

# Бутстрапирование эконометрических моделей\*

Расселл Дэвидсон<sup>†</sup>

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.

<sup>†</sup>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](mailto: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  $\alpha$  критическое значение – это  $(1 - \alpha)$ -квантиль распределения статистики при нулевой гипотезе. Для реализации  $\tau$  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  $\alpha$ -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  $\tau$  равно  $2 \min(F(\tau), 1 - F(\tau))$ .

Есть другие способы построения критических значений для двусторонних тестов. Если использовать  $\beta$ - и  $\gamma$ -квантили в качестве нижнего и верхнего критических значений соответственно, достаточно выполнения равенства  $1 - \gamma + \beta = \alpha$ , чтобы уровень значимости был равен  $\alpha$ . Можно также выбирать  $\beta$  и  $\gamma$ , 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  $\alpha$  в точности равна  $\alpha$ , когда нулевая гипотеза истинна. Если инференция основана на *P-значении*, то для любого  $\alpha$  между 0 и 1 вероятность получения *P-значения*, меньшего, чем  $\alpha$ , при истинности нулевой гипотезы, в точности равна  $\alpha$ .

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  $\alpha$  таков, что  $\alpha(B + 1)$  является целым числом. Чтобы убедиться в этом, заметим, что бутстраповские статистики, обозначаемые  $\tau_j^*$ ,  $j = 1, \dots, B$ , как и статистика  $\tau$ , 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  $\tau$ , 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  $\tau$ , а эта вероятностная масса есть просто  $r/B$ .

Тогда вероятность получения бутстраповского *P-значения*, меньшего, чем  $\alpha$ , равна  $\mathbb{P}(r < \alpha B)$ . Пусть  $\lceil \alpha B \rceil$  – наименьшее целое число, не меньшее, чем  $\alpha B$ . Тогда число возможных значений  $r$ , (строго) меньших  $\alpha B$ , равно  $\lceil \alpha B \rceil$ . Следовательно,  $\mathbb{P}(r < \alpha B) = \lceil \alpha B \rceil / (B + 1)$ . Эта вероятность равна  $\alpha$  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  $\tau$  асимптотически пивотальна для данной модели  $\mathbb{M}$ , ее распределение не должно слишком сильно меняться как функция конкретного DGP, скажем,  $\mu$ , в

рамках этой модели. Обычно можно показать, что расстояние между распределением  $\tau$  для DGP  $\mu$  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  $\mu_1$  и  $\mu_2$  также имеет порядок  $n^{-\delta}$ . Если расстояние между  $\mu_1$  и  $\mu_2$  также мало в некотором смысле, скажем, имеет порядок  $n^{-\epsilon}$ , то расстояние между распределениями  $\tau$  при  $\mu_1$  и при  $\mu_2$  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  $\mu^*$ . 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  $\mu^*$  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  $\mu^*$ .
- (iii) Сгенерировать  $B$  бутстраповских выборок, используя  $\mu^*$ , и для каждой из них рассчитать реализацию бутстраповской статистики  $\tau_j^*$ ,  $j = 1, \dots, B$ . Разумно выбирать  $B$  так, чтобы  $\alpha(B + 1)$  было целым числом для всех интересующих уровней значимости  $\alpha$ , обычно 1%, 5% и 10%.
- (iv) Рассчитать симулированное бутстраповское  $P$ -значение как долю бутстраповских статистик  $\tau_j^*$ , 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  $\alpha$ , 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  $\alpha$ ,  $\rho$  и  $\sigma^2$ . 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  $\alpha$ ,  $\rho$  и  $\sigma$ . 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  $\rho$  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

потребовать 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  $\tau$ , уже имеет форму асимптотического  $P$ -значения. Тогда отвержение на уровне значимости  $\alpha$  соответствует событию  $\tau < \alpha$ .

