ЭКОНОМЕТРИКА ДЛЯ ПРОДОЛЖАЮЩИХ

Курс лекций

Станислав Анатольев

Российская Экономическая Школа

КЛ/2002/004

Москва

2002–2003

Содержание

Введение

1 Описание курса

Курс служит введением в принципы современного искусства эконометрического оценивания и построения статистических выводов (инференции) как для кросс-данных, так и для временных рядов. Неудовлетворенность точным подходом заставляет нас рассмотреть две альтернативы: асимптотический и бутстраповский подходы. После изучения важных эконометрических тонкостей обоих подходов курс концентрируется на построении оценок в линейных моделях и изучении их свойств. Тем не менее, заключительная часть курса посвящена простым нелинейным моделям и методам. Акцент делается на концептуальную составляющую, нежели на математическую сложность, хотя последняя иногда неизбежна. Домашние задания по курсу содержат как теоретические задачи, так и практические задания, подразумевающие использование пакета GAUSS. Задания служат важным ингредиентом обучающего процесса, в котором часто будут встречаться теоретические и эмпирические примеры.

Выражаю благодарность Семёну Полбенникову за подготовку черновой версии конспектов, Людмиле Солнцевой за техническую помощь, и студентам Российской Экономической Школы за замечания и найденные недочёты. Материал подготовлен в рамках проекта "Совершенствование преподавания социально-экономического образования в ВУЗах", финансируемого Всемирным Банком и реализуемого Национальным Фондом Подготовки Кадров (НФПК).

2 Рекомендуемая литература

1. Anatolyev, Stanislav *Intermediate and Advanced Econometrics: Problems and Solutions*, Lecture Notes series, New Economic School, 2002

2. Hayashi, Fumio *Econometrics*, Princeton University Press, 2000

3. Goldberger, Arthur *A Course in Econometrics*, Harvard University Press, 1991

4. Greene, William *Econometric Analysis*, Prentice Hall, 4th edition, 2000

5. Potcher, Benedikt and Prucha, Ingmar *Basic elements of asymptotic theory*, in: *A Companion to Theoretical Econometrics*, edited by Baltagi, B., Blackwell Publishers, 2001

6. Horowitz, Joel *The bootstrap*, in: *Handbook of Econometrics*, vol. 5, Elsevier Science, North-Holland, 2001

I Приближенный подход к инференции

1 Сравнение точного и приближенного подходов

При эмпирическом анализе данных возникает ситуация, когда эконометрист, имея точечную оценку некоторого параметра, хочет изучить ее статистические свойства. Для этого ему необходимо знать распределение полученной оценки. Знать распределение всегда необходимо для построения доверительных интервалов и тестирования статистических гипотез. Существуют два подхода к вопросу о распределении оцениваемого параметра: *точный* и *приближенный*.

Точный подход основан на предположении об известности вида распределения данных. Остаётся лишь трансформировать его в распределение построенной статистики.

Пример. Пусть условное на X распределение вектора Y будет многомерным нормальным со средним $X\beta$ и дисперсией $\sigma^2 I_n$, т.е.

$$Y|X \sim \mathcal{N}(X\beta, \sigma^2 I_n).$$

Тогда МНК-оценка тоже имеет нормальное условное распределение:

$$\widehat{\beta}_{OLS} = (X'X)^{-1}X'Y|X \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}).$$

Недостатки точного подхода достаточно очевидны. *Во-первых*, чтобы использовать точный подход, необходимо сделать предположение о виде распределения данных. *Во-вторых*, точный подход обычно ограничивается использованием нормального распределения, линейных моделей и простых статистик, ибо в противном случае аналитический вывод распределения интересующей нас статистики обычно становится очень трудоемкой или вообще непосильной задачей.

Приближенный подход основан на аппроксимации истинного распределения исследуемой статистики. В настоящее время существуют два метода в приближенном подходе: *асимптотический* и *бутстраповский*.

Идея *асимптотического метода* в том, чтобы для построения приближенного распределения статистики использовать предельное (при стремлении размера выборки к бесконечности) распределение выборочных средних. Несомненным достоинством такого подхода является тот факт, что используемые предельные распределения обычно являются стандартными и затабулированными. С другой стороны, асимптотическая аппроксимация распределения оценки может быть плохой, и, более того, априори мы не можем знать, насколько она хороша или плоха. Для ответа на этот вопрос приходится использовать дополнительные симуляционные исследования. Кроме того, асимптотический метод в сложных ситуациях может потребовать значительных аналитических выкладок.

