятностью } (\sqrt{5}+1)/(2\sqrt{5}), \\ (\sqrt{5}+1)/2 & \text{с вероятностью } (\sqrt{5}-1)/(2\sqrt{5}). \end{cases}$$

которое было предложено в Mammen (1993). Более простое двухточечное распределение – это *распределение Радемахера*

$$s_t^* = \begin{cases} -1 & \text{с вероятностью } \frac{1}{2}, \\ 1 & \text{с вероятностью } \frac{1}{2}. \end{cases} \quad (20)$$

Davidson & Flachaire (2001) предлагают это более простое распределение, которое оставляет неизменной абсолютную величину каждого остатка в бутстраповском DGP, но придает ему произвольный знак. Авторы показывают с помощью симуляций, что выбор (20) часто приводит к более надежной инференции с помощью бутстрапа, чем другие альтернативы.

Векторные авторегрессии

Потенциальная проблема фридмановского парного бутстрапа состоит в том, что он обращается со всеми переменными, эндогенными и экзогенными, одинаково. Некоторые модели, однако, имеют более одной эндогенной переменной и, таким образом, за исключением частных случаев, когда правомерно действовать условно на некоторых из них, бутстраповский DGP должен порождать все эндогенные переменные одновременно. Это совсем несложно для таких моделей, как векторные авторегрессии (VAR). Типичная VAR-модель записывается как

$$\mathbf{Y}_t = \sum_{i=1}^p \mathbf{Y}_{t-i} \Pi_i + \mathbf{X}_t \mathbf{B} + \mathbf{U}_t, \quad t = p+1, \dots, n. \quad (21)$$

Здесь \mathbf{Y}_t и \mathbf{U}_t – вектора $1 \times m$, Π_i все являются матрицами $m \times m$, \mathbf{X}_t – вектор $1 \times k$, а \mathbf{B} – матрица $k \times m$. m элементов \mathbf{Y}_t – эндогенные переменные для наблюдения t . Элементы \mathbf{X}_t – экзогенные объясняющие переменные, хотя некоторые VAR-модели обходятся без экзогенных переменных, и в таких случаях $k = 0$. Элементы матриц Π_i , $i = 1, \dots, p$ и \mathbf{B} являются параметрами модели. Вектора \mathbf{U}_t имеют нулевое матожидание и обычно предполагаются взаимно независимыми, хотя межэлементно коррелированными; независимые элементы *одномоментной ковариационной матрицы* Σ размерности $m \times m$ также являются параметрами модели.

Среди гипотез, которые можно тестировать в рамках модели (21), – тесты на *причинность по Грэнжеру*; см. Granger (1969) или Davidson & MacKinnon (2004) с хрестоматийным описанием. Нулевая гипотеза для этих тестов – это *отсутствие причинности* по Грэнжеру, что налагает нулевые ограничения на подмножества элементов Π_i . При отсутствии ограничений модель (21) можно эффективно оценить, применив МНК к каждому уравнению по отдельности и взяв в качестве оценки ковариационной матрицы Σ эмпирическую ковариационную матрицу остатков. При наличии ограничений модель обычно оценивают с помощью метода максимального правдоподобия при предположении о совместной нормальности ошибок.

Бутстраповские DGP можно формировать для моделей, которые накладывают ограничения различного уровня. Во всех случаях матрицы Π_i , Σ и \mathbf{B} , если они присутствуют, следует полагать равными их оценкам при ограничениях. Кроме того, во всех случаях бутстраповские выборки следует рассматривать условно на первых p наблюдениях из исходной выборки, если не предполагается стационарность, когда первые p наблюдений каждой бутстраповской выборки следует вытягивать из стационарного распределения p последовательных m -мерных векторов $\mathbf{Y}_t, \dots, \mathbf{Y}_{t+p-1}$. Если предполагается нормальность ошибок, бутстраповские ошибки можно генерировать как IID-реализации многомерного $N(\mathbf{0}, \tilde{\Sigma})$ -распределения – необходимо получить декомпозицией Холецки $(m \times m)$ -матрицу \mathbf{A} , такую, что $\mathbf{A}\mathbf{A}' = \tilde{\Sigma}$, и сгенерировать \mathbf{U}_t^* как $\mathbf{A}\mathbf{V}_t^*$, где m элементов \mathbf{V}_t^* – случайные стандартные нормальные IID-величины. Если нежелательно предполагать нормальность, то можно ресэмплировать *вектора* остатков модели с ограничениями $\tilde{\mathbf{U}}_t$. Если нежелательно предполагать даже то, что \mathbf{U}_t являются IID, можно использовать дикий бутстрап, при котором каждый из векторов $\tilde{\mathbf{U}}_t$ умножается на скаляр s_t^* , IID-реализацию распределения с нулевым матожиданием и единичной дисперсией.

Одновременные уравнения

Ситуация оказывается немного сложнее для *модели одновременных уравнений*, в которой эндогенные переменные для данного наблюдения определяются как решение системы одновременных уравнений, также включающих экзогенные объясняющие переменные. Лагированные значения эндогенных переменных также могут выступать в качестве объясняющих переменных; они называются *предопределенными*. В их присутствии бутстраповский DGP надо основывать на рекурсивных симуляциях.

Модель одновременных уравнений можно записать как

$$\mathbf{Y}_t \Gamma = \mathbf{W}_t \mathbf{B} + \mathbf{U}_t, \quad (22)$$

где \mathbf{Y}_t и \mathbf{U}_t – вектора $1 \times m$, \mathbf{W}_t – вектор экзогенных или предопределенных объясняющих переменных $1 \times k$, Γ – матрица $m \times m$, а \mathbf{B} – матрица $k \times m$. Элементы Γ и \mathbf{B} , наряду с независимыми элементами одномоментной ковариационной матрицы Σ , являются параметрами модели. Чтобы эндогенные переменные \mathbf{Y}_t задавались с помощью (22), Γ должна быть невырожденной.

Система уравнений (22) называется *структурной формой* модели. *Приведенная форма* получается в результате решения уравнений структурной формы:

$$\mathbf{Y}_t = \mathbf{W}_t \mathbf{B} \Gamma^{-1} + \mathbf{V}_t, \quad (23)$$

где одномоментная ковариационная матрица \mathbf{V}_t равна $(\Gamma')^{-1} \Sigma \Gamma^{-1}$. Приведенную форму можно оценить без ограничений, применяя МНК отдельно к каждому уравнению системы

$$\mathbf{Y}_t = \mathbf{W}_t \Pi + \mathbf{V}_t,$$

где Π – матрица параметров $k \times m$. Однако часто структурная форма является *сверхидентифицированной*, когда на матрицы Γ и \mathbf{B} накладываются ограничения. Это имеет место, если нулевая гипотеза накладывает такие ограничения. Существует множество приемов оценивания при ограничениях любой из эквивалентных моделей (22) или (23). Когда используется стандартная асимптотика, асимптотическая эффективность достигается при использовании двух методов, *трехшагового метода наименьших квадратов* (ЗШМНК), и метода максимального правдоподобия при полной информации (ММППИ). Эти стандартные методы описаны в большинстве учебников по эконометрике.

Нелинейные модели

Бутстраповские DGP должны во всех случаях использовать эффективные оценки параметров при ограничениях, полученные с помощью ЗШМНК или ММППИ, с небольшим предпочтением ММППИ, имеющего свойства оптимальности более высокого порядка, не достигаемые при применении ЗШМНК. Бутстраповские ошибки можно генерировать из многомерного нормального распределения, с помощью ресэмплинга векторов остатков модели при ограничениях или с помощью процедуры дикого бутстрапа. См. Davidson & MacKinnon (2006b) с подробным обсуждением.

Бутстрапирование часто считается вычислительно интенсивной процедурой, хотя при нынешних способностях компьютеров и программного обеспечения это редко когда является серьезной проблемой на практике. Модели, требующие нелинейного оценивания, могут быть исключением для данного утверждения, поскольку алгоритмы, применяемые при нелинейном оценивании, могут не сходиться при малом числе итераций. Если это случается при оценивании модели на реальных данных, проблема не связана с бутстрапом, а возникает из-за взаимосвязи между моделью и данными. Проблема для бутстрапирования возникает, если процедура оценивания, которая работает на исходных данных, не срабатывает для одной или нескольких бутстраповских выборок.

В принципе нелинейное оценивание в контексте бутстрапа должно быть скорей проще, чем наоборот. Известен истинный бутстраповский DGP, и можно использовать истинные параметры для этого DGP в качестве начальных значений для итеративной процедуры, применяемой при нелинейном оценивании. В тех случаях, когда необходимо оценить две модели, с ограничением и без, можно использовать оценки модели, скажем, с ограничением в качестве начальных значений для оценивания модели без ограничений, используя таким образом специфические свойства конкретной бутстраповской выборки.

Когда любая нелинейная процедура повторяется тысячи раз, кажется, что все, что может пойти не так, пойдет не так по крайней мере один раз. В большинстве случаев аргументы из предыдущего абзаца применимы, но не всегда. Любая итеративная процедура может заикнуться, если она не сходится, со всякого рода нежелательными последствиями. Следовательно, хорошей практикой является установление умеренной верхней границы для числа разрешаемых итераций для каждой бутстраповской выборки.

Во многих случаях верхнюю границу в 3 или 4 итерации можно теоретически обосновать. Асимптотическая теория обычно дает *скорость сходимости* бутстраповского расхождения к нулю относительно размера выборки. Она также дает скорость сходимости метода Ньютона или квази-ньютоновского метода, применяемых в алгоритме оценивания. Если бутстраповское расхождение сходится к нулю как, скажем, $n^{-3/2}$, мало смысла в поиске численной точности с лучшей скоростью сходимости. Для большинства квази-ньютоновских методов, например алгоритма Гаусса–Ньютона, на каждой итерации уменьшается расстояние между текущими параметрами алгоритма и теми, к которым он сходится (в предположении, что он сходится), в $n^{-1/2}$ раз. Обычно возможно инициализировать алгоритм с параметрами, отличающимися от финальных на величину порядка $n^{-1/2}$. Через три итерации разница уже имеет порядок n^{-2} , более низкий, чем для бутстраповского расхождения. Таким образом, достигается тот же порядок точности, как тот, который был бы достижим, если бы итерации продолжались до сходимости согласно некоторому более строгому критерию. Поскольку бутстраповская инференция основана на среднем по бутстраповским повторам, этого достаточно для большинства целей.

Дальнейшие подробности по поводу идеи об ограничении числа итераций можно найти в Davidson & MacKinnon (1999), где также отмечено, что численные методы для подсчета статистик отношения правдоподобия сходятся еще быстрее, чем применяемые для статистик Вальда или множителей Лагранжа. Подробное описание асимптотической теории, стоящей за этой идеей, можно найти в Andrews (2002).

6 Доверительные множества

Справедливо будет сказать, что в статистической литературе по бутстрапу наибольшие усилия ушли на разработку бутстраповских методов построения доверительных интервалов, или, более общо, доверительных множеств. Инференция, основанная на доверительных множествах, в принципе эквивалентна инференции на основе тестирования гипотез, но на практике могут существовать препятствия для этой теоретической эквивалентности. Общее правило состоит в том, что стандартные методы тестирования гипотез с помощью бутстрапа работают лучше, чем стандартные методы построения бутстраповских доверительных множеств.

Процентильные методы

Рассмотрим случай инференции о скалярном параметре θ . Предположим, что для каждого возможного значения параметра существует тестовая статистика $\tau(\theta)$ с известным, хотя бы асимптотически, распределением, когда θ является истинным параметром. Доверительный

интервал при *уровне доверия* $1 - \alpha$, $0 < \alpha < 1$, – это множество значений θ , для которых гипотеза о том, что θ – истинный параметр, не отвергается на уровне значимости α с помощью теста на основе статистики $\tau(\theta)$. Пусть C обозначает таким образом найденный доверительный интервал. Тогда, если истинный параметр равен θ , имеем, возможно лишь приближенно, что

$$\mathbb{P}[\theta \in C] = 1 - \mathbb{P}[\tau(\theta) \in \text{Rej}(\alpha)] = 1 - \alpha, \quad (24)$$

где $\text{Rej}(\alpha)$ – область отвержения для теста на уровне значимости α . Такие обозначения используются для возможности рассмотрения случаев одно- и двусторонних доверительных интервалов, возникающих из одно- и двусторонних тестов, соответственно.

В обратную сторону, если для каждого уровня доверия $1 - \alpha$ имеется доверительный интервал C_α , тест на уровне значимости α отвергает гипотезу о том, что θ – истинный параметр, тогда и только тогда, когда $\theta \notin C_\alpha$. P -значение для этой гипотезы можно определить с помощью соотношения

$$P(\theta) = \max\{\alpha | \theta \in C_\alpha\}.$$

Весьма прямолинейный способ получения доверительного интервала для θ называется *процентильным методом*. Предполагается модель \mathbb{M} , и для каждого DGP $\mu \in \mathbb{M}$ существует соответствующий параметр $\theta(\mu)$, «истинный» параметр для DGP μ . Поскольку обычно имеются иные параметры помимо θ , их необходимо оценить вместе с θ , чтобы сформировать бутстраповский DGP. Для настоящего анализа не является необходимым различать возможные бутстраповские DGP; в любом случае все они используют оцененные параметры. Первый шаг состоит в получении оценок всех параметров для модели \mathbb{M} , которые должны быть настолько эффективными, насколько это возможно.

Рассмотрим сначала случай *равнохвостового* доверительного интервала. Пусть $q_{\alpha/2}$ и $q_{1-\alpha/2}$ – $\alpha/2$ и $(1 - \alpha/2)$ -квантили распределения $\hat{\theta} - \theta$, где θ – истинный параметр. Тогда ясно, что

$$\mathbb{P}[q_{\alpha/2} \leq \hat{\theta} - \theta \leq q_{1-\alpha/2}] = \alpha.$$

Эти неравенства эквивалентны

$$\hat{\theta} - q_{1-\alpha/2} \leq \theta \leq \hat{\theta} - q_{\alpha/2},$$

а отсюда ясно, что доверительный интервал с нижней границей $\hat{\theta} - q_{1-\alpha/2}$ и верхней границей $\hat{\theta} - q_{\alpha/2}$ содержит истинное значение θ с вероятностью α .

На следующем шаге генерируется множество бутстраповских выборок и для каждой из них рассчитывается оценка параметра, скажем θ^* . Поскольку истинное значение θ для бутстраповского DGP – это $\hat{\theta}$, можно использовать распределение $\theta^* - \hat{\theta}$ в качестве оценки распределения $\hat{\theta} - \theta$. В частности, $\alpha/2$ и $(1 - \alpha/2)$ -квантили распределения $\theta^* - \hat{\theta}$, скажем $q_{\alpha/2}^*$ и $q_{1-\alpha/2}^*$, дают процентильный доверительный интервал

$$C_\alpha^* = [\hat{\theta} - q_{1-\alpha/2}^*, \hat{\theta} - q_{\alpha/2}^*].$$

В качестве одностороннего доверительного интервала, не ограниченного справа, берется $[\hat{\theta} - q_{1-\alpha}^*, \infty)$, а в качестве не ограниченного слева $(-\infty, \hat{\theta} - q_\alpha^*]$. Отметим несколько контринтуитивный факт, что *верхний* квантиль распределения определяет *нижнюю* границу доверительного интервала и наоборот.

Процентильный интервал очень далек от наилучшего бутстраповского доверительного интервала. Первая причина в том, что почти во всех интересных случаях случайная величина $\hat{\theta} - \theta$ не является пивотальной даже асимптотически. Действительно, стандартная асимптотика дает предельное распределение $N(0, \sigma_\theta^2)$ с некоторой асимптотической дисперсией σ_θ^2 .

Если только $\sigma_{\hat{\theta}}^2$ не является постоянной для всех DGP из \mathbb{M} , то $\hat{\theta} - \theta$ не является асимптотически пивотальной.

По этой причине более популярным бутстраповским доверительным интервалом является *t-процентильный* интервал. Теперь предположим, что можно оценить дисперсию $\hat{\theta}$, таким образом, построить доверительный интервал на основе *стьюдентизированной* величины $(\hat{\theta} - \theta)/\hat{\sigma}_{\theta}$, которая во многих обстоятельствах является асимптотически нормальной, и, следовательно, асимптотически пивотальной. Пусть $q_{\alpha/2}$ и $q_{1-\alpha/2}$ – нужные квантили распределения $(\hat{\theta} - \theta)/\hat{\sigma}_{\theta}$, когда истинный параметр – это θ . Тогда

$$\mathbb{P} \left[q_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\hat{\sigma}_{\theta}} \leq q_{1-\alpha/2} \right] = \alpha.$$

Если квантили оцениваются с помощью квантилей распределения $(\theta^* - \hat{\theta})/\sigma_{\theta}^*$, где σ_{θ}^* – квадратный корень из оценки дисперсии, рассчитанной по бутстраповской выборке, получим *t-процентильный доверительный интервал*

$$C_{\alpha}^* = [\hat{\theta} - \hat{\sigma}_{\theta} q_{1-\alpha/2}^*, \hat{\theta} - \hat{\sigma}_{\theta} q_{\alpha/2}^*]. \quad (25)$$

Во многих случаях *t-процентильный интервал* работает гораздо лучше процентильного интервала. Для более полного обсуждения бутстраповских доверительных интервалов подобного рода см. Hall (1992).

Равнохвостовые доверительные интервалы не единственные, которые можно построить, используя процентильные или *t-процентильные* методы. Вспомним, что критические значения для тестов на уровне значимости α могут быть основаны на β и γ -квантилях для нижнего и верхнего критических значений, при условии что $1 - \gamma + \beta = \alpha$. Бутстраповское распределение редко бывает симметричным вокруг своей центральной точки (если оно не построено таким образом намеренно). Величины β и γ , минимизирующие расстояние между β -квантилем и γ -квантилем при ограничении $1 - \gamma + \beta = \alpha$, тогда не равны $\alpha/2$ и $1 - \alpha/2$, вообще говоря. Использование β и γ , полученных таким образом, ведет к *кратчайшему доверительному интервалу* при уровне значимости $1 - \alpha$.

Построение *t-процентильного интервала* следует правилу, по которому строится доверительное множество C для (24). Статистика $\tau(\theta)$ принимает вид $(\hat{\theta} - \theta)/\hat{\sigma}_{\theta}$, и область отвержения $\text{Rej}(\alpha)$ определяется из бутстраповского распределения $(\theta^* - \hat{\theta})/\sigma_{\theta}^*$. Золотое правило 1 соблюдается, поскольку бутстраповская статистика тестирует гипотезу, являющуюся истиной для бутстраповского DGP, а именно, что $\theta = \hat{\theta}$.

