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 уменьшение 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 обрисуем достаточные условия асимптотической нормальности двухшаговых ОММ-оценок.

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

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

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

$$\overline{\boldsymbol{\beta}}_T = \arg\min_{\beta\in B} G_T(\boldsymbol{\beta})' \Omega G_T(\boldsymbol{\beta}), \tag{2}$$

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 $\widehat{\Omega}_T\left(\overline{\boldsymbol{\beta}}_T\right)$ называется *оцененной взвешивающей.*

Определим

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

где $G_\infty(\boldsymbol{\beta})$ и Ω_∞ – пределы по вероятности (или почти наверное) $G_T(\boldsymbol{\beta})$ и $\widehat{\Omega}_T\left(\overline{\boldsymbol{\beta}}_T\right)$, соответственно.

Предположение A1

A1(i): $\sup_{\beta\in B} |G_T(\boldsymbol{\beta}) - G_\infty(\boldsymbol{\beta})| \xrightarrow{p} 0$ и $\widehat{\Omega}_T\left(\overline{\boldsymbol{\beta}}_T\right) \xrightarrow{p} \Omega_\infty\left(\boldsymbol{\beta}^\dagger\right)$, где матрица $\Omega_\infty\left(\boldsymbol{\beta}^\dagger\right)$ 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): $G_\infty(\boldsymbol{\beta}^\dagger)'\Omega_\infty\left(\boldsymbol{\beta}^\dagger\right) G_\infty(\boldsymbol{\beta}^\dagger) < G_\infty(\boldsymbol{\beta})'\Omega_\infty\left(\boldsymbol{\beta}^\dagger\right) G_\infty(\boldsymbol{\beta})$, т.е. имеется идентификация с единственностью.

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

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

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

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

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

(а) Пусть предположения A1(i)–(ii) выполнены. Тогда

$$\widehat{\boldsymbol{\beta}}_T \xrightarrow{p} \boldsymbol{\beta}^\dagger.$$

(б) Пусть предположения A1(i)–(vi) выполнены. Тогда

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

где

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

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 $H_0 : G_\infty(\boldsymbol{\beta}^\dagger) = 0$ против $H_A : G_\infty(\boldsymbol{\beta}^\dagger) \neq 0$, строят следующую статистику, известную как J-статистика:

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

и $J_T \xrightarrow{d} \chi^2_{p-k}$ 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 навытягивать 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*

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

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:

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

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,

$$\mathbb{P}\left(t_{\beta_i,T} \le x\right) - \mathbb{P}(t^*_{\beta_i,T} \le x) = O_P(T^{-(1+\delta)/2}),$$

в то время как

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

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

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

$$\mathbb{P}\left(t_{\beta_i,T} \le z_{T,\alpha}\right) = \alpha,$$

$$\mathbb{P}(t^*_{\beta_i,T} \le z^*_{T,\alpha}) = \alpha.$$

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

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

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 $\delta \le 1/2$.

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 ресэмплят 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 $X_1^*, X_2^*, \ldots, X_T^*$ будут ресэмпленными наблюдениями, и заметим, что $X_1^* = X_t$, $t = 1, \ldots, T$ с вероятностью $1/T$. Другими словами, $X_1^*, X_2^*, \ldots, X_T^*$ равен $X_{I_1}, X_{I_2}, \ldots, X_{I_T}$, где для $i = 1, \ldots, T$ I_i – случайная величина, принимающая значения $1, 2, \ldots, T$ с равными вероятностями $1/T$. Набор $X_1^*, X_2^*, \ldots, X_T^*$ составляет бутстраповскую выборку.

IID непараметрический бутстрап не работает с зависимыми наблюдениями. Причина в том, что ресэмпленные 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 $0, 1, \ldots, T - l$ с вероятностью $1/(T - l + 1)$, тогда первым блоком становится $X_{I_{1+1}}, \ldots, X_{I_1+l}$. Затем вытягивается другая дискретная случайная величина, скажем I_2, и вторым блоком длины l становится $X_{I_2+1}, \ldots, X_{I_2+l}$, и далее в том же духе. Наконец, вытягивается последняя дискретная случайная величина, скажем I_b, и последним блоком становится $X_{I_b+1}, \ldots, X_{I_b+l}$. Обозначим за X_t^* ресэмпленный ряд и заметим, что $X_1^*, X_2^*, \ldots, X_T^*$ соответствует $X_{I_1+1}, X_{I_1+2}, \ldots, X_{I_b+l}$, так что, условно на выборке, единственным случайным элементом является начало каждого блока. В частности, $X_1^*, \ldots, X_l^*$, $X_{l+1}^*, \ldots, X_{2l}^*, X_{T-l+1}^*, \ldots, X_T^*$, 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