Бутстраповский метод в качестве неизвестного истинного распределения данных использует эмпирическое распределение данных, т.е. как данные легли в выборке. В дальнейшем мы подробнее обсудим бутстраповский метод.

В настоящее время эконометристы предпочитают использовать приближенный подход, поскольку точный требует ряда очень сильных предположений о модели и виде распределения данных. Вряд ли разумно считать, что вид этого распределения известен исследователю.

2 Концепции асимптотической теории

Часто используемыми понятиями асимптотической теории являются *состоятельность*, *асимптотическая нормальность* и *асимптотическая эффективность*.

Пусть нас интересуют асимптотические свойства оценки $\widehat{\beta}_n$, полученной из выборки размера n. Поскольку мы предполагаем случайную природу исходных данных, построенная оценка будет случайной величиной. Таким образом, мы имеем последовательность случайных величин – для каждого n своя $\widehat{\beta}_n$. Оценка $\widehat{\beta}_n$ является

- *состоятельной*, если $\widehat{\beta}_n \xrightarrow{p} \beta$, где β – истинное значение оцениваемого параметра,
- *асимптотически нормальной*, если $n^\delta(\widehat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(\mu, \Sigma)$ для какого-то $\delta > 0$ (обычно $\delta = \frac{1}{2}$). Соответственно, n^δ называется *скоростью сходимости*, μ – *асимптотическим смещением*, Σ – *асимптотической дисперсионной матрицей*,
- будучи асимптотически нормальной и асимптотически несмещённой наряду с другой оценкой $\widetilde{\beta}_n$, *асимптотически более эффективной*, чем $\widetilde{\beta}_n$, если при

$$n^\delta(\widehat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$
$$n^\delta(\widetilde{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \widetilde{\Sigma}),$$

  матрица $\widetilde{\Sigma} - \Sigma$ является неотрицательно определенной.

Состоятельность оценки – необходимая норма, говорящая, что чем больше данных, тем ближе наша оценка к истинному параметру. Можно, конечно, представить себе несостоятельную оценку, обладающую желаемыми свойствами в маленьких выборках, но такая ситуация относится к разряду редких исключений.

Асимптотическая нормальность важна по той причине, что проверка статистических гипотез и построение доверительных интервалов требуют знания распределения оценки, а так как точного распределения мы не знаем, пользуемся асимптотическим, нормальным, распределением. Здесь фигурирует именно нормальное распределение, поскольку предельные теоремы, являющиеся сердцем асимптотитеской теории, говорят именно о нормальности эмпирических средних в бесконечных выборках.

Эффективность оценки желательна, поскольку чем более эффективна оценка, тем точнее она предсказывает истинный параметр. Расширяя определение, можно сказать, что дисперсия эффективной оценки мимнимальна среди дисперсий оценок из некоторого класса.

3 О последовательностях случайных величин

При предположении о случайной природе исходных данных построенные оценки являются, как правило, случайными величинами. Для дальнейшего исследования статистических свойств оценок дадим несколько определений и результатов, относящихся к типам сходимости последовательностей случайных величин.

Определение 1 (сходимость почти наверное). Матричная последовательность Z_n случайных величин сходится к случайной матрице Z *почти наверное* (или *с вероятностью единица*), т.е. $Z_n \xrightarrow{as} Z$, если

$$\mathbb{P}\left\{\lim_{n\to\infty} Z_n = Z\right\} = 1,$$

т.е. "почти каждая" траектория сходится к Z.

Определение 2 (сходимость по вероятности). Матричная последовательность Z_n случайных величин сходится к случайной матрице Z *по вероятности*, т.е. $Z_n \xrightarrow{p} Z$ или $p\lim Z_n = Z$, если

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}\left\{\|Z_n - Z\| > \varepsilon\right\} = 0,$$

т.е. вероятность "больших отклонений" от Z стремится к 0.