Доверительный интервал принимает простой вид (25) только потому, что тестовая статистика – простая функция от θ . Однако эта простота может повлечь издержки. Статистика $(\hat{\theta} - \theta)/\hat{\sigma}_{\theta}$ – это статистика Вальда, а известно, что статистика Вальда может иметь нежелательные свойства. Наихудшее из них заключается в том, что такие статистики неинвариантны к нелинейной перепараметризации. Например, если определить новый параметр ϕ соотношением $\phi = h(\theta)$, где h – монотонно возрастающая нелинейная функция, то доверительный интервал, основанный на вальдовской статистике $(\hat{\phi} - \phi)/\hat{\sigma}_{\phi}$, отличается от основанного на статистике $(\hat{\theta} - \theta)/\hat{\sigma}_{\theta}$. Аналогично, при данной нулевой гипотезе $\theta = \theta_0$, или, что эквивалентно, $\phi = \phi_0 = h(\theta_0)$, тест, бутстраповский или иной, на основе одной статистики может отвергать нулевую гипотезу, а на основе другой нет. См. анализ этого явления в Gregory & Veall (1985) и Lafontaine & White (1986).

Доверительные интервалы на основе статистик лучше

Доверительное множество не обязательно строить на основе теста Вальда. Предположим, что $\tau(\theta)$ – статистика отношения правдоподобия, или множителей Лагранжа, которая тестирует

гипотезу о том, что θ – истинное значение параметра. Эти статистики можно сделать инвариантными к перепараметризации. Они часто асимптотически распределены как хи-квадрат и, значит, отвергают нулевую гипотезу при больших значениях статистики. Доверительное множество C_α с номинальным уровнем доверия $1 - \alpha$ определяется как обычно:

$$C_\alpha = \{\theta | \tau(\theta) > q_{1-\alpha}\}, \quad (26)$$

где $q_{1-\alpha}$ – $(1 - \alpha)$ -квантиль номинального распределения, используемого для определения критической области – хи-квадрат распределения с соответствующим числом степеней свободы для теста на основе асимптотики или распределения, полученного бутстрапированием. Граничные точки доверительного множества (26) тогда удовлетворяют уравнению $\tau(\theta) = q_{1-\alpha}$.

В общем случае может быть затруднительно или даже невозможно получить аналитическое выражение для $\tau(\theta)$. Если это так, уравнение $\tau(\theta) = q_{1-\alpha}$, возможно, приходится решать с помощью численных методов. В регулярных случаях это уравнение имеет ровно два решения, нижнюю и верхнюю граничные точки доверительного интервала. В менее приятных случаях это уравнение может задавать неограниченный интервал или даже объединение несвязных интервалов, любой из которых может быть неограниченным. Доверительные множества, не являющиеся ограниченными интервалами, описаны в Dufour (1997).

Совместные доверительные области для более чем одного параметра также можно определить с помощью (26), интерпретируя θ как вектор. Вальдовские статистики порождают эллипсоидные области, которые легко описать. Но другие типы статистик могут давать доверительные области более сложных форм. В принципе нет разницы между доверительными множествами, основанными на асимптотике и основанными на бутстрапе. Во всех случаях номинальное распределение дает квантиль или квантили, характеризующие доверительную область.

Статистика Вальда по определению основана на оценивании модели при *альтернативной* гипотезе. Этот факт плохо вяжется с Золотым правилом 2. Но даже если использовать статистику множителей Лагранжа, основанную на оценивании модели при нулевой гипотезе, можно утверждать, что Золотое правило 2 также не выполняется. Проблема состоит в том, что для построения доверительного множества в принципе необходимо рассматривать бесконечное число нулевых гипотез; в (26) θ может изменяться в пределах открытого интервала, часто являющегося всей действительной прямой. На практике, при уверенности в том, что доверительное множество единственно, связно и является интервалом, достаточно найти два значения θ , удовлетворяющие $\tau(\theta) = q_{1-\alpha}$.

Соблюдение Золотого правила 2

Золотое правило 2 не соблюдается из-за предположения о том, что распределение $\tau(\theta)$ при DGP, для которого θ – истинный параметр, одинаково для всех θ . Если это так, то статистика называется пивотальной и дальнейших проблем нет. Но если статистика является лишь приблизительно пивотальной, ее распределение, когда истинное значение θ является граничной точкой доверительного интервала, не такое же, как в случае, когда истинный параметр – точечная оценка $\hat{\theta}$. Ведь истинный параметр для бутстраповского DGP – это $\hat{\theta}$.

Чтобы Золотое правило 2 полностью соблюдалось, уравнение, которое следует решить для поиска граничных точек доверительного интервала, должно выглядеть как

$$\tau(\theta) = q_{1-\alpha}(\theta), \quad (27)$$

где $q_{1-\alpha}(\theta)$ – $(1 - \alpha)$ -квантиль распределения $\tau(\theta)$, когда θ является истинным параметром. Если θ является единственным параметром, то возможно, хотя обычно и нелегко, решить (27) численно на основе симуляций. В общем случае, однако, ситуация еще сложнее. Если

помимо θ имеются другие параметры, которые в данном контексте можно назвать шумовыми, то, согласно Золотому правилу 2, следует использовать наилучшую возможную оценку этих параметров при нулевой гипотезе для бутстраповского DGP. То есть для каждого значения θ , рассматриваемого при поиске решения (27), следует переоценивать эти мешающие параметры при ограничении, что θ – истинный параметр и затем основывать бутстраповский DGP на θ и этих оценках при ограничении. Этот принцип лежит в основе так называемого *сеточного бутстрапа*, предложенного в работе Hansen (1999). Неудивительно, что этот метод является вычислительно очень сложным, но Хансен показывает, что он дает удовлетворительные результаты для авторегрессионной модели, когда другие бутстраповские доверительные интервалы дают ненадежную инференцию.

7 Заключительные замечания

Бутстрап является статистическим методом, способным обеспечить надежную инференцию в широком классе эконометрических моделей. В настоящем эссе основное внимание уделено инференции на основе бутстрапа. Хотя бутстрап можно использовать для многих других целей, инференция, в форме тестирования гипотез или построения доверительных множеств, является сферой, в которой применение бутстрапа имеет наиболее яркую пользу в эконометрической практике.

В настоящем эссе дается лишь набросок многочисленных способов применения бутстрапа в эконометрике. Ничего не говорится о трудных проблемах бутстрапирования зависимых данных или сложностях, возникающих при распределениях с тяжелыми хвостами. Оба эти вопроса в настоящее время являются активными областями исследований, и, надо надеяться, в ближайшем будущем их понимание улучшится.

Ясно, что теоретическое освоение бутстрапа все еще неполно. Многие симуляционные эксперименты показали, что бутстрап часто дает лучшие результаты, чем предсказывают имеющиеся теории. Несмотря на это, есть некоторые принципы, претенциозно сформулированные в настоящей работе как Золотые правила, которые могут помочь обеспечить надежную инференцию на основе бутстрапа. Эти правила отражают тот факт, что для инференции требуется как можно более точная характеристика распределения тестовой статистики, на которой основана инференция, при тестируемой нулевой гипотезе.

Прошло время с тех пор, как Beran (1988) отметил, что бутстрап обеспечивает более надежную инференцию, когда применяется для приблизительно пивотальных величин. На практике статистики, которые предположительно являются приблизительно пивотальными, могут иметь распределения, сильно зависящие от шумовых параметров. В некоторых случаях найти приблизительно пивотальные величины затруднительно. Бутстрап тем не менее может «срабатывать» даже в таких случаях, но не стоит ожидать, что он будет столь же надежен, как при более благоприятных обстоятельствах. Соблюдение Золотых правил, описанных в настоящем эссе, может увеличить надежность инференции даже в таких случаях.

Список литературы

- Andrews, D.W.K. (2002). Higher-order improvements of a computationally attractive k -step bootstrap for extremum estimators. *Econometrica* 70, 119–162.
- Athreya, K.B. (1987). Bootstrap of the mean in the infinite variance case. *Annals of Statistics* 15, 724–731.
- Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of American Statistical Association* 83, 687–697.
- Brown, B.W. and W. Newey (2002). Generalized method of moments, efficient bootstrapping, and improved inference. *Journal of Business & Economic Statistics* 20, 507–517.
- Davidson, R. & E. Flachaire (2001). The wild bootstrap, tamed at last. GREQAM Document de Travail 99A32.

- Davidson, R. & J.G. MacKinnon (1999). Bootstrap testing in nonlinear models. *International Economic Review* 40, 487–508.
- Davidson R. & J.G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford: Oxford University Press.
- Davidson R. & J.G. MacKinnon (2006a). Bootstrap methods in econometrics. Chapter 23 of *Palgrave Handbook of Econometrics*, Volume 1, *Econometric Theory*, eds T.C. Mills & K. Patterson. London: Palgrave-Macmillan.
- Davidson R. & J.G. MacKinnon (2006b). Bootstrap inference in a linear equation estimated by instrumental variables. Discussion Paper 1024, Queen's University.
- Dufour, J.-M., (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1387.
- Dufour, J.-M. & L. Khalaf (2001). Monte Carlo test methods in econometrics. Ch. 23 in *A Companion to Econometric Theory*, ed. B. Baltagi. Oxford: Blackwell Publishers, 494–519.
- Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics* 28, 181–187.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1–26.
- Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics* 34, 447–456.
- Flachaire, E. (1999). A better way to bootstrap pairs. *Economics Letters*, 64, 257–262.
- Foster, J.E., J. Greer & E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica* 52, 761–776.
- Freedman, D.A. (1981). Bootstrapping regression models. *Annals of Statistics* 9, 1218–1228.
- Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics* 31, 1208–11.
- Godambe, V.P. & M.E. Thompson (1978). Some aspects of the theory of estimating equations. *Journal of Statistical Planning and Inference* 2, 95–104.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Gregory, A.W. & M.R. Veall (1985). On formulating Wald tests for nonlinear restrictions. *Econometrica* 53, 1465–1468.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Hansen, B.E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics* 81, 594–607.
- Horowitz, J.L. (2001). The bootstrap. Ch. 52 in *Handbook of Econometrics* Vol. 5, eds J.J. Heckman & E.E. Leamer, Amsterdam: North-Holland, 3159–3228.
- Horowitz, J.L. (2003). The bootstrap in econometrics. *Statistical Science* 18, 211–218.
- Hu, F. & J.D. Kalbfleisch (2000). The estimating function bootstrap. *Canadian Journal of Statistics* 28, 449–481.
- Knuth, D.E. (1998). *The Art of Computer Programming*, Vol 2, *Seminumerical Algorithms*, 3rd edition. Addison-Wesley.
- Lafontaine, F. & K.J. White (1986). Obtaining any Wald statistic you want. *Economics Letters* 21, 35–40.
- Liu, R.Y. (1988). Bootstrap procedures under some non-I.I.D. models. *Annals of Statistics* 16, 1696–1708.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255–285.
- Owen, A.B. (2001). *Empirical Likelihood*. Chapman and Hall.
- Politis, D.N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science* 18, 219–230.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.
- Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1261–1295.

Bootstrapping econometric models

Russell Davidson

*McGill University and CIREQ, Canada
GREQAM, France*

The bootstrap is a statistical technique used more and more widely in econometrics. While it is capable of yielding very reliable inference, some precautions should be taken in order to ensure this. Two “Golden Rules” are formulated that, if observed, help to obtain the best the bootstrap can offer. Bootstrapping always involves setting up a bootstrap data-generating process (DGP). The main types of bootstrap DGP in current use are discussed, with examples of their use in econometrics. The ways in which the bootstrap can be used to construct confidence sets differ somewhat from methods of hypothesis testing. The relation between the two sorts of problem is discussed.

Keywords: bootstrap, hypothesis test, confidence set

JEL Classification: C10, C12, C15

Бутстрап-схемы для временных рядов^{*}

Питер Бюльман[†]

ETH Zürich, Цюрих, Швейцария

В настоящем эссе представлен обзор и сравнение схем блочного, решетчатого и локального бутстрапа для временных рядов, а также рассмотрены их теоретические свойства и поведение в конечных выборках. Материал подобран с целью дать новую и объективную картину некоторых аспектов бутстрапирования временных рядов. Универсальность блочного бутстрапа противопоставляется решетчатому бутстрапу. Обсуждаются преимущества и недостатки реализации бутстрап-схем на практике, и утверждается, что решетчатый бутстрап часто превосходит блочный метод. Локальный бутстрап, предназначенный для непараметрического сглаживания, легок в реализации, но иногда работает плохо.

1 Введение

Бутстрапирование можно рассматривать как симуляции статистики или статистической процедуры на основе оцененного распределения \hat{P}_n наблюдаемых данных X_1, \dots, X_n . В случае зависимых наблюдений объект \hat{P}_n является более сложным и значительно менее очевидным, чем в основополагающей статье Эфрона (Efron, 1979) для случая независимых данных. Здесь в основном рассматриваются методы блочного, решетчатого и локального бутстрапирования, которые в определенном смысле являются непараметрическими и не привязанными к какой-либо модели. Целью является получение четкой картины о сильных и слабых сторонах различных способов бутстрапирования временных рядов. Для этого исследуются как их теоретические свойства, так и их поведение в конечных выборках. До сих пор в литературе очень мало внимания уделялось общей перспективе при сравнении различных бутстрап-схем. В этом отношении настоящий *избирательный* подход дает ценное и новое понимание вопроса и отличает наш сравнительный анализ от работ Léger, Politis & Romano (1992), Efron & Tibshirani (1993, гл. 8.5–8.6), Shao & Tu (1995, гл. 9), Li & Maddala (1996) или Davison & Hinkley (1997, гл. 8).

Получение информации из данных формализуется с помощью оценки $\hat{\theta}$, скалярной, векторной или в пространстве кривых. Для инференции необходимо оценить выборочное распределение оценки $\hat{\theta}$ или ее пивотизированной или студентизированной версии. Для временных рядов данная задача гораздо сложнее, чем в случае независимых наблюдений, и методы, основанные на аналитических выкладках, быстро сталкиваются с затруднениями. Рассмотрим, например, оценку $\hat{\theta}$, асимптотически нормально распределенную вокруг конечномерного параметра θ : при определенных условиях и предположении о стационарности ряда данных X_1, \dots, X_n ,

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, \sigma_\infty^2) \text{ при } n \rightarrow \infty. \quad (1)$$

В отличие от случая независимых наблюдений асимптотическая дисперсия σ_∞^2 является бесконечномерным объектом, представляющим собой бесконечную сумму ковариаций, которая,

^{*}Перевод Б. Гершмана и С. Анатольева. Эссе является сокращенной и измененной версией статьи Bühlmann (2002). Цитировать как: Бюльман, Питер (2007) «Бутстрап-схемы для временных рядов», Квантиль, №3, стр. 37–56. Citation: Bühlmann, Peter (2007) “Bootstrap schemes for time series,” Quantile, No.3, pp. 37–56.

[†]Адрес: Seminar für Statistik, ETH Zürich, CH-8092 Zürich, Switzerland. Электронная почта: buhlmann@stat.math.ethz.ch

вообще говоря, не поддается оцениванию со скоростью сходимости $1/\sqrt{n}$. В простом случае, когда $\hat{\theta} = \bar{X}_n$,

$$\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} \mathbb{C}[X_0, X_k].$$

Асимптотической дисперсией в этом примере является (нормированное на 2π) значение функции спектральной плотности процесса, порождающего данные, в нуле. В другом примере, когда $\hat{\theta} = \text{med}(X_1, \dots, X_n)$,

$$\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} \mathbb{C}[\text{IF}(X_0), \text{IF}(X_k)], \quad \text{IF}(x) = \frac{\text{sign}(x - \theta)}{2f(\theta)},$$

где $\theta = F^{-1}(1/2)$ – медиана маржинального распределения F для X_t , имеющего функцию плотности f . Здесь фигурирует спектральная плотность процесса с функциями влияния $(\text{IF}(X_t))_{t \in \mathbb{Z}}$, то есть мгновенное неизвестное преобразование процесса $(X_t)_{t \in \mathbb{Z}}$. Было бы очень неудобно оценивать неизвестную функцию плотности f и θ для получения оценки $\text{IF}(\cdot)$, а затем ее спектральной плотности. Преимущество бутстрапирования в том, что оно позволяет состоятельно оценивать асимптотическую дисперсию и распределение $\sqrt{n}(\hat{\theta} - \theta)$ *автоматически*.

Состоятельность, или «точность первого порядка», заключается в требовании состоятельного оценивания предельного распределения $\hat{\theta}$. Точнее говоря, для оценки $\hat{\theta}$ из пространства \mathbb{R}^q

$$\sup_{x \in \mathbb{R}^q} |\mathbb{P}^*[a_n(\hat{\theta}^* - \theta^*) \leq x] - \mathbb{P}[a_n(\hat{\theta} - \theta) \leq x]| = o_P(1) \quad (n \rightarrow \infty), \quad (2)$$

где $a_n(\hat{\theta} - \theta)$ сходится к невырожденному предельному распределению. Символ ‘ \leq ’ определен покомпонентно, и, как обычно, звездочка * обозначает бутстраповский аналог. Центральное значение θ^* , являющееся константой условно на исходных наблюдениях X_1, \dots, X_n , обычно не выбирается равным $\hat{\theta}$, как в случае Эфроновского бутстрапа для случая независимых одинаково распределенных данных; подробности даны ниже при обсуждении конкретных методов бутстрапирования временных рядов.