Введем две функции номинального уровня значимости теста  $\alpha$  и DGP  $\mu$ . Первая из них – это *функция вероятности отвержения*, или ФВО. Значение этой функции – это истинная вероятность отвержения нулевой гипотезы при  $\mu$  с помощью теста на уровне значимости  $\alpha$  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  $\tau$  имеет носитель  $[0, 1]$  и абсолютно непрерывно по отношению к равномерному распределению на этом интервале.

Для конкретного  $\mu$  функция  $R(\alpha, \mu)$  – это просто функция распределения  $\tau$ , подсчитанная в  $\alpha$ . 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  $\alpha$ -квантиль распределения  $\tau$  при  $\mu$ . Кроме того, из определений (4) и (5) следует, что

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

для всех  $\alpha$  и  $\mu$ .

Далее абстрагируемся от случайности симуляции и предположим, что распределение  $\tau$  при бутстраповском DGP точно известно. Бутстраповское критическое значение для  $\tau$  при уровне значимости  $\alpha$  равно  $Q(\alpha, \mu^*)$ ; напомним, что  $\mu^*$  обозначает бутстраповский DGP. Это случайная величина, которая была бы неслучайной и равной  $\alpha$ , если бы  $\tau$  была полностью пивотальной. Если  $\tau$  приблизительно (например, асимптотически) пивотальна, реализации  $Q(\alpha, \mu^*)$  должны быть близки к  $\alpha$ . Это верно независимо от того, соответствует ли истинный DGP нулевой гипотезе, поскольку бутстраповский DGP  $\mu^*$  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  $\mu$  того, что бутстраповский тест отвергает нулевую гипотезу при номинальном уровне значимости  $\alpha$ , 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,  $\tau$  и  $\mu^*$ , необходимых для расчета бутстраповского  $P$ -значения  $R(\tau, \mu^*)$ . Первая из этих случайных переменных распределена как  $U[0, 1]$  при  $\mu$ ; это

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

Равномерность распределения  $p$  следует из того, что  $R(\cdot, \mu)$  – функция распределения  $\tau$  при  $\mu$ , и из предположения о том, что распределение  $\tau$  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  $\alpha$ , 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  $\mu$  условно на случайной величине  $p$  за  $F(r|p)$ . Тогда вероятность при  $\mu$  отвергнуть нулевую гипотезу бутстраповским тестом на уровне значимости  $\alpha$  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  $\alpha$  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  $\alpha$ . 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  $\mu$ , меньше  $\alpha$ -квантиля бутстраповского распределения условно на этом распределении. Матожидание  $r$  за вычетом  $\alpha$  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  $\alpha$ -квантиля  $\tau$  при  $\mu$ , тогда как точное выражение (10) учитывает тот факт, что грань на самом деле определяется бутстраповским DGP.

Если статистика  $\tau$  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  $\alpha$  одно и то же для всех  $\mu \in \mathbb{M}$ , аналогично и для  $Q(\alpha, \mu)$ . Обозначим предельные функции от  $\alpha$  (только) за  $R^\infty(\alpha)$  и  $Q^\infty(\alpha)$ . При предположении об абсолютной непрерывности распределения функции  $R^\infty$  и  $Q^\infty$  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  $\sigma^2$ .

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)<sup>1</sup>. 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  $\rho$ , 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

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

<sup>1</sup>См. «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  $\tau$ .