Определение 3 (сходимость в среднеквадратических отклонениях). Матричная последовательность Z_n случайных величин сходится к случайной матрице Z *в среднеквадратических отклонениях*, т.е. $Z_n \xrightarrow{ms} Z$, если

$$\lim_{n\to\infty} \mathbb{E}\left[\|Z_n - Z\|^2\right] = 0,$$

т.е. среднеквадратическая ошибка стремится к 0.

Определение 4 (сходимость по распределению). Матричная последовательность Z_n случайных величин сходится к случайной матрице Z *по распределению*, т.е. $Z_n \xrightarrow{d} Z$ или $Z_n \xrightarrow{d} D_Z$, где D_Z – распределение Z, если

$$\lim_{n\to\infty} \mathbb{P}\{Z_n \leq z\} = \mathbb{P}\{Z \leq z\}$$

для всех точек непрерывности z распределения D_Z.

Можно показать, что сходимость по вероятности следует из сходимости почти наверное или сходимости в среднеквадратических отклонениях. Сходимость по распределению, в свою очередь, вытекает из сходимости по вероятности.

Результат 1. $\{Z_n \xrightarrow{as} Z$ или $Z_n \xrightarrow{ms} Z\} \Rightarrow Z_n \xrightarrow{p} Z$.

Результат 2. $Z_n \xrightarrow{p} Z \Rightarrow Z_n \xrightarrow{d} Z$.

Результат 3. Если Z – константа, то $Z_n \xrightarrow{p} Z \Leftrightarrow Z_n \xrightarrow{d} Z$

Пример. Рассмотрим последовательность случайных величин

$$\{Z_n\} : \frac{Z}{1}, \frac{Z}{2}, \frac{Z}{3}, \frac{Z}{4}, \ldots, \frac{Z}{n}, \ldots,$$

где Z имеет стандартное нормальное распределение. Тогда $E(Z_n) = 0$ и $Var(Z_n) = \frac{1}{n^2}$. Таким образом $Z_n \xrightarrow{ms} 0$, а, следовательно, и $Z_n \xrightarrow{p} 0$.

4 О последовательностях функций случайных величин

Существует несколько полезных теорем, которые нам понадобятся впоследствии. Здесь они приведены без доказательства.

Теорема (Манна–Вальда). Пусть функция $g : \mathbb{R}^{k_1\times k_2} \to \mathbb{R}^{l_1\times l_2}$ непрерывна, а Z_n – последовательность случайных величин. Тогда

- если $Z_n \xrightarrow{as} Z$, то $g(Z_n) \xrightarrow{as} g(Z)$,
- если $Z_n \xrightarrow{p} Z$, то $g(Z_n) \xrightarrow{p} g(Z)$,
- если $Z_n \xrightarrow{ms} Z$ и g линейна, то $g(Z_n) \xrightarrow{ms} g(Z)$,
- если $Z_n \xrightarrow{d} Z$, то $g(Z_n) \xrightarrow{d} g(Z)$.

Замечание: Если Z константа, то для выполнения теоремы достаточна только локальная непрерывность функции g в точке Z.

Теорема (Слуцкого). Если последовательность случайных величин U_n сходится по вероятности к некоторой константе U, а последовательность случайных величин V_n сходится по распределению к случайной величине V, т.е. $U_n \xrightarrow{p} U$ и $V_n \xrightarrow{d} V$, то

- $U_n + V_n \xrightarrow{d} U + V$,
- $U_n V_n \xrightarrow{d} UV$,
- $U_n^{-1} V_n \xrightarrow{d} U^{-1} V$, если $Pr\{det(U_n) = 0\} = 0$.

Ещё раз обратим внимание на тот факт, что в теореме Слуцкого одна последовательность должна сходиться *по вероятности к константе*. Если это не так, то теорема, вообще говоря, не верна. Следующий пример демонстрирует это.

Пример. Пусть случайная величина Z имеет стандартное нормальное распределение, т.е. $Z \sim \mathcal{N}(0,1)$. Рассмотрим две последовательности случайных величин: $\{Z_n\} = \{Z, Z, Z, Z, \ldots\}$ и $\{X_n\} = \{Z, -Z, Z, -Z, \ldots\}$. Ясно, что $\{Z_n\} \xrightarrow{p} Z$ и $\{X_n\} \xrightarrow{d} Z$. Однако, $\{Z_n + X_n\} = \{2Z, 0, 2Z, 0, 2Z, \ldots\}$. Таким образом, последовательность суммы случайных величин не сходится вообще никуда, т.е. теорема Слуцкого неприменима.