Например, если $\hat{\theta}$ является выборочным средним или медианой, предельное распределение имеет вид $\mathcal{N}(0, \sigma_\infty^2)$, как в (1), с дисперсией σ_∞^2 , приведенной выше (при выполнении некоторых условий регулярности). Состоятельность тогда следует из того, что

$$\sqrt{n}(\hat{\theta}^* - \theta^*) \Rightarrow \mathcal{N}(0, \sigma_\infty^2) \text{ по вероятности при } n \rightarrow \infty,$$

то есть предельные распределения бутстрапованной и исходной оценок совпадают. Эта сходимость обычно требует среди прочего, чтобы бутстраповская дисперсия была асимптотически верной, то есть

$$n\mathbb{V}^*(\hat{\theta}^*) = \sigma_\infty^2 + o_P(1) \quad (n \rightarrow \infty).$$

Поскольку $\sigma_\infty^2 = \lim_n n\mathbb{V}(\hat{\theta})$, данное требование можно рассматривать как сходимость нормированных дисперсий

$$n\mathbb{V}^*(\hat{\theta}^*) - n\mathbb{V}(\hat{\theta}) = o_P(1) \quad (n \rightarrow \infty).$$

Состоятельность бутстрапа (2) обычно имеет место, если $\hat{\theta}$ асимптотически нормальна. Помимо аппроксимации распределения оценки $\hat{\theta}$, бутстрап позволяет аппроксимировать $\mathbb{V}(\hat{\theta})$ бутстраповской дисперсией $\mathbb{V}^*(\hat{\theta}^*)$. Точность оценивания распределения в (2) зависит от точности бутстраповской дисперсии

$$a_n^2 \mathbb{V}^*(\hat{\theta}^*) - a_n^2 \mathbb{V}(\hat{\theta}), \quad (3)$$

при условии, что существуют разложения Эджворта для $\hat{\theta}$ и $\hat{\theta}^*$. Обычно бесконечномерность предельной дисперсии делает задачу точного оценивания бутстраповской дисперсии гораздо сложнее, чем в случае независимых наблюдений.

Конечно, как и в случае независимых данных, бутстрапирование временных рядов дает преимущество в виде точности приближения более высокого порядка по сравнению с аппроксимацией с помощью нормального распределения как в (1). Для этого приближение оценивается для студентизированных версий $\hat{\theta}$, или же доверительный интервал корректируется с помощью BC_a [коррекция на смещение и увеличение скорости сходимости (Efron, 1987)] или двойного бутстрапа. Но в случае *конечной* выборки схемы первого порядка часто могут быть такими же точными, как их аналоги второго порядка, и (3) желательно надлежащим образом ограничить. Значительная часть настоящей статьи посвящена обсуждению точности первого порядка, но также затрагиваются аспекты приближения второго порядка.

2 Блочный бутстрап

Блочный бутстрап пытается имитировать поведение оценки $\hat{\theta}$ с помощью случайного независимого ресэмплинга $X_{t+1}, \dots, X_{t+\ell}$ последовательных наблюдений: группировка в блоки используется для сохранения исходной структуры временного ряда внутри блока. Подобная идея появилась в статье Hall (1985), но прорывом в блочном бутстрапе стала статья Künsch (1989), где в деталях объясняется, как и почему такая бутстрап-схема работает.

2.1 Процедура блочного бутстрапа

Правильное применение схемы блочного бутстрапа включает, во-первых, адаптацию к задаче. Предположим, что статистика $\hat{\theta}$ оценивает параметр θ , являющийся функционалом m -мерного маргинального распределения временного ряда. Например, автокорреляция первого порядка $\text{Cov}(X_0, X_1)$ для стационарного временного ряда является функционалом распределения (X_0, X_1) , что соответствует $m = 2$. Рассмотрим далее вектора последовательных наблюдений

$$Y_t = (X_{t-m+1}, \dots, X_t), \quad t = m, \dots, n \quad (4)$$

и создадим блочно-бутстраповскую выборку на основе этих векторизованных наблюдений следующим образом. Построим пересекающиеся блоки последовательных векторов $(Y_m, \dots, Y_{m+\ell-1}), (Y_{m+1}, \dots, Y_{m+\ell}), \dots, (Y_{n-\ell+1}, \dots, Y_n)$, где $\ell \in \mathbb{N}$ – длина блока. Для простоты сначала предположим, что число блоков $n - m + 1 = k\ell$, $k \in \mathbb{N}$. Далее осуществляем независимые вытягивания k блоков с возвращением,

$$Y_{S_1+1}, \dots, Y_{S_1+\ell}, Y_{S_2+1}, \dots, Y_{S_2+\ell}, \dots, Y_{S_k+1}, \dots, Y_{S_k+\ell}, \quad (5)$$

где точки начала блоков S_1, \dots, S_k независимо равномерно распределены на множестве $\{m - 1, \dots, n - \ell\}$ всех возможных начальных точек. Если число блоков $n - m + 1$ не кратно ℓ , ресэмплируют $k = \lfloor (n - m + 1) / \ell \rfloor + 1$ блоков, но от k -го блока используется только часть, чтобы в итоге получить $n - m + 1$ m -мерных векторов. Таким образом сформированную последовательность блоков m -мерных векторов в (5) можно назвать блочно-бутстраповской выборкой. Однако, как будет видно далее, блочно-бутстрапованная оценка не определяется принципом подстановки, и понятие бутстраповской выборки тогда не так прозрачно.

Хорошо определить блочно-бутстрапованную оценку не совсем просто. Векторизация в (4) обычно так соотносится с оценкой, что $\hat{\theta}$ симметрична относительно векторизованных наблюдений Y_m, \dots, Y_n . Например, часто оценку можно представить в виде

$$\hat{\theta} = T(F_n^{(m)}), \quad (6)$$

где $F_n^{(m)}(\cdot) = (n - m + 1)^{-1} \sum_{t=m}^n \mathbb{I}_{[Y_t \leq \cdot]}$ — эмпирическая функция распределения m -мерного маржинального распределения $(X_t)_{t \in \mathbb{Z}}$, а T — гладкий функционал.

Пример А. Для автокорреляции первого порядка $\theta = \text{Cov}(X_0, X_1)$ рассмотрим оценку $\hat{\theta} = \hat{R}(1)/\hat{R}(0)$, где $\hat{R}(k) = (n - 1)^{-1} \sum_{t=1}^{n-1} (X_t - \hat{\mu}_X)(X_{t+k} - \hat{\mu}_X)$ ($k \in \{0, 1\}$), $\hat{\mu}_X = (n - 1)^{-1} \sum_{t=1}^{n-1} X_t$. Эта оценка $\hat{\theta}$ симметрична относительно Y_2, \dots, Y_n , где $Y_t = (X_{t-1}, X_t)$, и имеет вид (6), причем $m = 2$. Отметим, что обычная оценка — это $\theta = \tilde{R}(0)/\tilde{R}(1)$, где $\tilde{R}(k) = n^{-1} \sum_{t=1}^n (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)$, что приблизительно равняется $\hat{\theta}$, если забыть об эффектах на краях.

Пример В. GM-оценки (обобщенные M-оценки) в AR(p)-модели можно записать в виде (6) для $m = p + 1$. Они определяются неявно, по аналогии с нормальными уравнениями:

$$\sum_{t=p+1}^n w_t \psi((X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p}) \sigma^{-1})(X_{t-1}, \dots, X_{t-p})^T = 0,$$

где $\psi : \mathbb{R} \rightarrow \mathbb{R}$, σ^2 — дисперсия инноваций, а $(w_t)_{t=p+1}^n$ — последовательность соответствующих весов; см. Martin & Yohai (1986). Помимо гауссовских ММП-оценок с $\psi(x) = x$, они включают оценки, робастные к выбросам в инновациях и лагированных значениях.

Блочно-бутстрапированная оценка, соответствующая (6), определяется как

$$\hat{\theta}^{*B} = T(F_n^{(m)*B}), \quad F_n^{(m)*B}(\cdot) = (n - m + 1)^{-1} \sum_{i=1}^k \sum_{t=S_i+1}^{S_i+\ell} \mathbb{I}_{[Y_t \leq \cdot]}, \quad (7)$$

где k и S_i означают то же, что и в (5). Центральное значение θ^* для блочного бутстрапа в (2) часто равняется $\mathbb{E}^{*B}[\hat{\theta}^{*B}]$, что, вообще говоря, отлично от $\hat{\theta}$. Это определение блочно-бутстрапированной оценки, данное в Künsch (1989), можно интерпретировать следующим образом. Если $\hat{\theta} = g_{n-m+1}(Y_m, \dots, Y_n)$ является симметричной функцией $g_{n-m+1}(\cdot)$ от $n - m + 1$ векторизованных наблюдений, то

$$\hat{\theta}^{*B} = g_{n-m+1}(Y_{S_1+1}, \dots, Y_{S_1+\ell}, Y_{S_2+1}, \dots, Y_{S_2+\ell}, \dots, Y_{S_k+1}, \dots, Y_{S_k+\ell}),$$

где использован принцип подстановки для векторизованных наблюдений. В частности, блочно-бутстрапированная оценка определяется для таких Y , которые попадают в множество исходных векторизованных наблюдений. Это будет *не так* без векторизации (4). Рис. 1 иллюстрирует недостаток наивного блочного бутстрапа, использующего $m = 1$ вместо правильного $m = 2$ в примере А. Наиболее яркий дефект псевдовыборки, полученной с помощью наивного блочного бутстрапа, состоит в появлении новых точек внутри прямоугольников в верхнем левом и нижнем правом углах диаграммы. Наивная блочно-бутстрапированная оценка (например, для автокорреляции в примере А), которая использует принцип подстановки с выборкой, полученной с помощью наивного блочного бутстрапа, может быть сильно подпорчена этими вновь созданными плохими точками. Как уже было сказано, этот недостаток отсутствует в определении блочного бутстрапа на основе векторизованных наблюдений (7).

Касательно процедуры блочного бутстрапа остаются как минимум две трудности, которые необходимо рассматривать в каждом отдельном случае.

- (1) Из-за неприменимости принципа подстановки часто необходимо изменение способа *вычисления* $\hat{\theta}^{*B}$, что может быть очень неудобно.
- (2) Векторизация как в (4) не всегда адекватна. Например, MA-параметр в MA(1)-модели или спектральная плотность стационарного процесса зависят от всего распределения процесса, что соответствует $m = \infty$.

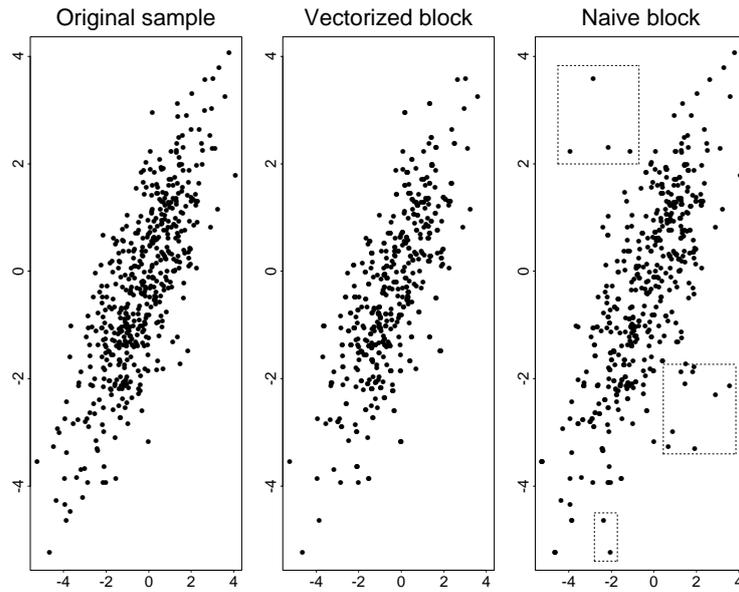


Рис. 1: Диаграммы пар наблюдений, отстоящих на один шаг, в бутстраповских выборках размера $n = 512$. Слева: исходная выборка (X_{t-1}, X_t) , $t = 2, \dots, n$. В центре: псевдовыборка для блочного бутстрапа Y_{S_i+j} , $i = 1, \dots, k = 64$, $j = 1, \dots, \ell = 8$ из (5) для $m = 2$. Справа: псевдовыборка для наивного блочного бутстрапа $(X_{t-1}^{*nB}, X_t^{*nB})$, $t = 2, \dots, n$, где X_t^{*nB} – последовательное t -ое значение из (5) для $m = 1$, $k = 64$, $\ell = 8$; точки внутри прямоугольников (и не только) не встречаются на диаграмме слева.

В тех случаях, когда проблема (1) или (2) становится слишком затруднительной, решение по умолчанию – игнорировать шаг (4), на котором происходит векторизация, и работать с наивным блочным бутстрапом (используя $m = 1$). В результате может произойти значительное снижение эффективности метода. Предложения для решения проблемы (2), главным образом для случая оценивания спектральной плотности, даны в Politis & Romano (1992) и Bühlmann & Künsch (1995).

2.2 Область применимости и точность

Блочный бутстрап предназначен для работы с общими стационарными процессами, порождающими данные, $(X_t)_{t \in \mathbb{Z}}$, где $X_t \in \mathbb{R}^d$ ($d \geq 1$) или принимает значения в пространстве категорий. С точки зрения асимптотики, длина блока ℓ должна расти при $n \rightarrow \infty$, но не слишком быстро. Если ограничиваться только процессами с коротким горизонтом зависимости (например, суммируемые автоковариации или коэффициенты перемешивания), применение блочного бутстрапа теоретически оправдано во многих обстоятельствах, например, для таких оценок, как в (6) с гладким T , см. Künsch (1989), Bühlmann (1994). Другие ссылки даны в разделе 7. Для случаев долгосрочной зависимости разработана некоторая теория и модификации для случая $\hat{\theta} = \bar{X}_n$: Lahiri (1993) показывает, что блочный бутстрап состоятелен, если \bar{X}_n имеет нормальное предельное распределение, но бутстрапированная статистика должна быть скорректирована на коэффициент, зависящий от обычно неизвестной скорости сходимости, например от параметра автомоделирования в автомодельных процессах. Если \bar{X}_n не имеет нормального распределения в качестве предельного из-за долгосрочной зависимости, Hall, Jing & Lahiri (1998) показывают состоятельность модифицированного метода блочных подвыборок. Когда наблюдения имеют маргинальное распределение с тяжелыми хвостами, в Lahiri (1995) показано, что блочный бутстрап с размером псевдовыборки $m \ll n$ состоятелен для случая $\hat{\theta} = \bar{X}_n$.

Что касается точности блочного бутстрапа, рассмотрим вначале оценивание асимптотической дисперсии $\hat{\theta}$. Künsch (1989) доказал, что для среднего квадрата ошибки (MSE)

$$\mathbb{E}[(n\mathbb{V}^{*B}(\hat{\theta}^{*B}) - n\mathbb{V}(\hat{\theta}))^2] \propto n^{-2/3}, \quad (8)$$

что достигается при оптимальной с точки зрения скорости сходимости длине блока $\ell \propto n^{1/3}$. Заметим, что это соответствует (3) при $a_n = \sqrt{n}$. Основные предположения для этого результата требуют, чтобы $n\mathbb{V}(\hat{\theta})$ сходилась к невырожденной предельной дисперсии, T в (6) был достаточно гладким, а также выполнения некоторых условий перемешивания для стационарного процесса $(X_t)_{t \in \mathbb{Z}}$. Несколько удивительно, что скорость $n^{-2/3}$ не зависит от «степени зависимости», то есть того, как быстро автокорреляции, или, более общо, коэффициенты перемешивания, падают по мере увеличения лага между наблюдениями. В частности, даже если автоковариации и коэффициенты перемешивания убывают экспоненциально, скорость для MSE все равно равна $n^{-2/3}$. Таким образом, оценка дисперсии с помощью блочного бутстрапа не адаптирует скорость сходимости к степени зависимости соответствующего процесса. Объяснение отсутствию этой адаптивности было дано уже в Künsch (1989): оценка дисперсии блочным бутстрапом асимптотически эквивалентна оценке спектральной плотности в нуле с треугольным окном:

$$n\mathbb{V}^{*B}(\hat{\theta}^{*B}) \approx \sum_{k=-\ell}^{\ell} \left(1 - \frac{|k|}{\ell}\right) \hat{R}_{\text{IF}}(k), \quad (9)$$

где $\hat{R}_{\text{IF}}(k)$ – эмпирическая ковариация $(\text{IF}(Y_t; F^{(m)}))_{t=m}^n$ порядка k , где $\text{IF}(\cdot; F^{(m)})$ – функция влияния оценки при соответствующем истинном m -мерном маргинальном распределении $F^{(m)}$; функция влияния $\text{IF}(\cdot; F^{(m)})$ – это преобразование, которое асимптотически линейризует подходяще регулярную оценку, см. (10). Но треугольная форма окна $1 - |k|/\ell$ ($k = -\ell, \dots, 0, \dots, \ell$) делает невозможным улучшение по сравнению со скоростью MSE $n^{-2/3}$. Конический блочный бутстрап, предложенный в Künsch (1989, формула (2.12)) и более подробно изученный в Paparoditis & Politis (2001), преодолевает это ограничение.

Что касается построения доверительных областей, Götze & Künsch (1996) доказали, что распределение надлежащим образом определенной студентизированной версии $\hat{\theta}$ можно аппроксимировать блочным бутстрапом с точностью, близкой к $O_P(n^{-2/3})$, используя длину блока $\ell \propto n^{1/3}$.¹ Как и при оценивании дисперсии, степень точности не может быть улучшена для временных рядов с геометрически убывающими свойствами зависимости. Götze & Künsch (1996) также представляют модификацию предложенной Эфроном (Efron, 1987) корректировки BC_a (удаление смещения и увеличение скорости сходимости). В конечных выборках метод, точный во втором приближении, не всегда может быть выгодным. К сожалению, нелегко судить по данным о целесообразности применения техники второго порядка. Двойное блочное бутстрапирование для корректировки доверительной области, полученной с помощью бутстрапа первого порядка, непросто, так как зависимость нарушается в местах соединения блоков, см. Davison & Hall (1993) и Choi & Hall (2000).

2.3 Выбор длины блока ℓ

Оптимальная длина блока, будучи параметром настройки блочного бутстрапа, зависит по крайней мере от трех вещей: процесса, порождающего данные, бутстрапируемой статистики и цели использования бутстрапа, например, оценивания смещения, дисперсии или распределения.

¹Эту скорость можно улучшить до приблизительно $O_P(n^{-3/4})$ при использовании для студентизации оценки дисперсии, которая принимает отрицательные значения с положительной вероятностью.

Рассмотрим сначала оценивание с помощью блочного бутстрапа дисперсии оценки $\hat{\theta}$ вида (6). В этом случае

$$\hat{\theta} \approx (n - m + 1)^{-1} \sum_{t=m}^n \text{IF}(Y_t; F^{(m)}), \quad (10)$$

где $\text{IF}(\cdot; F^{(m)})$ – функция влияния $\hat{\theta}$ в $F^{(m)}$. На основе этой линеаризации формулу (9) можно переписать в виде

$$n \mathbb{V}^{*B}(\hat{\theta}^{*B}) \approx 2\pi \hat{f}_{\text{IF}}(0), \quad (11)$$

где $\hat{f}_{\text{IF}}(\lambda)$ ($0 \leq \lambda \leq \pi$) – оценка спектральной плотности с треугольным окном при частоте λ и ширине окна ℓ^{-1} , основанная на функциях влияния $(\text{IF}(Y_t; F^{(m)}))_{t=m}^n$. Таким образом, длина блока имеет интересную интерпретацию как величина, обратная к ширине окна при оценке спектральной плотности. Отсюда следует, что асимптотически оптимальная в смысле MSE длина блока для оценки дисперсии есть

$$\ell_{opt} \propto n^{1/3}.$$

Bühlmann & Künsch (1999) предлагают оценивать ℓ_{opt} с помощью итеративной схемы подстановки для выбора локально оптимальной ширины окна при оценке спектральной плотности при частоте 0, используя асимптотическую эквивалентность в (11).