$$\begin{aligned}\widehat{\Omega}_T\left(\overline{\boldsymbol{\beta}}_T\right)^{-1} &= \frac{1}{T}\sum_{t=1}^{T} g_t\left(\overline{\boldsymbol{\beta}}_T\right) g_t\left(\overline{\boldsymbol{\beta}}_T\right)' \\ &+ \frac{1}{T}\sum_{t=\kappa+1}^{T}\sum_{j=1}^{\kappa} g_t\left(\overline{\boldsymbol{\beta}}_T\right) g_{t-j}\left(\overline{\boldsymbol{\beta}}_T\right)' + \frac{1}{T}\sum_{t=1}^{T-\kappa}\sum_{j=1}^{\kappa} g_t\left(\overline{\boldsymbol{\beta}}_T\right) g_{t+j}\left(\overline{\boldsymbol{\beta}}_T\right)'. \end{aligned} \tag{7}$$

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

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

определена в (7), но с заменой $\overline{\boldsymbol{\beta}}_T$ на $\widehat{\boldsymbol{\beta}}_T$, и пусть

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

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

Пусть

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

где

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

и

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

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

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

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

Теперь определим бутстраповский аналог $\overline{\boldsymbol{\beta}}_T$ как

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

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

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

где

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

Таким образом, $\widehat{\Omega}_T^*(\overline{\boldsymbol{\beta}}_T^*)$ – бутстраповский аналог $\widehat{\Omega}_T\left(\overline{\boldsymbol{\beta}}_T\right)$.

Бутстраповская дисперсионная матрица задается как

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

где

$$D^*_T(\widehat{\boldsymbol{\beta}}^*_T) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g^{**}_t(\widehat{\boldsymbol{\beta}}^*_T)}{\partial \boldsymbol{\beta}'}.$$

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

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

где $\widehat{\sigma}^{*2}_{ii,T}$ – (i,i)-ый элемент матрицы $\widehat{\Sigma}^*_T$. Но хотя $\widehat{\sigma}^{*2}_{ii,T}$ и есть бутстраповский аналог $\widehat{\sigma}^2_{ii,T}$, эта величина не совпадает с $\mathbb{V}^*[T^{1/2}(\widehat{\boldsymbol{\beta}}^*_{i,T} - \widehat{\boldsymbol{\beta}}_{i,T})]$. 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 $\widehat{\sigma}^{*2}_{ii,T}$ верно имитирует $\widehat{\sigma}^2_{ii,T}$ (т.е. $\mathbb{E}^*\left[\widehat{\sigma}^{*2}_{ii,T}\right] = \widehat{\sigma}^2_{ii,T}$), но $\widehat{\sigma}^{*2}_{ii,T}$ не равно $\mathbb{V}^*[T^{1/2}(\widehat{\boldsymbol{\beta}}^*_{i,T} - \widehat{\boldsymbol{\beta}}_{i,T})]$.

Таким образом, нужен корректирующий множитель. Определим

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

где

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

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

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

Теперь рассмотрим подправленную бутстраповскую статистику

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

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 необходимые добавить к A1 выше.

Предположение A2

A2(i): $\mathbb{E}[g_t\left(\boldsymbol{\beta}^{\dagger}\right)g_{t+j}\left(\boldsymbol{\beta}^{\dagger}\right)'] = 0$ for all $j > \kappa$.

A2(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).

A2(iii): Разложение Эджворта для t-статистики и ее бутстраповского аналога существуют.

A2(iv): Для $f_t(\beta) = \left(g_t\left(\boldsymbol{\beta}\right), g_t\left(\boldsymbol{\beta}\right)g_{t-j}\left(\boldsymbol{\beta}\right)', \partial^i g_t\left(\boldsymbol{\beta}\right)/\partial\boldsymbol{\beta}^i, \partial^i g_t\left(\boldsymbol{\beta}\right)g_{t-j}\left(\boldsymbol{\beta}\right)'/\partial\boldsymbol{\beta}^i,\ j \le \kappa,\ i \ge d_1\right)$, производные вплоть до порядка d_2 имеют конечные моменты и удовлетворяют условию Липшица.