Теорема (Дельта Метод). Пусть последовательность случайных векторов Z размерности $k \times 1$ удовлетворяет условию $\sqrt{n}(Z_n - Z) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, где $Z = p\lim Z_n$ – константа, а функция $g\colon \mathbb{R}^k \to \mathbb{R}^l$ непрерывно дифференцируема в точке Z. Тогда

$$\sqrt{n}(g(Z_n) - g(Z)) \xrightarrow{d} \mathcal{N}(0, G\Sigma G'),$$

где $G = \frac{\partial g(z)}{\partial z'}|_{z=Z}$.

Примеры 1 и 2 демонстрируют применение теоремы Манна–Вальда и Дельта Метода на практике.

Пример 1. Пусть $\overline{x} \xrightarrow{p} \mu$ и $\sqrt{n}(\overline{x} - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. Рассмотрим непрерывно дифференцируемую функцию $g(x) = x'x$. По теореме Манна–Вальда

$$\Sigma^{-1/2}\sqrt{n}(\overline{x} - \mu) \xrightarrow{d} \mathcal{N}(0, I_k),$$

где $(\Sigma^{-1/2})'\Sigma^{-1/2} = \Sigma^{-1}$. Таким образом, получаем следующий результат:

$$\sqrt{n}(\overline{x} - \mu)'\Sigma^{-1}(\overline{x} - \mu) \xrightarrow{d} \chi^2(k).$$

Используя Дельта Метод и учитывая, что $G = \frac{\partial(x'x)}{\partial x'}|_{\mu} = 2x'|_{\mu} = 2\mu'$, получим:

$$\sqrt{n}(\overline{x}'\overline{x} - \mu'\mu) \xrightarrow{d} \mathcal{N}(0, 4\mu'\Sigma\mu).$$

Пример 2. Пусть

$$\sqrt{n}\left[\begin{pmatrix}\overline{x}_1\\ \overline{x}_2\end{pmatrix} - \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\right] \xrightarrow{d} \mathcal{N}(0, I_2).$$

Рассмотрим непрерывно дифференцируемую функцию $g\binom{x_1}{x_2} = \frac{x_1}{x_2}$. По теореме Манна–Вальда,

$$\frac{\overline{x}_1 - \mu_1}{\overline{x}_2 - \mu_2} \xrightarrow{d} \frac{\mathcal{N}(0,1)}{\mathcal{N}(0,1)},$$

т.е. интересующая нас величина имеет распределение Коши. Применяя же Дельта Метод, имеем:

$$G = \left.\frac{\partial\left(\frac{x_1}{x_2}\right)}{\partial(x_1, x_2)}\right|_{\binom{\mu_1}{\mu_2}} = \left(\frac{1}{\mu_2}, -\frac{\mu_1}{\mu_2}\right),$$

так что

$$\sqrt{n}\left[\frac{\overline{x}_1}{\overline{x}_2} - \frac{\mu_1}{\mu_2}\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{1 + \left(\frac{\mu_1}{\mu_2}\right)^2}{\mu_2^2}\right).$$

5 Предельные теоремы для независимых наблюдений

Основными инструментами построения статистических выводов в асимптотическом подходе являются *Законы Больших Чисел* (ЗБЧ) и *Центральные Предельные Теоремы* (ЦПТ). ЗБЧ представляет собой результат о сходимости выборочного среднего к популяционному среднему, ЦПТ дает представление о предельном распределении определенным образом нормированного центрированного выборочного среднего. Существует довольно большое количество формулировок ЗБЧ и ЦПТ. Нас будут интересовать предельные теоремы для двух основных случаев: случай *независимых наблюдений* и случай *стационарных эргодичных временных рядов*. Далее приводятся ЗБЧ и ЦПТ для независимых или серийно нескоррелированных скалярных случайных величин.