Более общий метод, который к тому же применим для выбора оптимальной длины блока ℓ при оценивании распределения, был предложен в Hall, Horowitz & Jing (1995). Авторы рассматривают поведение блочного бутстрапа при разных значениях длины блоков для подвыборок размера $m \ll n$ и получают оптимальную длину блока для размера подвыборки m . Затем оцененная оптимальная длина блока выводится путем экстраполяции Ричардсона до размера исходной выборки n . Этот метод требует спецификации размера подвыборки m , что менее критично, чем выбор длины блока. Подобная техника использования подвыборок является очень общей, но может быть не очень эффективной. В частности, если оценка $\hat{\theta}$ сильно нелинейная, свойства метода в подвыборках могут быть очень плохими; подобная неэффективность в аналогичных обстоятельствах иллюстрируется в подразделе 4.2.

Что касается оценки смещения с помощью блочного бутстрапа, Lahiri (1999) показывает, что асимптотически оптимальные в смысле MSE длины блоков одинаковы для оценивания смещения и дисперсии: оцененные длины блоков для дисперсии могут, таким образом, быть использованы при оценивании смещения.

Автоматический выбор длины блока по крайней мере так же труден в контексте временных рядов, как выбор *локального* параметра настройки наподобие ширины окна. Хуже того, формула (11) указывает на эквивалентность задаче выбора ширины окна только в асимптотике: линеаризация в (10) может иметь значительный эффект в конечных выборках. Более того, длина блока ℓ не имеет практической интерпретации, а диагностические инструменты для ее выбора пока что не разработаны.

3 AR-решетчатый бутстрап для стационарных линейных временных рядов

Решетчатый бутстрап основан на идее решетчатого приближения (Grenander, 1981) процесса $(X_t)_{t \in \mathbb{Z}}$ классом (полу)параметрических моделей. В таком случае бутстрап является не чем иным, как симуляцией из оцененного с помощью решетки процесса.

Временной ряд называется линейным и обратимым, если он представим в виде авторегрессии бесконечного порядка $\text{AR}(\infty)$,

$$X_t - \mu_X = \sum_{j=1}^{\infty} \phi_j (X_{t-j} - \mu_X) + \epsilon_t \quad (t \in \mathbb{Z}), \quad (12)$$

где $\mu_X = \mathbb{E}[X_t]$, $(\epsilon_t)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций, $\mathbb{E}[\epsilon_t] = 0$, и ϵ_t не зависят от $\{X_s; s < t\}$. Это определение корректно, если, например, $\mathbb{E}[\epsilon_t^2] < \infty$ и $\sum_{j=1}^{\infty} \phi_j^2 < \infty$.

3.1 Процедура AR-решетчатого бутстрапа

AR-решетчатое приближение строится с помощью AR(p) моделей

$$X_t - \mu_X = \sum_{j=1}^p \phi_j (X_{t-j} - \mu_X) + \epsilon_t \quad (t \in \mathbb{Z}),$$

где μ_X и ϵ_t определены как в (12). Для исходных данных сначала выбирается порядок авторегрессии \hat{p} , например, с помощью критерия AIC для гауссовских инноваций, см. Shibata (1980). Оставшийся интересующий исследователя объект $\eta_{\hat{p}} = (\mu_X, (\phi_1, \dots, \phi_{\hat{p}}), F_{\epsilon})$ является полупараметрическим. Здесь F_{ϵ} обозначает распределение IID-инноваций ϵ_t . Оценки выбираются следующим образом:

$$\hat{\mu}_X = 1/n \sum_{t=1}^n X_t, \quad (\hat{\phi}_1, \dots, \hat{\phi}_{\hat{p}}) \text{ по методу Юла-Уолкера,}$$

$$\hat{F}_{\epsilon}(x) = \hat{\mathbb{P}}[\epsilon_t \leq x] = (n - \hat{p})^{-1} \sum_{t=\hat{p}+1}^n \mathbb{I}_{[R_t - \bar{R}_{\bullet} \leq x]}, \quad R_t = X_t - \sum_{j=1}^{\hat{p}} \hat{\phi}_j X_{t-j},$$

где \bar{R}_{\bullet} – среднее доступных остатков R_t ($t = \hat{p} + 1, \dots, n$).

Оценки $\hat{p}, \hat{\eta}_{\hat{p}}$ характеризуют распределение $\hat{P}_{n;AR}$ для авторегрессионного процесса. Оно может быть представлено в виде следующего AR(\hat{p})-уравнения:

$$X_t^{*AR-S} - \hat{\mu}_X = \sum_{j=1}^{\hat{p}} \hat{\phi}_j (X_{t-j}^{*AR-S} - \hat{\mu}_X) + \epsilon_t^* \quad (t \in \mathbb{Z}), \quad (13)$$

где $(\epsilon_t^*)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций, с маржинальным распределением $\epsilon_t^* \sim \hat{F}_{\epsilon}$.

Тогда AR-решетчатая бутстраповская выборка – это конечная выборка X_1^*, \dots, X_n^* из процесса (13), имеющего распределение $\hat{P}_{n;AR}$. На практике вычисление происходит так. Начиная с $(X_{-u}^*, \dots, X_{-u+\hat{p}-1}^*) = (\hat{\mu}_X, \dots, \hat{\mu}_X)$ при большом u , например, $u = 1000$. Далее симулируем X_t^* для $t = -u + \hat{p}, \dots, 0, 1, \dots, n$ в соответствии с (13). Поскольку процесс в (13) является (с большой вероятностью) марковским и с геометрическим перемешиванием, значения X_1^*, \dots, X_n^* из симулированной выборки являются очень хорошей аппроксимацией для выборки из стационарного распределения процесса (13). Далее AR-решетчато-бутстрапированная оценка $\hat{\theta}^{*AR-S}$ строится с помощью правила подстановки. Записав $\hat{\theta} = h_n(X_1, \dots, X_n)$ как функцию от исходных данных X_1, \dots, X_n , определим

$$\hat{\theta}^{*AR-S} = h_n(X_1^{*AR-S}, \dots, X_n^{*AR-S}). \quad (14)$$

Подобная схема бутстрапа была предложена в Kreiss (1992) и впоследствии исследована в Bühlmann (1997), Bickel & Bühlmann (1999) и Choi & Hall (2000).

Центральное значение θ^* в (2) для AR-решетчатого бутстрапа находится следующим образом. Параметр θ является функционалом истинного процесса $(X_t)_{t \in \mathbb{Z}} \sim P$, тогда θ^{*AR-S} – тот же самый функционал, вычисленный для оцененного распределения $\hat{P}_{n;AR}$, которое генерирует бутстрапированный процесс в (13).

Пример А (продолжение). Для оценки автокорреляции первого порядка $\theta^{*AR-S} = \text{Corr}^{*AR-S}(X_t^{*AR-S}, X_{t+1}^{*AR-S})$.

Заметим, что, вообще говоря, $\mathbb{E}^{*AR-S}[\hat{\theta}^{*AR-S}] \neq \theta^{*AR-S}$. Вычисление θ^{*AR-S} можно осуществить с помощью быстрой оценки Монте-Карло:

- (1) Сгенерировать одну очень длинную реализацию $X_1^{*AR-S}, \dots, X_v^{*AR-S}$, $v \gg n$.
- (2) Использовать $\hat{\theta}_v^{*AR-S} = h_v(X_1^{*AR-S}, \dots, X_v^{*AR-S})$ в качестве Монте-Карло аппроксимации для θ^{*AR-S} .

Оправдание аппроксимации на шаге (2) дает формула (2), означающая, что $\hat{\theta}_v^{*AR-S}$ сходится к θ^{*AR-S} со скоростью $a_v^{-1} \ll a_n^{-1}$ (в предположении, что $a_n(\hat{\theta}_n - \theta)$ сходится к невырожденному распределению).

3.2 Область применимости и точность

Процедура AR-решетчатого бутстрапа сильно полагается на предположение о том, что данные X_1, \dots, X_n являются конечной реализацией процесса $AR(\infty)$ из (12). В такой постановке состоятельность $\hat{\theta}$ в (2), являющейся гладкой функцией от средних, доказана в Bühlmann (1997); этот результат расширен в Bickel & Bühlmann (1999) для $\hat{\theta}$, имеющей вид (6). Согласно этим результатам, аппроксимирующий порядок авторегрессии должен асимптотически расти, хотя не слишком быстро, при $n \rightarrow \infty$. Представление $AR(\infty)$ включает важный класс ARMA-моделей

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \sum_{k=1}^q \psi_k \epsilon_{t-k} + \epsilon_t \quad (t \in \mathbb{Z})$$

с обратимым генерирующим MA-полиномом, т.е. корни $\Psi(z) = 1 + \sum_{k=1}^q \psi_k z^k$ ($z \in \mathbb{C}$) лежат вне единичного круга $\{z \in \mathbb{C}; |z| \leq 1\}$. Здесь $(\epsilon_t)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций; также должны выполняться некоторые дополнительные условия регулярности, стандартные для теории ARMA-моделей. Но, конечно, существует множество процессов, не представимых в виде $AR(\infty)$, например, приведенные ниже нелинейная авторегрессия второго порядка $AR(2)$, см. (19), или билинейная модель (20). К сожалению, тестирование на линейность или представимость в виде $AR(\infty)$ весьма деликатно: Bickel & Bühlmann (1997) показывают, что замыкание линейных или AR-процессов на удивление велико. Этот факт отражает сложность суждения по конкретному набору данных о том, годится ли для применения AR-решетчатый бутстрап.

В классе линейных обратимых временных рядов, определенных в (12), AR-решетчатый бутстрап имеет высокую точность: теоретические и практические исследования показывают, что он обычно превосходит более общую процедуру блочного бутстрапа из раздела 2. В Bühlmann (1997) показано, что для $\hat{\theta} = \bar{X}_n$ и при аппроксимации порядка авторегрессии \hat{p} с помощью критерия AIC,

$$n\mathbb{V}^{*AR-S}(\bar{X}_n^{*AR-S}) - n\mathbb{V}(\bar{X}_n) = O_P(n^{-(v-2)/(2v)}),$$

если истинные параметры авторегрессии $(\phi_j)_{j \in \mathbb{N}}$ убывают как $\phi_j \leq const \cdot j^{-v}$ ($v > 2$). В частности, если ϕ_j убывают экспоненциально, то

$$n\mathbb{V}^{*AR-S}(\bar{X}_n^{*AR-S}) - n\mathbb{V}(\bar{X}_n) = O_P(n^{-1/2+\kappa}) \text{ для любого } \kappa > 0. \quad (15)$$

Эти два результата показывают, что метод *автоматически адаптируется* к характеру убывания соответствующей структуры зависимости, что является весьма желательным свойством, которым не обладает блочный бутстрап, см. (8). Эти результаты об адаптивности важны не только в асимптотике, но также могут быть использованы при симуляциях в конечных выборках, см. подраздел 4.1.

AR-решетчатый бутстрап очень точен не только при оценивании дисперсии: Choi & Hall (2000) выводят свойства второго порядка для построения доверительных областей. Они предлагают калибровать полученную область первого порядка с помощью двойного бутстрапирования, основываясь на идеях, берущих начало в работах Hall (1986), Veran (1987) и Loh

(1987). Рассмотрим построение двустороннего доверительного интервала, покрывающего θ с вероятностью $1 - \alpha$. Интервал первого порядка имеет вид $[\hat{\theta} - \hat{r}_{1-\alpha/2}, \hat{\theta} - \hat{r}_{\alpha/2}]$, где \hat{r}_{α} — α -квантиль распределения $\hat{\theta}^{*AR-S} - \theta^{*AR-S}$ условно на X_1, \dots, X_n .

Теперь рассмотрим дополнительную корректировку исходного номинального уровня покрытия с использованием двойного бутстрапа. На основе $X_1^{*AR-S}, \dots, X_n^{*AR-S}$ применяем AR-решетчатый бутстрап и получаем $X_1^{**AR-S}, \dots, X_n^{**AR-S}$. Тогда \hat{r}_{α}^{*AR-S} — α -квантиль распределения $\hat{\theta}^{**AR-S} - \theta^{**AR-S}$ условно на $X_1^{*AR-S}, \dots, X_n^{*AR-S}$.

Пусть

$$\hat{a}(1 - q) = \mathbb{P}^{*AR-S}[\hat{\theta}^{*AR-S} - \hat{r}_{1-q/2}^{*AR-S} \leq \theta^{*AR-S} \leq \hat{\theta}^{*AR-S} - \hat{r}_{q/2}^{*AR-S}], \quad (16)$$

что измеряет действительную степень покрытия при номинальном уровне $1 - q$ для бутстрапа второго уровня, основанного на бутстрапированных данных первого уровня $[\theta^{*AR-S} - \text{константа, зависящая только от } X_1, \dots, X_n]$. Далее рассмотрим $\hat{s}_{1-\alpha} = \hat{a}^{-1}(1 - \alpha)$, т.е. $(1 - \alpha)$ -квантиль \hat{a} , рассматриваемой как функция распределения, который корректирует номинальный уровень покрытия $1 - \alpha$ до $\hat{s}_{1-\alpha}$. Теперь берем

$$[\hat{\theta} - \hat{r}_{\{1-(1-\hat{s}_{1-\alpha})/2\}}, \hat{\theta} - \hat{r}_{\{(1-\hat{s}_{1-\alpha})/2\}}] \quad (17)$$

в качестве двустороннего, полученного с помощью двойного бутстрапа доверительного интервала для θ с номинальным уровнем покрытия $1 - \alpha$. Как показано в Choi & Hall (2000), этот интервал второго порядка корректен. Отметим, что явное (сложное) оценивание дисперсии при зависимых данных не является необходимым для студентизации. Choi & Hall (2000) приводят результаты симуляций, которые показывают, что этот интервал второго порядка может дать значительное улучшение и «никогда» не оказывается значительно хуже построения первого порядка.

3.3 Выбор аппроксимирующего порядка авторегрессии

Предлагается использовать AR-решетчатую аппроксимацию вместе с процедурой выбора модели с помощью критерия AIC при гауссовских инновациях. Shibata (1980) доказал оптимальность критерия AIC для прогнозирования в AR(∞)-моделях. Более того, формула (15) и предшествующая ей основаны на AIC и объясняют, почему этот критерий является хорошим выбором при оценивании дисперсии $\hat{\theta} = \bar{X}_n$.

По аналогии с задачей выбора оптимальной длины блока в блочном бутстрапе, оптимальный порядок авторегрессии, вообще говоря, зависит от истинного процесса, порождающего данные, бутстрапируемой статистики и цели применения бутстрапа. Критерий AIC имеет приятное свойство автоматически выбирать более высокий порядок для моделей с большей степенью зависимости. Ничего неизвестно о том, как адаптировать выбор порядка в AR-решетчатой аппроксимации к бутстрапируемой статистике или различным случаям оценивания дисперсии или распределения с помощью бутстрапа.

Параметр настройки процедуры AR-решетчатого бутстрапа, а именно выбор AR-модели, имеет интерпретацию и позволяет проводить диагностическую проверку, включая графические процедуры для AR-остатков. Это отличает его от параметров настройки типа ширины окна, таких как длина блока в подразделе 2.3, которые не имеют интерпретации и нелегко диагностируются по данным. Наш практический опыт говорит о том, что выбор аппроксимирующего порядка авторегрессии достаточно *нечувствителен* в плане свойств AR-решетчатого бутстрапа при условии, что выбранный порядок разумен.

4 Блочный и AR-решетчатый бутстрап в действии

4.1 AR-решетчатый и блочный бутстрап для симулированных временных рядов

Для сравнения двух схем бутстрапа мы рассматриваем симуляционные эксперименты с двумя разными процессами, но для одной и той же статистики – выборочной медианы $\hat{\theta} = \text{med}(X_1, \dots, X_n)$, представителя простых нелинейных оценок. Размер выборки $n = 512$. Кроме того, параметр настройки для AR-решетчатого бутстрапа выбирается путем минимизации критерия AIC; $\ell = 8 = n^{1/3}$ в соответствии с простым правилом выбора оптимальной асимптотической скорости для оценивания дисперсии, а $\hat{\ell}$ – из Bühlmann & Künsch (1999) для блочно-бутстраповской дисперсии, как указано в подразделе 2.3.

В первом эксперименте рассмотрим линейный ARMA(1, 1)-процесс

$$X_t = -0.8X_{t-1} - 0.5\epsilon_{t-1} + \epsilon_t, \quad (18)$$

где $(\epsilon_t)_{t \in \mathbb{Z}}$ – IID-последовательность, независимая от $\{X_s; s < t\}$, $\epsilon_t \sim t_6$. Эта модель представима в виде AR(∞)-процесса (12).