Тогда имеем:

Теорема 2 (из теоремы 2 в Andrews, 2002)

(a) Пусть A1 и A2 выполнены, причем $d_1 \ge 5$ и $d_2 \ge 4$. Пусть $l = T^{\gamma}$, и предположим, что $0 \le \xi \le 1/2 - \gamma$ и $\xi < \gamma$, тогда

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

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

(b) Пусть A1 и A2 выполнены, причем $d_1 \ge 4$ и $d_2 \ge 3$. Пусть $l = T^{\gamma}$, и предположим, что $0 \le \xi \le 1/2 - \gamma$ и $\xi < \gamma$, тогда

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

Доказательство теоремы 2 основано на следующих шагах. Во-первых, $t_{\beta_i,T}$ можно приблизить гладкой функцией, скажем G, от $f_t(\boldsymbol{\beta}^{\dagger})$, определенной в A2(iv), а бутстраповскую статистику без корректирующего множителя $t^*_{\beta_i,T}$ можно приблизить $G(f^*_t(\widehat{\boldsymbol{\beta}}_T))$, где $f^*_t(\widehat{\boldsymbol{\beta}}_T)$ определяется так же, как и $f_t(\widehat{\boldsymbol{\beta}}_T)$, но с заменой выборочных функций моментов на бутстраповские. Тогда, если выполнено A2(iii), для $G\left(f_t(\boldsymbol{\beta}^{\dagger})\right)$ и $G(f^*_t(\widehat{\boldsymbol{\beta}}_T))$ существуют разложения Эджворта, и, если выполнены моментные и липшицевы условия в A2(iv), разница между первыми двумя членами разложений Эджворта для $G\left(f_t(\boldsymbol{\beta}^{\dagger})\right)$ и $G(f^*_t(\widehat{\boldsymbol{\beta}}_T))$ сходится к нулю достаточно быстро. Наконец, если $\xi < \gamma$, корректирующий множитель сходится к единице достаточно быстро, гарантируя, что разложения Эджворта для скорректированной бутстраповской статистики $\widetilde{t}^*_{\beta_i,T}$ 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 $T \to \infty$, корректирующий множитель $\tau_{ii,T} \to 1$. 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 мартингальных приращений, как в случае динамической (верно специфицированной) модели, то $\kappa = 0$, хотя все равно нужно использовать длину блока l, удовлетворяющую $l \to \infty$, 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 $\widetilde{z}^*_{T,\alpha/2}$. Стандартный подход – построить B бутстраповских статистик, скажем $\widetilde{t}^{*(j)}_{\beta_i,T}$, $j = 1,\ldots,B$, и взять $\widetilde{z}^*_{T,B,\alpha/2}$ в качестве $(1-\alpha/2)$-квантили эмпирического распределения $(\widetilde{t}^{*(1)}_{\beta_i,T},\ldots,\widetilde{t}^{*(B)}_{\beta_i,T})$. Проблема в том, как выбрать B достаточно большим, чтобы гарантировать, что инференции, основанные на $\widetilde{z}^*_{T,\alpha/2}$ и $\widetilde{z}^*_{T,B,\alpha/2}$, 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 $\widehat{\beta}^*_T$ может быть трудоемким, так как оно включает решение B нелинейных оптимизационных задач. Davidson & MacKinnon (1999) предложили альтернативную k-шаговую оценку. По сути, нужно положить $\widehat{\beta}^{*(0)}_T = \widehat{\beta}_T$ и совершить k шагов по направлению к $\widehat{\beta}^*_T$ согласно, например, алгоритму Ньютона–Рафсона. Andrews (2002, теорема 1) показал, что инференция, полагающаяся на $z^*_{T,k,\alpha/2}$, т.е. на критические значения, основанные на $\widehat{\beta}^{*(k)}_T$, приводит к того же порядка рафинированию, как и инференция, основанная на $\widehat{\beta}^*_T$, 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 имелась бы и в бутстраповской, и в первоначальной выборках. Другими словами, статистика подсчитывается с удалением πl наблюдений, непосредственно предшествующих точкам соприкосновения $l+1, 2l+1, \ldots, (b-1)l+1$, где, по мере того как $T \to \infty$, $\pi \to 0$ и $\pi l \to \infty$. Поскольку выборка обладает свойством сильного перемешивания, $l(1-\pi)$-е и $(l+1)$-е наблюдения становятся независимыми по мере того как $\pi l \to \infty$. 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 jacknife 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 a overview of some alternative bootstrap procedures which provide improvements over the block bootstrap refinements.