Теорема Колмогорова (независимые одинаково распределенные наблюдения). Пусть случайные величины $\{Z_n\}_{i=1}^{\infty}$ независимы и одинаково распределены. Кроме того, пусть существует математическое ожидание $E|Z_i|$. Тогда

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{as} E[Z_i].$$

Теорема Колмогорова (независимые неоднородные наблюдения). Пусть случайные величины $\{Z_n\}_{i=1}^{\infty}$ независимы и имеют конечные дисперсии σ_i^2. Если $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty$, то

$$\frac{1}{n}\sum_{i=1}^{n} Z_i - E\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] \xrightarrow{as} 0.$$

Теорема Чебышева (нескоррелированные наблюдения). Пусть случайные величины $\{Z_n\}_{i=1}^{\infty}$ нескоррелированы, т.е. $Cov(Z_i, Z_j) = 0$ для $i \neq j$. Если $\frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 \underset{n\to\infty}{\to}$

0, то
$$\frac{1}{n}\sum_{i=1}^{n} Z_i - E\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] \xrightarrow{p} 0.$$

Теорема Линдберга-Леви (независимые одинаково распределенные наблюдения). Пусть случайные величины $\{Z_n\}_{i=1}^{\infty}$ независимы и одинаково распределены с математическим ожиданием $E[Z_i] = \mu$ и дисперсией $Var[Z_i] = \sigma^2$. Тогда:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Теорема Ляпунова (независимые неоднородные наблюдения). Пусть случайные величины $\{Z_n\}_{i=1}^{\infty}$ независимы с математическим ожиданием $E[Z_i] = \mu_i$, дисперсией $Var[Z_i] = \sigma_i^2$ и третьим центральным моментом $E[|Z_i - \mu_i|^3] = \nu_i$. Тогда, если

$$\frac{\left(\sum_{i=1}^{n} \nu_i\right)^{1/3}}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}} \underset{n\to\infty}{\to} 0,$$

то

$$\frac{\sum_{i=1}^{n}(Z_i - \mu_i)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

6 Асимптотические статистические выводы

Идея построения статистических выводов при помощи асимптотического метода довольно очевидна. Вместо точного распределения оценки берется асимптотическое, на основании которого строятся распределения тестовых статистик.

Пример. Пусть
$$\sqrt{n}(\overline{Z}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

В данном случае мы имеем дело с выборочным средним $\overline{Z}_n$, которое согласно ЦПТ имеет асимптотически нормальное распределение. Заметим, что в данном случае распределение зависит от неизвестного параметра σ^2, поэтому статистика $\overline{Z}_n$ является *асимптотически непивотальной статистикой*.

Определение. Статистика называется (асимптотически) *пивотальной*, если ее (асимптотическое) распределение не зависит от неизвестных параметров.

Возвращаясь к нашему примеру, мы можем получить пивотальную статистику, построив состоятельную оценку дисперсии $\widehat{\sigma}^2$:

$$\frac{\sqrt{n}(\overline{Z}_n - \mu)}{\widehat{\sigma}} = \frac{\sqrt{n}(\overline{Z}_n - \mu)}{\sigma}\frac{\sigma}{\widehat{\sigma}} \xrightarrow{d} \mathcal{N}(0, 1),$$

так как согласно ЦПТ $\sqrt{n}(\overline{Z}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$, а в силу состоятельности оценки $\widehat{\sigma}^2$, $\sigma/\widehat{\sigma} \xrightarrow{p} 1$. Теперь, зная асимптотическое распределение построенной статистики можно построить доверительный интервал. Так асимптотический доверительный интервал для μ будет

$$\left[\overline{Z}_n - \frac{\widehat{\sigma}}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}(0,1)}, \overline{Z}_n + \frac{\widehat{\sigma}}{\sqrt{n}} q_{1-\frac{\alpha}{2}}^{\mathcal{N}(0,1)}\right].$$

Предположим теперь, что нам нужно протестировать гипотезу $H_0 : \mu = \mu^0$. Согласно построенному нами α-процентному доверительному интервалу гипотеза будет отвергаться, если $\sqrt{n}|\overline{Z}_n - \mu^0|/\widehat{\sigma} > q_{1-\frac{\alpha}{2}}^{\mathcal{N}(0,1)}$. В противном случае гипотеза принимается. Осталось построить состоятельную оценку дисперсии. Оказывается, выборочная дисперсия будет состоятельной оценкой для дисперсии:

$$\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mu)^2 - (\overline{Z}_n - \mu)^2} \xrightarrow{p} \sigma,$$

поскольку из ЗБЧ следует, что $\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mu)^2 \xrightarrow{p} E[(Z_i - \mu)^2] = \sigma^2$ и $(\overline{Z}_n - \mu)^2 \xrightarrow{p} 0$.