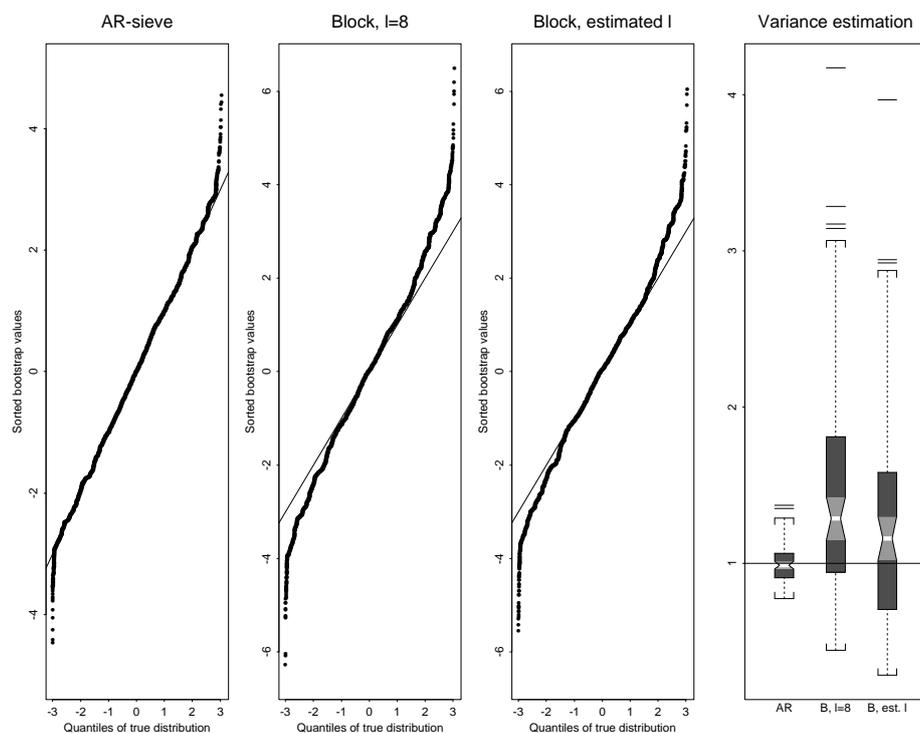


Рис. 2: Линейная модель (18), $n = 512$. Бутстраповское распределение и оценивание дисперсии $(\hat{\theta} - \mathbb{E}[\hat{\theta}])/\sigma_n$ с помощью $(\hat{\theta}^* - \mathbb{E}^*[\hat{\theta}^*])/\sigma_n$ для $\hat{\theta} = \text{med}(X_1, \dots, X_n)$ [$\sigma_n = (\mathbb{V}[\hat{\theta}])^{1/2}$]. Три графика слева: QQ-графики с целевыми значениями, обозначенными прямой. График справа: диаграммы размаха с целевым значением, обозначенными горизонтальной прямой. Всего 100 симуляций, 500 бутстраповских повторов для каждой симуляции.

На Рис. 2 показано качество бутстраповской аппроксимации для выборочной медианы в модели (18). AR-решетчатый бутстрап явно превосходит блочный бутстрап. Оценивание длины блока немного улучшает качество по сравнению с использованием фиксированной длины блока $\ell = 8 = n^{1/3}$: типичные значения $\hat{\ell}$ в 100 симуляциях равны 22, 16 и 30, что соответствует медианному значению и нижнему и верхнему квартилям. Наилучшее поведение AR-решетчатого бутстрапа неудивительно: в данном случае мы используем его преимущества,

обсужденные в подразделе 3.2. Данный результат указывает на количественный выигрыш в случае, когда истинный процесс не является AR-моделью конечного порядка и, следовательно, не является элементом аппроксимирующей решетки (для любого конечного размера выборки), но представим в виде процесса $AR(\infty)$ в (12). Как уже отмечено в Bühlmann (1997), преимущество AR-решетчатого бутстрапа обычно более значительно, если автоковариации процесса характеризуются некоторым затухающим псевдопериодическим убыванием, что верно для модели (18). Это свойство можно диагностировать графически, глядя на оцененные автоковариации.

Второй эксперимент проводился для нелинейного экспоненциального $AR(2)$ -процесса с гетероскедастичными инновациями:

$$\begin{aligned} X_t &= (0.5 + 0.9 \exp(-X_{t-1}^2))X_{t-1} - (0.8 - 1.8 \exp(-X_{t-1}^2))X_{t-2} + \sigma_t \epsilon_t, \\ \sigma_t^2 &= 0.5 + 0.1X_{t-1}^2 + 0.05\sigma_{t-1}^2 1_{[X_{t-1} \leq 0]} + 0.5 \exp(-\sigma_{t-1}^2) 1_{[X_{t-1} > 0]}, \end{aligned} \quad (19)$$

где $(\epsilon_t)_{t \in \mathbb{Z}}$ – IID-последовательность, независимая от $\{X_s; s < t\}$, $\epsilon_t \sim t_6/\sqrt{1.5}$. Этот процесс не представим в виде $AR(\infty)$ из (12).

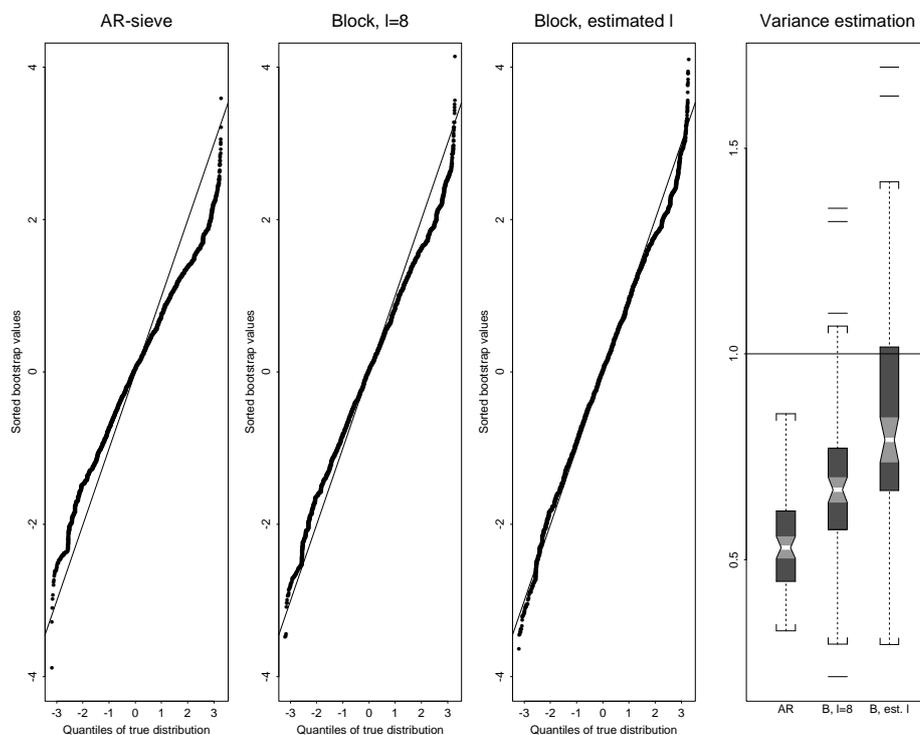


Рис. 3: Нелинейная модель (19), $n = 512$. Бутстраповское распределение и оценивание дисперсии $(\hat{\theta} - \mathbb{E}[\hat{\theta}])/\sigma_n$ с помощью $(\hat{\theta}^* - \mathbb{E}^*[\hat{\theta}^*])/\sigma_n$ для $\hat{\theta} = \text{med}(X_1, \dots, X_n)$ [$\sigma_n = (\mathbb{V}[\hat{\theta}])^{1/2}$]. Три графика слева: QQ-графики с целевыми значениями, обозначенными прямой. График справа: диаграммы размаха с целевым значением, обозначенными горизонтальной прямой. Всего 100 симуляций, 500 бутстраповских повторов для каждой симуляции.

На Рис. 3 показано качество бутстраповской аппроксимации для выборочной медианы в модели (19). AR-решетчатый бутстрап, являющийся асимптотически несостоятельным из-за нелинейности модели (19), дает явное смещение, и блочный бутстрап имеет превосходство. Как и в линейном случае (18) выше, использование оцененной длины блока $\hat{\ell}$ дает улучшение по сравнению с фиксированной длиной блока $\ell = 8 = n^{1/3}$: типичные значения $\hat{\ell}$ в 100 симуляциях равны 11, 7 и 14, что соответствует медианному значению и нижнему и верхнему квартилям. Результаты на Рис. 3 вновь дают количественное представление о выигрыше при использовании блочного бутстрапа в нелинейной модели.

4.2 Сравнение с методом подвыборок

Использование блочных подвыборок – это весьма общая техника оценивания моментов распределения оценки $\hat{\theta}$. Основная идея состоит в расчете оценки $\hat{\theta} = h_n(X_1, \dots, X_n)$ для большого числа подвыборок из ℓ последовательных наблюдений (блоков)

$$\hat{\theta}_{\ell,t} = h_\ell(X_{t-\ell+1}, \dots, X_t), \quad t = \ell, \dots, n.$$

Аппроксимации распределения и дисперсии с помощью подвыборок строятся следующим образом:

$$(n - \ell + 1)^{-1} \sum_{t=\ell}^n 1_{[a_\ell(\hat{\theta}_{\ell,t} - \hat{\theta}) \leq x]} \approx \mathbb{P}[a_n(\hat{\theta} - \theta) \leq x],$$

$$(n - \ell + 1)^{-1} \sum_{t=\ell}^n (a_\ell(\hat{\theta}_{\ell,t} - \hat{\theta}))^2 \approx a_n^2 \mathbb{V}(\hat{\theta}),$$

где $(a_n)_{n \in \mathbb{N}}$ – как в (2). Главным преимуществом данного метода по сравнению с бутстрапом

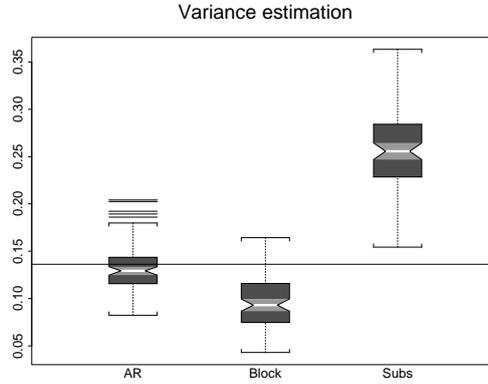


Рис. 4: Линейная модель (18), $n = 512$. Оценивание дисперсии $n\mathbb{V}[\hat{\theta}]$ для оцененной автокорреляции первого порядка $\hat{\theta} = \hat{\rho}(1)$ (горизонтальная прямая). AR, Block и Subs обозначают AR-решетчатый бутстрап, блочный бутстрап и метод подвыборок, соответственно; длина блоков для последних двух методов $\ell = 8$. Всего 100 симуляций, 500 бутстраповских повторов для каждой симуляции.

является весьма общие обстоятельства, в которых аппроксимация с помощью метода подвыборок является состоятельной. Детали описаны в Politis, Romano & Wolf (1999, раздел 3). Однако расчет оценки $\hat{\theta}$ в подвыборках гораздо меньшего размера $\ell \ll n$ и масштабирование ее поведения до размера исходной выборки n может быть проблематичным, если $\hat{\theta}$ сильно нелинейна, а размер выборки n не слишком велик. Простой, но впечатляющий пример – эмпирический аналог автокорреляции первого порядка

$$\hat{\theta} = \hat{\rho}(1) = \hat{R}(1)/\hat{R}(0),$$

как в примере А.

На Рис. 4 показаны результаты оценивания дисперсии $\hat{\rho}(1)$ в модели (18) при размере выборки $n = 512$. AR-решетчатый бутстрап дает наилучший результат, так как модель линейная; блочный бутстрап явно превосходит технику подвыборок при длине блоков $\ell = 8 = n^{1/3}$ для обоих методов (выбранной в соответствии с простым правилом выбора верной асимптотической скорости для обоих методов). В целом, не рекомендуется использовать технику подвыборок, когда известно, что бутстрап состоятелен. Техника подвыборок является интересным инструментом для сложных процедур с $\hat{\theta}$ (при большом объеме данных), когда бутстраповские схемы могут дать сбой.

5 Локальный бутстрап для оценок условного среднего

Решетчатый и блочный бутстрап дают разумные результаты для разнообразных оценок $\hat{\theta}$, если процесс, порождающий данные, принадлежит к подходящему для соответствующей бутстрап-схемы классу, как обсуждалось выше. Несколько удивительно, что некоторые схемы бутстрапа, основанные на независимом ресэмплинге, можно использовать для класса непараметрических оценок $\hat{\theta}$, имеющих более медленную скорость сходимости, чем $1/\sqrt{n}$, например, для $\hat{\theta}(\cdot)$ – (ядерной) сглаживающей оценки условного матожидания $\theta(\cdot) = \mathbb{E}[X_t | X_{t-1} = \cdot]$ стационарного процесса.

Сконцентрируемся на интервальном оценивании с помощью так называемого локального бутстрапа для условного матожидания $\theta(x) = \mathbb{E}[X_t | X_{t-1} = x]$ ($x \in \mathbb{R}$) стационарного процесса с областью действительных значений $(X_t)_{t \in \mathbb{Z}}$. Этот случай, выбранный ради простоты изложения, легко можно расширить на общий – $\mathbb{E}[f(X_t) | X_{t-i_1} = x_1, \dots, X_{t-i_p} = x_p]$ при заданном множестве p лагированных индексов $t - i_1, \dots, t - i_p$ и $f : \mathbb{R} \rightarrow \mathbb{R}$. Имея данные X_1, \dots, X_n , рассмотрим ядерную оценку

$$\hat{\theta}_h(x) = \frac{\sum_{t=2}^n W_{t,h}(x) X_t}{\sum_{t=2}^n W_{t,h}(x)}, \quad W_{t,h}(x) = K\left(\frac{x - X_{t-1}}{h}\right)$$

с шириной окна h .

Для бутстрапирования $\hat{\theta}_h(x)$ можно формировать псевдовыборки в рамках модели локальной регрессии

$$X_t^{*L} \sim \hat{F}_{X_{t-1},b} (t = 2, \dots, n) \text{ независимо от } X_s^{*L} (s \neq t),$$

где $\hat{F}_{x,b}(\cdot) = \sum_{t=2}^n W_{t,b}(x) \mathbb{I}_{[X_t \leq \cdot]} / \sum_{t=2}^n W_{t,b}(x)$ – оценка условной функции распределения X_t при $X_{t-1} = x$; b – это (контрольная) ширина окна, а веса $W_{t,b}(\cdot)$ определены выше. Тогда ресэмплинг осуществляется независимо от оцененной $\{\hat{F}_{x,b}(\cdot); x \in \mathbb{R}\}$, которой позволено меняться *локально*. Бутстрапированная ядерная оценка $\hat{\theta}_h(x)$ тогда получается из данных регрессионного типа $(X_1, X_2^{*L}), (X_2, X_3^{*L}), \dots, (X_{n-1}, X_n^{*L})$:

$$\hat{\theta}_h^{*L}(x) = \frac{\sum_{t=2}^n W_{t,h}(x) X_t^{*L}}{\sum_{t=2}^n W_{t,h}(x)}.$$

Такой подход был рассмотрен в Neumann & Kreiss (1998) и Paparoditis & Politis (2000).

Локальный бутстрап работает, поскольку асимптотическое распределение ядерной оценки $\hat{\theta}_h(x)$ является гауссовским, зависящим только от маргинального распределения X_t , условного распределения X_t на X_{t-1} и известной формы ядра; см. Robinson (1983). Локальный бутстрап способен состоятельно оценить все эти неизвестные. Верно это только в асимптотике, и для любого конечного размера выборки n уже дисперсия $\hat{\theta}_h(x)$ определенным образом зависит от n -мерного распределения $(X_t)_{t \in \mathbb{Z}}$. По построению, локальный бутстрап не может ухватывать зависимости за пределами условного распределения X_t на X_{t-1} . Этот недостаток отсутствует при использовании бутстрап-схем для зависимых наблюдений, например, блочного бутстрапа, точность оценивания $\mathbb{V}[\hat{\theta}_h(x)]$ выше при блочном бутстрапе, чем при локальном.

Neumann & Kreiss (1998) и Neumann (1998) строят (используя соответствующий локальный бутстрап) состоятельные доверительные области для $\theta(x)$, являющиеся *совместными* для x . Их скорости сходимости равны $1/\sqrt{n\bar{h}}$, как для поточечного случая. Это очень важный результат, так как аналитические совместные аппроксимации приближаются к предельному распределению экстремальных значений с очень медленной скоростью $1/\log(n)$: аналитический подход с предельным распределением гораздо хуже, чем схема локального бутстрапа.

5.1 Область применимости и выбор параметра настройки

Доказано, что локальный бутстрап состоятелен, если $(X_t)_{t \in \mathbb{Z}}$ – процесс с малым горизонтом зависимости, см. Paparoditis & Politis (2000) и Ango Nze, Bühlmann & Doukhan (2002).

Параметром настройки локального бутстрапа является контрольная ширина окна b . Простое решение – выбрать $b = h$, где h – предварительно выбранная ширина окна для оценивания $\hat{\theta}_h(x)$. Если b более высокого порядка, чем h , асимптотически нетривиальное смещение $\mathbb{E}[\hat{\theta}_h(x)] - \theta(x)$ можно оценить с помощью локального бутстрапа, см. Paparoditis & Politis (2000). Контрольная ширина окна играет роль при оценивании условного распределения X_t на X_{t-1} : эта задача довольно проста для двумерных распределений. Процедура не является чувствительной к спецификации этой контрольной ширины окна.

5.2 Локальный и блочный бутстрап для симулированных временных рядов

Интересно узнать, имеет ли преимущество, с точки зрения конечных выборок, бутстрап-схема, принимающая во внимание временную структуру, например блочный бутстрап. Рассмотрим

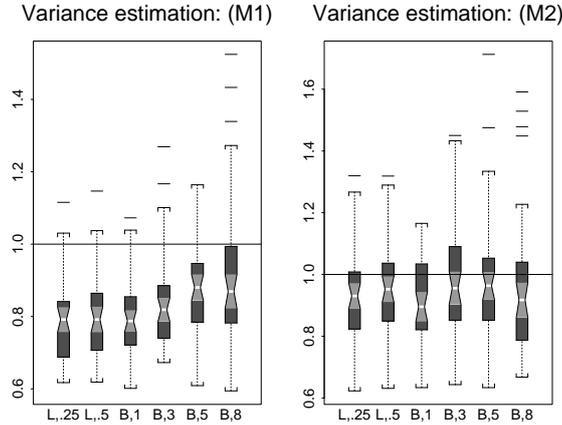


Рис. 5: Бутстраповские оценки дисперсии $\mathbb{V}^*[\hat{\theta}_h^*(x)]/\mathbb{V}[\hat{\theta}_h(x)]$ при $x = 1, 60$ и $x = 0, 76$ для (M1) и (M2) соответственно (целевой уровень указан горизонтальной прямой); L соответствует локальному бутстрапу с контрольной шириной окна 0,25 и 0,5; B соответствует блочному бутстрапу с длиной блока 1, 3, 5 и 8. Размер выборки $n = 512$.