### 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  $\alpha$  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  $\alpha$  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  $\Delta_0$ . 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  $\Delta_0$ . 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  $\infty$ , 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  $\beta$  или  $\beta_2$  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  $\beta_1$  содержит элементы  $\beta$ , не принадлежащие  $\beta_2$ , а  $\tilde{\beta}_1$  – МНК-оценка модели *с ограничением*. Аналогично,  $\mathbf{X}_{s1}$  содержит элементы  $\mathbf{X}_s$ , коэффициенты при которых являются элементами  $\beta_1$ . 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  $1 \times m$ ,  $\Pi_i$  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  $\Pi_i$ ,  $i = 1, \dots, p$  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*  $\Sigma$  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  $\Pi_i$ . При отсутствии ограничений модель (21) можно эффективно оценить, применив МНК к каждому уравнению по отдельности и взяв в качестве оценки ковариационной матрицы  $\Sigma$  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  $\Pi_i$ ,  $\Sigma$  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  $1 \times k$ ,  $\Gamma$  – матрица  $m \times m$ , а  $\mathbf{B}$  – матрица  $k \times m$ . Элементы  $\Gamma$  и  $\mathbf{B}$ , наряду с независимыми элементами одномоментной ковариационной матрицы  $\Sigma$ , являются параметрами модели. Чтобы эндогенные переменные  $\mathbf{Y}_t$  задавались с помощью (22),  $\Gamma$  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  $\Pi$  – матрица параметров  $k \times m$ . Однако часто структурная форма является *сверхидентифицированной*, когда на матрицы  $\Gamma$  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 заиклиться, 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  $\theta$ . Предположим, что для каждого возможного значения параметра существует тестовая статистика  $\tau(\theta)$  с известным, хотя бы асимптотически, распределением, когда  $\theta$  является истинным параметром. Доверительный

интервал при *уровне доверия*  $1 - \alpha$ ,  $0 < \alpha < 1$ , – это множество значений  $\theta$ , для которых гипотеза о том, что  $\theta$  – истинный параметр, не отвергается на уровне значимости  $\alpha$  с помощью теста на основе статистики  $\tau(\theta)$ . Пусть  $C$  обозначает таким образом найденный доверительный интервал. Тогда, если истинный параметр равен  $\theta$ , 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  $\alpha$ . 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  $C_\alpha$ , тест на уровне значимости  $\alpha$  отвергает гипотезу о том, что  $\theta$  – 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  $\theta$  называется *процентильным методом*. Предполагается модель  $\mathbb{M}$ , и для каждого DGP  $\mu \in \mathbb{M}$  существует соответствующий параметр  $\theta(\mu)$ , «истинный» параметр для DGP  $\mu$ . Поскольку обычно имеются иные параметры помимо  $\theta$ , их необходимо оценить вместе с  $\theta$ , 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  $\theta$  – 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  $\theta$  с вероятностью  $\alpha$ .

На следующем шаге генерируется множество бутстраповских выборок и для каждой из них рассчитывается оценка параметра, скажем  $\theta^*$ . Поскольку истинное значение  $\theta$  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  $\sigma_\theta^2$ .

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  $\theta$ . 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  $\sigma_{\theta}^*$  – 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  $\alpha$  могут быть основаны на  $\beta$  и  $\gamma$ -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  $\beta$  и  $\gamma$ , минимизирующие расстояние между  $\beta$ -квантилем и  $\gamma$ -квантилем при ограничении  $1 - \gamma + \beta = \alpha$ , тогда не равны  $\alpha/2$  и  $1 - \alpha/2$ , вообще говоря. Использование  $\beta$  и  $\gamma$ , 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  $\theta$ . 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  $\phi$  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  $\theta$  – 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  $C_\alpha$  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  $\theta$  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)  $\theta$  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  $\theta$ , 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  $\theta$  – истинный параметр, одинаково для всех  $\theta$ . Если это так, то статистика называется пивотальной и дальнейших проблем нет. Но если статистика является лишь приблизительно пивотальной, ее распределение, когда истинное значение  $\theta$  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  $\theta$  является истинным параметром. Если  $\theta$  является единственным параметром, то возможно, хотя обычно и нелегко, решить (27) численно на основе симуляций. В общем случае, однако, ситуация еще сложнее. Если

помимо  $\theta$  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  $\theta$ , рассматриваемого при поиске решения (27), следует переоценивать эти мешающие параметры при ограничении, что  $\theta$  – истинный параметр и затем основывать бутстраповский DGP на  $\theta$  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*