7 Асимптотика для стационарных временных рядов

До сих пор мы рассматривали асимптотические свойства оценок в случае независимых наблюдений, и если у нас есть последовательность Z_1, Z_2, Z_3, ..., Z_n, мы могли сказать, что у нас имеется n наблюдений. В случае временных рядов (наблюдений во времени) это, вообще говоря, не так. Каждая траектория $Z_1, Z_2, Z_3, \ldots, Z_T$ представляет собой в общем случае *одно наблюдение*, а из одного наблюдения делать статистические выводы проблематично. Поэтому на природу исходных данных приходится накладывать какую-то структуру. Часто для этого используются предположения о *стационарности и эргодичности* ряда. Грубо говоря, *стационарность* – это "стабильность" распределения Z_t во времени, а *эргодичность* – это "потеря памяти", или "асимптотическая независимость от начальных данных". Дадим более четкие определения:

Определение. Временной ряд называется *строго стационарным*, если совместное распределение $Z_t, Z_{t-1}, \ldots, Z_{t-k}$ не зависит от t для любых k.

Поскольку точное определение эргодичности использует понятия теории меры, дадим интуитивное определение:

"Определение". Временной ряд Z_t называется *эргодичным*, если Z_t и Z_{t+k} асимптотически независимы при $k \to \infty$.

Приведем примеры различных стационарных или нестационарных и эргодичных или неэргодичных временных рядов.

Пример 1 (стационарные эргодичные ряды).

- $Z_t \sim iid$, независимые одинаково распределенные наблюдения,
- $\varepsilon_t \sim iid(0, \sigma^2)$, сильный "белый шум",
- $AR(1) : z_t = \rho z_{t-1} + \varepsilon_t$, $|\rho| < 1$,
- $MA(1) : z_t = \varepsilon_t + \theta\varepsilon_{t-1}$.

Пример 2 (нестационарные неэргодичные ряды).

- $z_t = z_{t-1} + \varepsilon_t$, "случайное блуждание". Здесь дисперсия наблюдений растет со временем: $Var(z_t) = Var(z_{t-1}) + \sigma_\varepsilon^2$, т.е. ряд не может быть стационарен. Кроме того, начальные данные "не забываются" со временем: $z_t = z_0 + \sum_{i=1}^{t} \varepsilon_i$, и ряд неэргодичен.

Пример 3 (стационарные неэргодичные ряды).

- Пусть $z \sim \mathcal{N}(0, 1)$ и $z_t = z + \varepsilon_t$, где ε_t и z независимы. Очевидно, что ряд z_t стационарен, но неэргодичен.

Пример 4 (нестационарные эргодичные ряды).

- Сезонный ряд: $z_t = s(\tau, t) + \varepsilon_t$, где $s(\tau, t) = s(\tau, t + \tau)$.

Результат. Если случайный процесс z_t является стационарным и эргодичным, и если $Y_t = f(z_t, z_{t-1} \ldots)$ есть случайная величина, то Y_t также является стационарным и эргодичным рядом.

Определение. *Информацией* в момент времени t называются все реализовавшиеся значения z_k вплоть до z_t, т.е. $I_t = \{z_t, z_{t-1}, \ldots\}$.

Определение. Ряд z_t называется *последовательностью мартингальных приращений* (ПМП) по отношению к своему прошлому, если $E[z_t|I_{t-1}] = 0$.

Сформулируем ЗБЧ и ЦПТ для временных рядов.

Теорема Биркоффа–Хинчина (зависимые наблюдения). Пусть ряд $\{Z_t\}_{t=-\infty}^{+\infty}$ стационарен и эргодичен. Кроме того, пусть $E[|Z_t|] < \infty$, тогда

$$\frac{1}{T}\sum_{t=1}^{T} Z_t \xrightarrow{as} E[Z_t]$$

при $T \to \infty$.