симуляционный эксперимент на основе билинейной модели:

$$X_t = 0.5\epsilon_{t-1}X_{t-1} + \epsilon_t, \quad (20)$$

где $(\epsilon_t)_{t \in \mathbb{Z}}$ – последовательность IID-инноваций, где ϵ_t независимы от $\{X_s; s < t\}$. Рассмотрим случаи, когда

$$(M1) \quad \epsilon_t \text{ IID } \sim U(\{-1, 1\}), \text{ то есть } \mathbb{P}[\epsilon_t = 1] = \mathbb{P}[\epsilon_t = -1] = 1/2,$$

$$(M2) \quad \epsilon_t \text{ IID } \sim U([-1, 1]).$$

Обе модели (M1) и (M2) характеризуются слабой формой зависимости. Эмпирически мы обнаружили, что оценивание $\mathbb{V}[\hat{\theta}_h(x)]$ сложнее для модели с дискретными инновациями (M1), чем для (M2).

Рис. 5 отражает результаты оценивания $\mathbb{V}[\hat{\theta}_h(x)]$ со стандартным гауссовским ядром K и разумной шириной окна $h = 0, 25$. Размер выборки $n = 512$. Графически определить существенные различия затруднительно. Количественное описание имеет следующий вид. Для модели (M1) блочный бутстрап с $\ell = 5$ ('B,5') дает наилучшие результаты в смысле MSE:

он дает приблизительно на 40% меньшую MSE, чем наилучший вариант локального бутстрапа с $b = 0,5$ ('L,.5'): двусторонний парный тест Уилкоксона отдает предпочтение схеме 'B,5' с P -значением 0,002 при нулевой гипотезе о равных MSE. Сравнение любого варианта локального бутстрапа с любым другим вариантом блочного бутстрапа с длинами блоков $\ell = 1, 3, 8$ не дает значительных отличий. Для (M2) блочный бутстрап с $\ell = 1$ (что является регрессионным бутстрапом при независимости) дает наилучшие результаты с MSE, на приблизительно 9% меньшей, чем для наилучшего локального бутстрапа с $b = 0,5$, но различие незначимо. За исключением варианта 'B,8' с неразумно большой длиной блоков (значительно уступающего локальному бутстрапу), сравнение любого варианта локального бутстрапа с любыми вариантами блочного бутстрапа при $\ell = 1, 3, 5$ не выявляет значимого различия. Для конкретной билинейной модели со слабой степенью зависимости можно сделать следующий вывод. В более простом случае (M2) локальный и блочный бутстрап одинаково хороши (исключая случай неразумной длины блоков $\ell = 8$). Это не так для немного более сложного случая (M1), когда блочный бутстрап всегда так же хорош, как локальный, или даже лучше него, при условии, что имеется удачное правило выбора длины блока вблизи $\ell = 3$. Ожидается более явное преимущество блочного бутстрапа по сравнению с локальным, если данные характеризуются большей степенью зависимости.

6 Заключение

Среди схем блочного, решетчатого и локального бутстрапа блочный бутстрап является наиболее общим методом. Другое его преимущество – простая реализация процедуры ресэмплинга, не сложнее, чем в эфроновском бутстрапе для независимых наблюдений. Недостатки же метода заключаются в следующем. Блочно-бутстраповскую выборку не следует рассматривать как разумную выборку, имитирующую порождающий данные процесс: она нестационарна и искажена в местах соединения взятых из выборки блоков. Это означает, что правило подстановки для бутстрапирования оценки $\hat{\theta}$ не годится. Предварительная векторизация данных очень рекомендуется, но это может потребовать изменения бутстрапированной оценки и процедуры ее подсчета. Как общая непараметрическая схема, блочный бутстрап может уступать другим методам для различных классов стационарных временных рядов, например, для линейных. Точность второго порядка для доверительного интервала достигается путем стьюдентизации и BC_a -коррекции; последний метод дает маргинальное улучшение в симулированном примере. Двойное бутстрапирование не дает больших надежд, так как блочный бутстрап на первой итерации нарушает зависимость в местах соединения блоков.

Решетчатый бутстрап в целом отталкивается от разумной модели временного ряда. Это приводит к двум преимуществам: правило подстановки используется для определения и расчета бутстраповской оценки, а двойной бутстрап потенциально приводит к точности более высокого порядка. Хорошие схемы решетчатого бутстрапа, такие как AR-решетчатый бутстрап, адаптируются к степени зависимости временного ряда: их точность растет по мере снижения степени зависимости, см. формулу (15). Это не так для блочного бутстрапа, как видно из формулы (8). Кроме того, решетчатый бутстрап в целом менее чувствителен к выбору модели для решетки, чем блочный бутстрап к длине блоков.

AR-решетчатый бутстрап, естественно, является наилучшим, если процесс, порождающий данные, является линейным временным рядом, представимым в виде $AR(\infty)$, см. (12). Метод легок в реализации благодаря простоте подгонки AR-модели.

Локальный бутстрап из раздела 5 ограничивается процедурами непараметрического оценивания, имеющими более низкую скорость сходимости, чем $1/\sqrt{n}$. Хотя он реализуется аналогично регрессионному бутстрапу в случае независимости, он состоятелен и, следовательно, устойчив к некоторым формам зависимости. Его преимуществом является простота, поскольку не требуется специфицировать параметр настройки, управляющий степенью за-

висимости порождающего данные процесса. С другой стороны, это также свидетельствует о его слабости и неспособности имитировать зависимость подобающим образом: этот метод может давать результаты похуже, чем блочный бутстрап.

7 Другие результаты и примечания к списку литературы

В дополнение к приведенному избирательному изложению дадим ссылки на дополнительную литературу. Efron & Tibshirani (1993, гл. 8.5–8.6), Shao & Tu (1995, гл. 9), Li & Maddala (1996), Davison & Hinkley (1997, гл. 8) обсуждают методы бутстрапирования для зависимых наблюдений с другой точки зрения, чем наш сравнительный обзор.

Литература по блочному бутстрапу на настоящий момент обширна. Обзор более ранних исследований в этой области содержится в Léger, Politis & Romano (1992). Переработка результатов Künsch (1989) при минимальных предположениях содержится в Radulović (1996). Различные результаты для эмпирических процессов приводятся в Bühlmann (1994, 1995) и Peligrad (1998). Lahiri (1996) доказывает корректность второго порядка для блочного бутстрапа в случае, когда $\hat{\theta}$ является М-оценкой в линейной регрессионной модели с зависимым шумом. Техника блочного бутстрапа также применима к пространственным процессам, см. Politis & Romano (1993). Версия блочного бутстрапа, достигающая стационарности бутстраповской выборки (так называемый стационарный бутстрап), была предложена в Politis & Romano (1994a). Lahiri (1999) строго показывает, что блочный бутстрап лучше стационарного. Carlstein, Do, Hall, Hesterberg & Künsch (1998) предлагают схему связывания блоков при формировании псевдovyборки: они утверждают, что для случая оценивания дисперсии $\hat{\theta} = \overline{X}_n$ такая процедура имеет меньшую среднеквадратическую ошибку.

Связанными с блочным бутстрапом являются методы, использующие подвыборки. Работу Carlstein (1986) можно рассматривать как предшественницу блочного бутстрапа для оценивания дисперсии. В своей выдающейся статье Politis & Romano (1994b) показали, что метод подвыборок применим гораздо в более общем случае, чем методы блочного бутстрапа, а именно, практически во всех случаях, когда $\hat{\theta}$ имеет некоторое невырожденное предельное распределение. Künsch (1989) утверждает, что для случая, когда статистика $\hat{\theta}$ асимптотически нормальна, блочный бутстрап превосходит метод подвыборок. Другие результаты в методе подвыборок можно найти в книге Politis, Romano & Wolf (1999).

Основанное на моделях бутстрапирование изучалось для множества случаев: Freedman (1984) – для AR-модели; Kreiss & Franke (1992) – для ARMA-моделей; Paparoditis & Politis (2002) – для марковских моделей. Непараметрическая AR(1)-модель с гетероскедастичными инновациями обсуждается в Franke, Kreiss, Mammen & Neumann (2002): бутстрап, основанный на моделях, можно использовать для точного построения совместных доверительных интервалов для авторегрессионной функции $m(x) = \mathbb{E}[X_t | X_{t-1} = x]$. Отметим, что того же можно достигнуть (в первом порядке) с помощью локального бутстрапа из раздела 5.

Для AR-решетчатого бутстрапа результаты для эмпирических процессов даны в Bickel & Bühlmann (1999) путем введения слабого понятия перемешивания для бутстрапируемых процессов. Нестационарный случай $X_t = m_t + Z_t$ ($t \in \mathbb{Z}$), где $(m_t)_{t \in \mathbb{Z}}$ – медленно меняющийся детерминистский тренд, а $(Z_t)_{t \in \mathbb{Z}}$ – AR(∞)-шум, изучается в Bühlmann (1998), где обсуждаются AR-решетчато-бутстраповские доверительные интервалы для тренда.

Объединение решетчатых или модельных методов с блочным бутстрапом было предложено в Davison & Hinkley (1997, гл. 8.2); авторы называют такую процедуру ‘последующее зачернение’. Идея состоит в том, чтобы предварительно отбелить временной ряд с помощью модельного или решетчатого подхода, а затем применить блочный бутстрап к, возможно, менее зависимым, отбеленным остаткам: блочная формировка подвыборок из этих остатков и обращение операции отбеливания тогда дают зачерненную бутстраповскую выборку.

Другой метод бутстрапирования стационарных линейных временных рядов был предложен

в Dahlhaus & Janas (1996); авторы независимым образом ресэмплируют значения периодограммы в частотной области соответственно оценке спектральной плотности. По построению, такой ресэмплинг затрагивает только структуру автоковариаций, и состоятельность таким образом ограничивается только линейными временными рядами. Идея независимого ресэмплинга в частотной области прежде появлялась в Franke & Härdle (1992) при бутстрапировании оценки спектральной плотности; ее модификация для бутстрап-схемы локального типа рассматривается в Paparoditis & Politis (1999).

Список литературы

- Ango Nze, P., P. Bühlmann & P. Doukhan (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics* 30, 397–430.
- Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* 74, 457–468.
- Bickel, P.J. & P. Bühlmann (1997). Closure of linear processes. *Journal of Theoretical Probability* 10, 445–479.
- Bickel, P.J. & P. Bühlmann (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* 5, 413–446.
- Bühlmann, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics* 22, 995–1012.
- Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and their Applications* 58, 247–265.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3, 123–148.
- Bühlmann, P. (1998). Sieve bootstrap for smoothing in non-stationary time series. *Annals of Statistics* 26, 48–83.
- Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science* 17, 52–72.
- Bühlmann, P. & H.R. Künsch (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics* 22, 35–54.
- Bühlmann, P. & H.R. Künsch (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis* 31, 295–310.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* 14, 1171–1179.
- Carlstein, E., K.-A. Do, P. Hall, T. Hesterberg & H.R. Künsch (1998). Matched-block bootstrap for dependent data. *Bernoulli* 4, 305–328.
- Choi, E. & P. Hall (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *Journal of Royal Statistical Society (Series B)* 62, 461–477.
- Dahlhaus, R. & D. Janas (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Annals of Statistics* 24, 1934–1963.
- Davison, A.C. & P. Hall (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* 35, 215–224.
- Davison, A.C. & D.V. Hinkley (1997). *Bootstrap Methods and their Applications*. Cambridge University Press.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1–26.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of American Statistical Association* 82, 171–185.
- Efron, B. & R.J. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall.
- Franke, J. & W. Härdle (1992). On bootstrapping kernel spectral estimates. *Annals of Statistics* 20, 121–145.
- Franke, J., J.-P. Kreiss, E. Mammen & M.H. Neumann (2002). Properties of the nonparametric autoregressive bootstrap. *Journal of Time Series Analysis* 23, 555–585.
- Freedman, D.A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals of Statistics* 12, 827–842.
- Götze, F. & H.R. Künsch, (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914–1933.

- Grenander, U. (1981). *Abstract Inference*. Wiley, New York.
- Hall, P. (1985). Resampling a coverage pattern. *Stochastic Processes and their Applications* 20, 231–246.
- Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics* 14, 1431–1452.
- Hall, P., J.L. Horowitz & B.-Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* 82, 561–574.
- Hall, P., B.-Y. Jing & S.N. Lahiri (1998). On the sampling window method under long range dependence. *Statistica Sinica* 8, 1189–1204.
- Kreiss, J.-P. (1992). Bootstrap procedures for $AR(\infty)$ -processes. В *Bootstrapping and Related Techniques*, Eds. Jöckel, K.-H., G. Rothe & W. Sendler, стр. 107–113. Springer, Heidelberg.
- Kreiss, J.-P. & J. Franke (1992). Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis* 13, 297–317.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Lahiri, S.N. (1993). On the moving block bootstrap under long range dependence. *Statistics and Probability Letters* 18, 405–413.
- Lahiri, S.N. (1995). On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tail random variables. *Annals of Statistics* 23, 1331–1349.
- Lahiri, S.N. (1996). On Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models. *Journal of Multivariate Analysis* 56, 42–59.
- Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics* 27, 386–404.
- Léger, C., D.N. Politis & J.P. Romano (1992). Bootstrap technology and applications. *Technometrics* 34, 378–398.
- Li, H. & G.S. Maddala (1996). Bootstrapping time series models. *Econometric Reviews* 15, 115–158.
- Loh, W. (1987). Calibrating confidence coefficients. *Journal of American Statistical Association* 82, 155–162.
- Martin, R.D. & V.J. Yohai (1986). Influence functionals for time series. *Annals of Statistics* 14, 781–818 (обсуждение на стр. 819–855).
- Neumann, M.H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Annals of Statistics* 26, 2014–2048.
- Neumann, M.H., J.-P. Kreiss (1998). Regression-type inference in nonparametric autoregression. *Annals of Statistics* 26, 1570–1613.
- Paparoditis, E. & D.N. Politis (1999). The local bootstrap for periodogram statistics. *Journal of Time Series Analysis* 20, 193–222.
- Paparoditis, E. & D.N. Politis (2000). The local bootstrap for kernel estimators under general dependence conditions. *Annals of the Institute of Statistical Mathematics* 52, 139–159.
- Paparoditis, E. & D.N. Politis (2001). Tapered block bootstrap. *Biometrika* 88, 1105–1119.
- Paparoditis, E. & D.N. Politis (2002). The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference* 108, 301–328.
- Peligrad, M. (1998). On the blockwise bootstrap for empirical processes for stationary sequences. *Annals of Probability* 26, 877–901.
- Politis, D.N. & J.P. Romano (1992). A general resampling scheme for triangular arrays of α -mixing random variables. *Annals of Statistics* 20, 1985–2007.
- Politis, D.N. & J.P. Romano (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis* 47, 301–328.
- Politis, D.N. & J.P. Romano (1994a). The stationary bootstrap. *Journal of American Statistical Association* 89, 1303–1313.
- Politis, D.N. & J.P. Romano (1994b). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22, 2031–2050.
- Politis, D.N., J.P. Romano & M. Wolf (1999). *Subsampling*. Springer.

- Radulović, D. (1996). The bootstrap of the mean for strong mixing sequences under minimal conditions. *Statistics and Probability Letters* 28, 65–72.
- Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
- Shao, J. & D. Tu (1995). *The Jackknife and Bootstrap*. Springer.
- Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.

Bootstrap schemes for time series

Peter Bühlmann

ETH Zürich, Zürich, Switzerland

We review and compare block, sieve and local bootstraps for time series and thereby illuminate theoretical aspects of the procedures as well as their performance on finite-sample data. Our view is selective with the intention of providing a new and fair picture of some particular aspects of bootstrapping time series. The generality of the block bootstrap is contrasted with the sieve bootstrap. We discuss implementational advantages and disadvantages, and argue that the sieve often outperforms the block method. Local bootstraps, designed for nonparametric smoothing problems, are easy to use and implement but exhibit in some cases low performance.

Бутстраповское рафинирование тестов, основанных на ОММ*

Валентина Корради[†]

Университет Уорвика, Ковентри, Великобритания

Настоящее эссе содержит краткое обозрение рафинирований высокого порядка, свойственных бутстрапу тестов, основанных на обобщенном методе моментов. Во-первых, мы вкратце описываем асимптотическое поведение двухшаговых ОММ-оценок. Во-вторых, мы эвристически поясняем, почему инференция, основанная на бутстраповских критических значениях, более точна, чем инференция, полагающаяся на асимптотическую нормальность. В-третьих, мы сжато описываем непараметрические методы ресэмплинга. В-четвертых, мы обрисовываем, как использование критических значений, основанных на блочном бутстрапе, приводит к уменьшению ошибки вероятности отклонения гипотез для основанных на ОММ-оценках t -тестов. Наконец, мы даем обзор некоторых альтернативных бутстраповских процедур, которые приводят к улучшению блочно-бутстраповского рафинирования.

1 Введение

Данное эссе содержит краткое обозрение рафинирований высокого порядка, свойственных бутстрапу тестов, основанных на обобщенном методе моментов (ОММ). Из экспериментов Монте-Карло видно, что в умеренного размера выборках тесты, основанные на ОММ, такие как t -тесты и тесты на сверхидентифицирующие ограничения, часто страдают резкими искажениями уровня. Хорошо известно, что при определенных условиях регулярности инференция, основанная на бутстраповских критических значениях, приводит к уменьшению ошибки вероятности отклонения (ОВО) нулевой гипотезы, то есть к уменьшению разницы между номинальным и фактическим уровнями теста. Тем не менее, изучение свойств высокого порядка у бутстрапа ОММ-оценок для зависимых данных началось сравнительно недавно. Hall & Horowitz (1996) показали, что использование бутстраповских критических значений действительно дает уменьшения ОВО, а Andrews (2002) определил порядок величины уменьшения ОВО бутстрапом.

Эссе организовано следующим образом. Во-первых, мы вкратце опишем асимптотическое поведение двухшаговых ОММ-оценок. Во-вторых, мы эвристически поясним, почему инференция, основанная на бутстраповских критических значениях, более точна, чем инференция, полагающаяся на асимптотическую нормальность. В-третьих, мы сжато опишем непараметрические методы ресэмплинга, подходящие для случая, когда распределение данных неизвестно. В-четвертых, следуя Andrews (2002), мы обрисуем бутстраповское рафинирование для t -тестов, основанных на ОММ, для случая зависимых наблюдений. В-пятых, мы обратимся к некоторым вопросам практической реализации бутстраповских оценок. Наконец, мы упомянем, каким образом блочно-блочный бутстрап (Andrews, 2004) и марковский бутстрап (Horowitz, 2003) приводят к улучшенному рафинированию в случае марковских или приблизительно марковских процессов.

*Перевод С. Анатольева. Данное эссе основано на лекциях по курсу «Продвинутая эконометрика», читаемых автором докторантам университета Уорвика. Цитировать как: Корради, Валентина (2007) «Бутстраповское рафинирование тестов, основанных на ОММ», Квантиль, №3, стр. 57–66. Citation: Corradi, Valentina (2007) “Bootstrap refinements for GMM based tests,” *Quantile*, No.3, pp. 57–66.

[†]Адрес: Department of Economics, University of Warwick, Coventry CV4 7AL, UK. Электронная почта: V.Corradi@warwick.ac.uk

2 Асимптотическая нормальность ОММ

Сначала мы обрисовем достаточные условия асимптотической нормальности двухшаговых ОММ-оценок.

Пусть $g_t(\beta) = g(y_t, X_t, \beta)$, где $\beta \in \mathbf{B} \subseteq \mathbb{R}^k$, а g принимает значения в \mathbb{R}^p , причем $p \geq k$. Определим двухшаговую ОММ-оценку как

$$\begin{aligned} \hat{\beta}_T &= \arg \min_{\beta \in \mathbf{B}} \left(\frac{1}{T} \sum_{t=1}^T g_t(\beta) \right)' \hat{\Omega}_T(\bar{\beta}_T) \left(\frac{1}{T} \sum_{t=1}^T g_t(\beta) \right) \\ &= \arg \min_{\beta \in \mathbf{B}} G_T(\beta)' \hat{\Omega}_T(\bar{\beta}_T) G_T(\beta), \end{aligned} \quad (1)$$

где $\bar{\beta}_T$ – оценка первого шага, определяемая как

$$\bar{\beta}_T = \arg \min_{\beta \in B} G_T(\beta)' \Omega_{G_T}(\beta), \quad (2)$$

где обычно $\Omega = \mathbf{I}_p$, единичная матрица $p \times p$. Когда $p = k$, у нас *точно идентифицированный* ОММ, а когда $p > k$, – *сверхидентифицированный* ОММ. Матрица $\hat{\Omega}_T(\bar{\beta}_T)$ называется *оцененной взвешивающей*.

Определим

$$\beta^\dagger = \arg \min_{\beta \in B} G_\infty(\beta)' \Omega_\infty(\beta^\dagger) G_\infty(\beta),$$

где $G_\infty(\beta)$ и Ω_∞ – пределы по вероятности (или почти наверное) $G_T(\beta)$ и $\hat{\Omega}_T(\bar{\beta}_T)$, соответственно.

Предположение А1

A1(i): $\sup_{\beta \in B} |G_T(\beta) - G_\infty(\beta)| \xrightarrow{p} 0$ и $\hat{\Omega}_T(\bar{\beta}_T) \xrightarrow{p} \Omega_\infty(\beta^\dagger)$, где матрица $\Omega_\infty(\beta^\dagger)$ строго положительно определена, а \mathbf{B} есть компактное множество в \mathbb{R}^k , то есть вектор функций моментов и взвешивающая матрица удовлетворяют равномерному закону больших чисел.

A1(ii): $G_\infty(\beta^\dagger)' \Omega_\infty(\beta^\dagger) G_\infty(\beta^\dagger) < G_\infty(\beta)' \Omega_\infty(\beta^\dagger) G_\infty(\beta)$, т.е. имеется идентификация с единственностью.

A1(iii): $G_\infty(\beta^\dagger) = 0$, то есть существует (единственный) вектор β^\dagger , для которого условия на моменты выполнены.

A1(iv): $G_T(\beta)$ дифференцируема во внутренности \mathbf{B} , а $\hat{\beta}_T$ и β^\dagger находятся внутри \mathbf{B} .

A1(v): $\nabla_\beta G_T(\beta) - \mathbf{D}_\infty(\beta) \xrightarrow{p} 0$ равномерно по всем β в окрестности β^\dagger , а $\mathbf{D}_\infty(\beta)$ имеет полный ранг k и равномерно непрерывна по β для всех β в окрестности β^\dagger .

A1(vi): $\sqrt{T} \Omega_\infty(\beta^\dagger)^{-1/2} G_T(\beta^\dagger) \xrightarrow{d} N(0, I_p)$, причем $\Omega_\infty(\beta^\dagger) = \lim_{T \rightarrow \infty} \mathbb{V}[\sqrt{T} G_T(\beta^\dagger)]$, то есть функции моментов, оцененные в β^\dagger , подчиняются центральной предельной теореме.

Теорема 1 (см., например, Hansen, 1982)

(а) Пусть предположения А1(i)–(ii) выполнены. Тогда

$$\hat{\beta}_T \xrightarrow{p} \beta^\dagger.$$

(б) Пусть предположения А1(i)–(vi) выполнены. Тогда

$$\hat{\Sigma}_T^{-1/2} \sqrt{T} (\hat{\beta}_T - \beta^\dagger) \xrightarrow{d} N(0, I_k),$$

где

$$\hat{\Sigma}_T = \left(\nabla_\beta G_T(\hat{\beta}_T)' \hat{\Omega}_T(\hat{\beta}_T) \nabla_\beta G_T(\hat{\beta}_T) \right)^{-1}. \quad (3)$$

Таким образом, если нужно протестировать гипотезу $H_0 : \beta_i^\dagger = \beta_{0,i}^\dagger$ против $H_A : \beta_i^\dagger \neq \beta_{0,i}^\dagger$, можно просто построить t-статистику

$$t_{\beta_i,T} = \frac{\sqrt{T} \left(\widehat{\beta}_{i,T} - \beta_{0,i}^\dagger \right)}{\widehat{\sigma}_{ii,T}},$$

где $\widehat{\sigma}_{ii,T}^2 - (i, i)$ -ый элемент матрицы $\widehat{\Sigma}_T$. При выписанных выше предположениях $t_{\beta_i,T} \xrightarrow{d} N(0, 1)$, когда нулевая гипотеза истинна.

Аналогично можно построить тест Вальда для множественных ограничений или тестирования сверхидентифицирующих ограничений. Чтобы протестировать $H_0 : G_\infty(\beta^\dagger) = 0$ против $H_A : G_\infty(\beta^\dagger) \neq 0$, строят следующую статистику, известную как J-статистика:

$$J_T = TG_T(\widehat{\beta}_T)' \widehat{\Omega}_T(\widehat{\beta}_T) G_T(\widehat{\beta}_T),$$

и $J_T \xrightarrow{d} \chi_{p-k}^2$ при истинности H_0 .

Стандартным подходом является инференция на основе асимптотических критических значений, то есть на нормальных для t-тестов и хи-квадрат для J-тестов. Но насколько хорошо приближение нормальным? Некоторые эксперименты Монте-Карло указывают на то, что разница между номинальным и реальным уровнями теста довольно существенна для умеренного размера выборок. Отсюда вопрос: нельзя ли приближение улучшить? Нельзя ли уменьшить ошибку вероятности отклонения нулевой гипотезы? Мы увидим, что *бутстраповские критические значения* могут рафинировать асимптотические критические значения в разнообразных ситуациях.

Сначала мы очертим логику бутстрапа, а затем посмотрим, как использование бутстрапа может привести к более точной инференции в контексте тестов, основанных на ОММ-оценках. Ради краткости мы ограничим внимание t-статистиками.

3 Почему бутстрап работает: эвристика

Идея бутстрапа – прикинуться, что выборка и есть популяция, так что можно навтыгивать из выборки столько бутстраповских выборок, сколько душе угодно, и таким образом построить много бутстраповских статистик. Затем эмпирическое распределение последних можно использовать для получения более точных критических значений. Здесь и далее, пусть $t_{\beta_i,T}^*$ будет бутстраповским аналогом $t_{\beta_i,T}$.

При определенных мягких условиях регулярности (см., например, Hall, 1992, главы 2 и 3) можно выразить распределение t-статистики в виде лидирующего слагаемого, которое есть стандартная нормальная кумулятивная функция распределения, плюс другие слагаемые, ухватывающие отклонения от нормальности. Имеем *разложение Эджворта*

$$\mathbb{P}(t_{\beta_i,T} \leq x) = \Phi(x) + T^{-1/2} p_1(x) \phi(x) + T^{-1} p_2(x) \phi(x) + T^{-3/2} p_3(x) \phi(x) + \dots, \quad (4)$$

где $\Phi(x)$ и $\phi(x)$ – кумулятивная функция и плотность стандартного нормального распределения, оцененные в x , $p_1(x)$ – полином по x , зависящий от центрального третьего момента, $p_2(x)$ – полином по x , зависящий от четвертого момента минус 3, и т.д. Таким образом, $p_1(x)$ ухватывает отклонения от нормальности в форме скошенности, $p_2(x)$ ухватывает отклонения от нормальности в смысле эксцесса. Последующие слагаемые ухватывают более сложные отклонения и эффекты высокого порядка. Из (4) видно, что порядок приближения нормальным распределением $T^{-1/2}$.

Аналогично можно записать разложение Эджворта для $t_{\beta_i,T}^*$, учитывая, что бутстраповские моменты – это выборочные моменты:

$$\mathbb{P}(t_{\beta_i,T}^* \leq x) = \Phi(x) + T^{-1/2} p_1^*(x) \phi(x) + T^{-1} p_2^*(x) \phi(x) + T^{-3/2} p_3^*(x) \phi(x) + \dots, \quad (5)$$

где $p_1^*(x)$, $p_2^*(x)$ и т.д. – бутстраповские аналоги $p_1(x)$, $p_2(x)$ и т.д. При некоторых мягких предположениях, в основном касающихся существования достаточного количества моментов, равномерно по x , для $i = 1, 2, \dots$,

$$p_i^*(x) - p_i(x) = O_p(T^{-\delta/2}), \text{ где } 0 < \delta < 1.$$

Таким образом,

$$\mathbb{P}(t_{\beta_i, T} \leq x) - \mathbb{P}(t_{\beta_i, T}^* \leq x) = O_P(T^{-(1+\delta)/2}),$$

в то время как

$$\mathbb{P}(t_{\beta_i, T} \leq x) - \Phi(x) = O(T^{-1/2}).$$

Таким образом, при приближении $\mathbb{P}(t_{\beta_i, T} \leq x)$ стандартным нормальным распределением ошибка имеет порядок $O(T^{-1/2})$, в то время как при приближении $\mathbb{P}(t_{\beta_i, T} \leq x)$ бутстраповским распределением $\mathbb{P}(t_{\beta_i, T}^* \leq x)$ ошибка имеет порядок $O_P(T^{-(1+\delta)/2})$. В итоге бутстраповское распределение дает более точное приближение, чем нормальное.

На практике сравнивают не $\mathbb{P}(t_{\beta_i, T} \leq x)$ с $\Phi(x)$, а $t_{\beta_i, T}$ с z_α , $\alpha\%$ -ным критическим значением стандартного нормального распределения. Пусть $z_{T, \alpha}$ и $z_{T, \alpha}^*$ определяются как

$$\mathbb{P}(t_{\beta_i, T} \leq z_{T, \alpha}) = \alpha,$$

$$\mathbb{P}(t_{\beta_i, T}^* \leq z_{T, \alpha}^*) = \alpha.$$

Имея разложение Эджворта, всегда можно вывести *разложение Корнуса–Фишера* обращением:

$$z_{T, \alpha} = z_\alpha + T^{-1/2}q_1(z_\alpha) + T^{-1}q_2(z_\alpha) + T^{-3/2}q_3(z_\alpha) + \dots, \quad (6)$$

где $q_1(\alpha)$, $q_2(\alpha)$ – вновь полиномы по α , ухватывающие скошенность и эксцесс, и

$$z_{T, \alpha}^* = z_\alpha + T^{-1/2}q_1^*(z_\alpha) + T^{-1}q_2^*(z_\alpha) + T^{-3/2}q_3^*(z_\alpha) + \dots,$$

где $q_1^*(z_\alpha)$ и $q_2^*(z_\alpha)$ – бутстраповские аналоги $q_1(z_\alpha)$ и $q_2(z_\alpha)$. Так что, если для $i = 1, 2$ $q_i(z_\alpha) - q_i^*(z_\alpha) = O_P(T^{-\delta/2})$ при $0 < \delta < 1$, то для *равнохвостового t-теста*

$$\mathbb{P}(t_{\beta_i, T} < -z_{T, \alpha/2}^* \text{ или } t_{\beta_i, T} > z_{T, 1-\alpha/2}^*) = \alpha + O(T^{-(1+\delta)/2}),$$

в то время как для *симметричного t-теста*

$$\mathbb{P}(|t_{\beta_i, T}| < z_{T, \alpha/2}^*) = \alpha + O(T^{-1-\delta/2}).$$

Поэтому говорят, что инференция, основанная на бутстраповских критических значениях, точнее, чем если полагаться на асимптотические нормальные критические значения. Причина более «высокого» рафинирования симметричного теста в том, что в симметричном случае лидирующий член в разложении Эджворта из-за его нечетности равен нулю. Таким образом, улучшение ОВО при использовании бутстраповских критических значений имеет порядок $T^{-\delta/2}$, где $\delta \leq 1/2$.

Реальное значение, достигаемое δ , т.е. «степень» рафинирования, зависит от типа реализуемого бутстрапа, что в свою очередь зависит от длины памяти в данных.

4 Непараметрический бутстрап

Когда генерирующий данные процесс неизвестен, обычно ресэмплият в непараметрическом духе. Простейшая форма бутстрапа – *IID непараметрический бутстрап*, подходящий для IID-наблюдений.

Имея T наблюдений, из выборки вытягивают одно наблюдение за раз, с возвращением. Пусть $X_1^*, X_2^*, \dots, X_T^*$ будут ресэмплированными наблюдениями, и заметим, что $X_t^* = X_{I_t}$, $t = 1, \dots, T$ с вероятностью $1/T$. Другими словами, $X_1^*, X_2^*, \dots, X_T^*$ равен $X_{I_1}, X_{I_2}, \dots, X_{I_T}$, где для $i = 1, \dots, T$ I_i – случайная величина, принимающая значения $1, 2, \dots, T$ с равными вероятностями $1/T$. Набор $X_1^*, X_2^*, \dots, X_T^*$ составляет бутстраповскую выборку.

IID непараметрический бутстрап не работает с зависимыми наблюдениями. Причина в том, что ресэмплированные наблюдения IID, в то время как первоначальные наблюдения – нет. С одной стороны, хотелось бы вытягивать «блоки» данных, достаточно длинные, чтобы сохранить структуру зависимости, свойственную первоначальной выборке, а с другой стороны, хотелось бы иметь достаточно большое количество блоков, независимых друг от друга. Наиболее используемый метод ресэмплинга данных временных рядов – блочный бутстрап с перекрытием, предложенный в Künsch (1989).

Пусть $T = bl$, где b обозначает количество блоков, а l обозначает длину каждого блока. Сначала вытягивается дискретная равномерно распределенная случайная величина I_1 , принимающая значения $0, 1, \dots, T - l$ с вероятностью $1/(T - l + 1)$, тогда первым блоком становится $X_{I_1+1}, \dots, X_{I_1+l}$. Затем вытягивается другая дискретная случайная величина, скажем I_2 , и вторым блоком длины l становится $X_{I_2+1}, \dots, X_{I_2+l}$, и далее в том же духе. Наконец, вытягивается последняя дискретная случайная величина, скажем I_b , и последним блоком становится $X_{I_b+1}, \dots, X_{I_b+l}$. Обозначим за X_t^* ресэмплированный ряд и заметим, что $X_1^*, X_2^*, \dots, X_T^*$ соответствует $X_{I_1+1}, X_{I_1+2}, \dots, X_{I_b+l}$, так что, условно на выборке, единственным случайным элементом является начало каждого блока. В частности, $X_1^*, \dots, X_l^*, X_{l+1}^*, \dots, X_{2l}^*, X_{T-l+1}^*, \dots, X_T^*$, условно на выборке, можно интерпретировать как b IID-блоков из дискретного равномерного распределения.

Для случая, когда генерирующее данные распределение неизвестно, но процесс марковский или приблизительно марковский, Horowitz (2003) предложил использовать *марковский бутстрап*. В этом случае строится ядерная оценка условной плотности, и бутстраповские выборки вытягиваются согласно этой оценке плотности.

5 Улучшения высокого порядка для бутстраповских ОММ-тестов

Рассмотрим оценку, определенную в (1) в случае, когда

$$\begin{aligned} \hat{\Omega}_T(\bar{\beta}_T)^{-1} &= \frac{1}{T} \sum_{t=1}^T g_t(\bar{\beta}_T) g_t(\bar{\beta}_T)' \\ &+ \frac{1}{T} \sum_{t=\kappa+1}^T \sum_{j=1}^{\kappa} g_t(\bar{\beta}_T) g_{t-j}(\bar{\beta}_T)' + \frac{1}{T} \sum_{t=1}^{T-\kappa} \sum_{j=1}^{\kappa} g_t(\bar{\beta}_T) g_{t+j}(\bar{\beta}_T)'. \end{aligned} \quad (7)$$

Непосредственно видно, что $\hat{\Omega}_T(\bar{\beta}_T) - \Omega_\infty(\beta^\dagger) = o_p(1)$, если $\mathbb{E}[g_t(\beta^\dagger) g_{t+j}(\beta^\dagger)'] = 0$ при всех $j > \kappa$ для конечного κ . Действительно, бутстраповское рафинирование в случае $\kappa = \kappa_T$ для $\kappa_T \rightarrow \infty$ по мере того как $T \rightarrow \infty$ было показано в Inoue & Shintani (2006) для случая линейных условий на моменты.

Обозначим за $\hat{\sigma}_{ii,T}^2$ (i, i) -ый элемент матрицы $\hat{\Sigma}_T$, определенной в (3), где матрица $\hat{\Omega}_T(\hat{\beta}_T)$

определена в (7), но с заменой $\bar{\beta}_T$ на $\hat{\beta}_T$, и пусть

$$t_{\beta_i, T} = \frac{\sqrt{T} \left(\hat{\beta}_{i, T} - \beta_i^\dagger \right)}{\hat{\sigma}_{ii, T}}. \quad (8)$$

Положим $g_t(\beta) = g(y_t, X_t, \beta)$. Ресэмплим с возвращением b блоков длины l , соблюдая $bl = T - \kappa$, из $(\tilde{y}_t, \tilde{X}_t)$, где $\tilde{y}_t = (y_t, \dots, y_{t+\kappa})$ и аналогично $\tilde{X}_t = (X_t, \dots, X_{t+\kappa})$, и получаем в результате $(\tilde{y}_t^*, \tilde{X}_t^*)$, где $\tilde{y}_t^* = (y_t^*, \dots, y_{t+\kappa}^*)$, и $\tilde{X}_t^* = (X_t^*, \dots, X_{t+\kappa}^*)$. Здесь и далее \mathbb{E}^* и \mathbb{V}^* обозначают взятие среднего и дисперсии по отношению к бутстраповскому распределению вероятностей, условно на выборке.

Пусть

$$g_t^{**}(\beta) = g(y_t^*, X_t^*, \beta) - \mathbb{E}^*[g(y_t^*, X_t^*, \hat{\beta}_T)], \quad (9)$$

где

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}^*[g(y_t^*, X_t^*, \hat{\beta}_T)] = \frac{1}{T-l+1} \sum_{t=1}^T w_t g(y_t, X_t, \hat{\beta}_T)$$

и

$$w_t = \frac{t}{l}, \quad t = 1, \dots, l-1,$$

$$w_t = 1, \quad t = l, \dots, T-l+1,$$

$$w_t = \frac{T-t+1}{l}, \quad t = T-l+2, \dots, T.$$

Вес w_t меньше единицы для первых и последних l наблюдений, поскольку их вытянуть шансов меньше. Заметим, что в общем случае у $g(y_t^*, X_t^*, \hat{\beta}_T)$ ненулевое среднее, даже если у $g(y_t, X_t, \beta^\dagger)$ оно нулевое, отсюда необходимость рецентрировать бутстраповские условия на моменты. В самом деле, $\mathbb{E}^*[g_t^{**}(\hat{\beta}_T)] = 0$.

Теперь определим бутстраповский аналог $\bar{\beta}_T$ как

$$\bar{\beta}_T^* = \arg \min_{\beta \in \mathbf{B}} \left(\frac{1}{T} \sum_{t=1}^T g_t^{**}(\beta) \right)' \Omega \left(\frac{1}{T} \sum_{t=1}^T g_t^{**}(\beta) \right),$$

где $g_t^{**}(\beta)$ определено в (9). Определим также бутстраповский аналог $\hat{\beta}_T$ как

$$\begin{aligned} \hat{\beta}_T^* &= \arg \min_{\beta \in \mathbf{B}} \left(\frac{1}{T} \sum_{t=1}^T g_t^{**}(\beta) \right)' \hat{\Omega}_T^*(\bar{\beta}_T^*) \left(\frac{1}{T} \sum_{t=1}^T g_t^{**}(\beta) \right) \\ &= \arg \min_{\beta \in \mathbf{B}} G_T^{**}(\beta)' \hat{\Omega}_T^*(\bar{\beta}_T^*) G_T^{**}(\beta), \end{aligned} \quad (10)$$

где

$$\begin{aligned} \hat{\Omega}_T^*(\bar{\beta}_T^*)^{-1} &= \frac{1}{T} \sum_{t=1}^T g_t^{**}(\bar{\beta}_T^*) g_t^{**}(\bar{\beta}_T^*)' \\ &\quad + \frac{1}{T} \sum_{t=\kappa+1}^T \sum_{j=1}^{\kappa} g_t^{**}(\bar{\beta}_T^*) g_{t-j}^{**}(\bar{\beta}_T^*)' + \frac{1}{T} \sum_{t=1}^{T-\kappa} \sum_{j=1}^{\kappa} g_t^{**}(\bar{\beta}_T^*) g_{t+j}^{**}(\bar{\beta}_T^*)'. \end{aligned}$$

Таким образом, $\hat{\Omega}_T^*(\bar{\beta}_T^*)$ – бутстраповский аналог $\hat{\Omega}_T(\bar{\beta}_T)$.

Бутстраповская дисперсионная матрица задается как

$$\widehat{\Sigma}_T^* = \left(D_T^*(\widehat{\beta}_T^*)' \widehat{\Omega}_T^*(\widehat{\beta}_T^*) D_T^*(\widehat{\beta}_T^*) \right)^{-1},$$

где

$$D_T^*(\widehat{\beta}_T^*) = \frac{1}{T} \sum_{t=1}^T \frac{\partial g_t^{**}(\widehat{\beta}_T^*)}{\partial \beta'}.$$

Нам интересно тестирование $H_0 : \beta_i = \beta_i^\dagger$ против $H_A : \beta_i \neq \beta_i^\dagger$. Бутстраповский аналог $t_{\beta_i, T}$, определенной в (8), есть

$$t_{\beta_i, T}^* = \frac{\sqrt{T} \left(\widehat{\beta}_{i, T}^* - \widehat{\beta}_{i, T} \right)}{\widehat{\sigma}_{ii, T}^*}, \quad (11)$$

где $\widehat{\sigma}_{ii, T}^{*2}$ — (i, i) -ый элемент матрицы $\widehat{\Sigma}_T^*$. Но хотя $\widehat{\sigma}_{ii, T}^{*2}$ и есть бутстраповский аналог $\widehat{\sigma}_{ii, T}^2$, эта величина не совпадает с $\mathbb{V}^*[T^{1/2}(\widehat{\beta}_{i, T}^* - \widehat{\beta}_{i, T})]$. Почему? Зависимость в выборочной и бутстраповской функциях моментов не одна и та же. Это происходит из-за так называемой *проблемы точек соприкосновения*. Блоки независимы условно на выборке, и последнее наблюдение в блоке и первое наблюдение последующего блока нескоррелированы, хотя это не так в первоначальной выборке. Всего точек соприкосновения b (столько же, сколько и блоков), и их надо учесть. Суть в том, что $\widehat{\sigma}_{ii, T}^{*2}$ верно имитирует $\widehat{\sigma}_{ii, T}^2$ (т.е. $\mathbb{E}^*[\widehat{\sigma}_{ii, T}^{*2}] = \widehat{\sigma}_{ii, T}^2$), но $\widehat{\sigma}_{ii, T}^{*2}$ равно $\mathbb{V}^*[T^{1/2}(\widehat{\beta}_{i, T}^* - \widehat{\beta}_{i, T})]$.

Таким образом, нужен корректирующий множитель. Определим

$$\begin{aligned} \widetilde{\sigma}_{ii, T}^2 &= \left(D_T(\widehat{\beta}_T)' \widetilde{\Omega}_T(\widehat{\beta}_T) D_T(\widehat{\beta}_T) \right)^{-1} D_T(\widehat{\beta}_T)' \\ &\quad \times \widetilde{\Omega}_T(\widehat{\beta}_T) \widetilde{\Omega}_T(\widehat{\beta}_T)^{-1} \widetilde{\Omega}_T(\widehat{\beta}_T) D_T(\widehat{\beta}_T) \left(D_T(\widehat{\beta}_T)' \widetilde{\Omega}_T(\widehat{\beta}_T) D_T(\widehat{\beta}_T) \right)^{-1}, \end{aligned}$$

где

$$\begin{aligned} \widetilde{\Omega}_T(\widehat{\beta}_T)^{-1} &= \mathbb{E}^* \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T g_t^{**}(\widehat{\beta}_T) g_s^{**}(\widehat{\beta}_T)' \right] \\ &= \frac{1}{l(T-l-1)} \sum_{t=0}^{T-l} \sum_{j=1}^l \sum_{i=1}^l g_{t+j}^{**}(\widehat{\beta}_T) g_{t+i}^{**}(\widehat{\beta}_T)'. \end{aligned}$$

Заметим, что $\widetilde{\sigma}_{ii, T}^2 = \mathbb{V}^*[T^{1/2}(\widetilde{\beta}_{i, T}^* - \widetilde{\beta}_{i, T})]$. Корректирующий множитель задается как

$$\tau_{ii, T} = \frac{\widehat{\sigma}_{ii, T}}{\widetilde{\sigma}_{ii, T}}.$$

Теперь рассмотрим подправленную бутстраповскую статистику

$$\widetilde{t}_{\beta_i, T}^* = \frac{\sqrt{T} \left(\widehat{\beta}_{i, T}^* - \widehat{\beta}_{i, T} \right) \widehat{\sigma}_{ii, T}}{\widehat{\sigma}_{ii, T}^* \widetilde{\sigma}_{ii, T}}, \quad (12)$$

то есть произведение бутстраповского аналога t -статистики и корректирующего множителя. Заметим, что в случае IID-бутстрапа нет проблемы точек соприкосновения, и поэтому необходимости в корректирующем множителе нет.

Предположения A1 недостаточно для бутстраповского рафинирования. В то время как полный список достаточных условий содержится в предположениях 1–5 в Andrews (2002),

ниже мы лишь опишем в общих чертах предположения, которые необходимо добавить к А1 выше.

Предположение А2

А2(i): $\mathbb{E}[g_t(\beta^\dagger) g_{t+j}(\beta^\dagger)'] = 0$ for all $j > \kappa$.

А2(ii): (y_t, X_t) – стационарный ряд с сильным перемешиванием и экспоненциально убывающими коэффициентами перемешивания (см., например, предположение 1 в Andrews, 2002 или в Hall & Horowitz, 1996).

А2(iii): Разложение Эджворта для t-статистики и ее бутстраповского аналога существуют.

А2(iv): Для $f_t(\beta) = (g_t(\beta), g_t(\beta) g_{t-j}(\beta)', \partial^i g_t(\beta) / \partial \beta^i, \partial^i g_t(\beta) g_{t-j}(\beta)' / \partial \beta^i, j \leq \kappa, i \geq d_1)$, производные вплоть до порядка d_2 имеют конечные моменты и удовлетворяют условию Липшица.

Тогда имеем:

Теорема 2 (из теоремы 2 в Andrews, 2002)

(а) Пусть А1 и А2 выполнены, причем $d_1 \geq 5$ и $d_2 \geq 4$. Пусть $l = T^\gamma$, и предположим, что $0 \leq \xi \leq 1/2 - \gamma$ и $\xi < \gamma$, тогда

$$\mathbb{P}\left(|t_{\beta_i, T}| < \tilde{z}_{T, \alpha/2}^*\right) = \alpha + O\left(T^{-(1+\xi)}\right),$$

где $\tilde{z}_{T, \alpha/2}^*$ таково, что $\mathbb{P}\left(\tilde{t}_{\beta_i, T}^* \leq \tilde{z}_{T, \alpha/2}^*\right) = \alpha/2$, где $t_{\beta_i, T}$ и $\tilde{t}_{\beta_i, T}^*$ определены в (8) и (12).

(б) Пусть А1 и А2 выполнены, причем $d_1 \geq 4$ и $d_2 \geq 3$. Пусть $l = T^\gamma$, и предположим, что $0 \leq \xi \leq 1/2 - \gamma$ и $\xi < \gamma$, тогда

$$\mathbb{P}\left(t_{\beta_i, T} < -\tilde{z}_{T, \alpha/2}^* \text{ или } t_{\beta_i, T} > \tilde{z}_{T, 1-\alpha/2}^*\right) = \alpha + O\left(T^{-1/2+\xi}\right).$$

Доказательство теоремы 2 основано на следующих шагах. Во-первых, $t_{\beta_i, T}$ можно приблизить гладкой функцией, скажем G , от $f_t(\beta^\dagger)$, определенной в А2(iv), а бутстраповскую статистику без корректирующего множителя $t_{\beta_i, T}^*$ можно приблизить $G(f_t^*(\hat{\beta}_T))$, где $f_t^*(\hat{\beta}_T)$ определяется так же, как и $f_t(\hat{\beta}_T)$, но с заменой выборочных функций моментов на бутстраповские. Тогда, если выполнено А2(iii), для $G(f_t(\beta^\dagger))$ и $G(f_t^*(\hat{\beta}_T))$ существуют разложения Эджворта, и, если выполнены моментные и липшицевы условия в А2(iv), разница между первыми двумя членами разложений Эджворта для $G(f_t(\beta^\dagger))$ и $G(f_t^*(\hat{\beta}_T))$ сходится к нулю достаточно быстро. Наконец, если $\xi < \gamma$, корректирующий множитель сходится к единице достаточно быстро, гарантируя, что разложения Эджворта для скорректированной бутстраповской статистики $\tilde{t}_{\beta_i, T}^*$ и для $t_{\beta_i, T}$ сближаются.

Из теоремы 2 сразу видно, что если положить $\gamma = 1/4$, т.е. $l = T^{1/4}$, то ξ можно сделать сколь угодно близким к $1/4$. Таким образом, бутстраповское улучшение в ОВО имеет порядок $T^{-\xi}$, где $0 < \xi < 1/4$. Условие $\xi < \gamma$ гарантирует, что по мере того как $T \rightarrow \infty$, корректирующий множитель $\tau_{ii, T} \rightarrow 1$. Как уже упомянуто, в случае IID-наблюдений в корректировке нет необходимости, так что условие $\xi < \gamma$ не требуется. Таким образом, можно положить $\gamma = 0$ (т.е. $l = 1$), так что $\xi = 1/2$, что приводит к улучшению в ОВО порядка $T^{-1/2}$.

Если функция моментов – последовательность мартингалных приращений, как в случае динамической (верно специфицированной) модели, то $\kappa = 0$, хотя все равно нужно использовать длину блока l , удовлетворяющую $l \rightarrow \infty$, для того чтобы ухватить зависимость в более высоких (чем вторые) моментах.

6 Как построить бутстраповские критические значения

(а) На практике мы не знаем бутстраповское критическое значение $\tilde{z}_{T,\alpha/2}^*$. Стандартный подход – построить B бутстраповских статистик, скажем $\hat{t}_{\beta_i,T}^{*(j)}$, $j = 1, \dots, B$, и взять $\tilde{z}_{T,B,\alpha/2}^*$ в качестве $(1 - \alpha/2)$ -квантили эмпирического распределения $(\hat{t}_{\beta_i,T}^{*(1)}, \dots, \hat{t}_{\beta_i,T}^{*(B)})$. Проблема в том, как выбрать B достаточно большим, чтобы гарантировать, что инференции, основанные на $\tilde{z}_{T,\alpha/2}^*$ и $\tilde{z}_{T,B,\alpha/2}^*$, приводили к одним и тем же улучшениям высокого порядка. Вопрос оптимального выбора количества бутстраповских повторов рассмотрен, например, в Davidson & MacKinnon (2000) и Andrews & Buchinski (2000).

(б) Построение бутстраповской статистики требует выбора длины блока l . Адаптивная процедура выбора l предложена в Hall, Horowitz & Jing (1995).

(в) Вычисление распределения бутстраповской оценки $\hat{\beta}_T^*$ может быть трудоемким, так как оно включает решение B нелинейных оптимизационных задач. Davidson & MacKinnon (1999) предложили альтернативную k -шаговую оценку. По сути, нужно положить $\hat{\beta}_T^{*(0)} = \hat{\beta}_T$ и совершить k шагов по направлению к $\hat{\beta}_T^*$ согласно, например, алгоритму Ньютона–Рафсона. Andrews (2002, теорема 1) показал, что инференция, полагающаяся на $\hat{z}_{T,k,\alpha/2}^*$, т.е. на критические значения, основанные на $\hat{\beta}_T^{*(k)}$, приводит к того же порядка рафинированию, как и инференция, основанная на $\hat{\beta}_T^*$, при $k \geq 3$ или $k \geq 4$ в зависимости от того, рассматривается симметричный или же равнохвостовой тест.

7 Улучшенное рафинирование

Блочно-блочный бутстрап

Как утверждает теорема 2, блочный бутстрап дает рафинирование ОВО вплоть до порядка $T^{-\xi}$, где $\xi < 1/4$, в то время как IID-бутстрап рафинирует до порядка $T^{-1/2}$. Одна из причин – это упомянутая выше проблема точек соприкосновения. Andrews (2004) предложил строить блочные статистики так, чтобы одна и та же проблема точек соприкосновения имела бы и в бутстраповской, и в первоначальной выборках. Другими словами, статистика подсчитывается с удалением πl наблюдений, непосредственно предшествующих точкам соприкосновения $l+1, 2l+1, \dots, (b-1)l+1$, где, по мере того как $T \rightarrow \infty$, $\pi \rightarrow 0$ и $\pi l \rightarrow \infty$. Поскольку выборка обладает свойством сильного перемешивания, $l(1 - \pi)$ -е и $(l+1)$ -е наблюдения становятся независимыми по мере того как $\pi l \rightarrow \infty$. Тогда уже нет необходимости в корректирующем множителе, и $\gamma > \xi$ уже не требуется. Таким образом, можно выбрать $\gamma < 1/4$, допуская этим $\xi > 1/4$. Тем не менее, необходимо выбирать достаточно большую длину блока, чтобы ухватить зависимость в данных.

Марковский бутстрап

Если генерирующий данные процесс марковский, или его можно приблизить марковским, стоит воспользоваться марковским бутстрапом, предложенном в Horowitz (2003). Суть его в том, что по выборочным наблюдениям строится ядерная оценка условной плотности; затем бутстраповские выборки вытягиваются из оцененной условной плотности. При мягких условиях регулярности марковский бутстрап обеспечивает рафинирование ОВО порядка $T^{1/2-\varepsilon}$ для сколь угодно малого ε .

Список литературы

- Andrews, D.W.K. & M. Buchinski (2000). A three step method for choosing the number of bootstrap repetition. *Econometrica* 68, 23–51.
- Andrews, D.W.K. (2002). Higher order improvements of a computationally attractive k -step bootstrap for extremum estimators. *Econometrica* 70, 119–162.
- Andrews, D.W.K. (2004). The block-block bootstrap: Improved asymptotic refinements. *Econometrica* 72, 673–700.
- Davidson, R. & J.G. Mackinnon (1999). Bootstrap testing in nonlinear models. *International Economic Review* 40, 487–508.
- Davidson, R. & J.G. Mackinnon (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews* 19, 55–68.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Hall, P., J.L. Horowitz & B.Y. Jing (1995). On blocking rules for bootstrap with dependent data. *Biometrika* 82, 561–574.
- Hall, P. & J.L. Horowitz (1996). Bootstrap critical values for tests based on generalized method of moments estimators. *Econometrica* 64, 891–916.
- Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Horowitz, J.L. (2003). Bootstrap methods for Markov processes. *Econometrica* 71, 1049–1082.
- Inoue, A. & M. Shintani (2006). Bootstrapping GMM estimators for time series. *Journal of Econometrics* 133, 531–555.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.

Bootstrap refinements for GMM based tests

Valentina Corradi

University of Warwick, United Kingdom

This essay provides a brief review about bootstrap higher order refinements for tests based on generalized method of moments estimators. First, we briefly describe the asymptotic behavior of two-step GMM estimators. Second, we give a heuristic argument for why inference based on bootstrap critical values is more accurate than that based on asymptotic normality. Third, we briefly summarize nonparametric resampling methods. Fourth, we outline how critical values based on the block bootstrap reduce the error in the rejection probability for t-tests based on GMM estimators. Finally, we give an overview of some alternative bootstrap procedures which provide improvements over the block bootstrap refinements.