Теорема Биллингслея (последовательность мартингальных приращений). Пусть ряд $\{Z_t\}_{t=-\infty}^{+\infty}$ стационарен, эргодичен и является ПМП по отношению к своему прошлому. Кроме того, пусть $\sigma^2 = E[Z_t^2] < \infty$, тогда

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

при $T \to \infty$.

"Теорема" (зависимые наблюдения). Пусть ряд $\{Z_t\}_{t=-\infty}^{+\infty}$ стационарен и эргодичен. Кроме того, пусть

$$\sigma^2 = \sum_{j=-\infty}^{+\infty} Cov[Z_t, Z_{t-j}] < \infty.$$

Тогда при определенных условиях,

$$\sqrt{T} \left(\frac{1}{T} \sum_{t=1}^{T} Z_t - E[Z_t] \right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

при $T \to \infty$.

Приведем примеры использования изложеных выше теорем для исследования асимптотических свойств оценок на временных рядах.

Пример. Рассмотрим авторегрессионный процесс первого порядка $AR(1)$:

$$x_t = \rho x_{t-1} + \varepsilon_t, \ |\rho| < 1, \ \varepsilon_t \sim iid(0, \sigma^2).$$

Нас интересуют асимптотические свойства МНК-оценки

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} x_{t-1} x_t}{\sum_{t=2}^{T} x_{t-1}^2} = \rho + \frac{\sum_{t=2}^{T} x_{t-1} \varepsilon_t}{\sum_{t=2}^{T} x_{t-1}^2}.$$

По теореме Биркоффа–Хинчина,

$$\frac{1}{T-1} \sum_{t=2}^{T} x_{t-1} \varepsilon_t \xrightarrow{p} E[x_{t-1} \varepsilon_t] = 0,$$

$$\frac{1}{T-1} \sum_{t=2}^{T} x_{t-1}^2 \xrightarrow{p} E[x_{t-1}^2].$$

Следовательно, по теореме Слуцкого оценка $\widehat{\rho}$ является состоятельной оценкой, т.е. $\widehat{\rho} \xrightarrow{p} \rho$.

Теперь найдем асимптотическое распределение МНК-оценки:

$$\sqrt{T} \, (\widehat{\rho} - \rho) = \frac{\frac{1}{\sqrt{T-1}} \sum_{t=2}^{T} x_{t-1} \varepsilon_t}{\frac{1}{T-1} \sum_{t-2}^{T} x_{t-1}^2} \sqrt{\frac{T}{T-1}}.$$

Очевидно, что $\sqrt{\frac{T}{T-1}} \underset{n\to\infty}{\to} 1$, а $\frac{1}{T-1}\sum_{t-2}^{T} x_{t-1}^2 \xrightarrow{p} E[x_{t-1}^2]$. Покажем, что последовательность $x_{t-1}\varepsilon_t$ является последовательностью мартингальных приращений по отношению к своему прошлому, т.е. информационному множеству $I_{t-1} = \{x_{t-2}\varepsilon_{t-1}, x_{t-3}\varepsilon_{t-2}\ldots\}$:

$$E[x_{t-1}\varepsilon_t|I_{t-1}] = E[E[x_{t-1}\varepsilon_t|x_{t-1}, x_{t-2}\varepsilon_{t-1}\ldots]|I_{t-1}] = 0,$$

т.е. последовательность $x_{t-1}\varepsilon_t$ является ПМП. Таким образом, мы можем применить ЦПТ Биллингслея:

$$\frac{1}{\sqrt{T-1}}\sum_{t=2}^{T} x_{t-1}\varepsilon_t \xrightarrow{d} \mathcal{N}(0, E[x_{t-1}^2\varepsilon_t^2]).$$

Заметив, что $E[x_t^2] = Var[x_t] = \rho^2 Var[x_{t-1}] + \sigma^2 = \frac{\sigma^2}{1-\rho^2}$, получим окончательный результат:

$$\sqrt{T}\,(\widehat{\rho}-\rho) \xrightarrow{d} \mathcal{N}(0, 1-\rho^2).$$

Соответствующая пивотальная статистика будет

$$\frac{\sqrt{T}\,(\widehat{\rho}-\rho)}{\sqrt{1-\widehat{\rho}^2}} \xrightarrow{d} \mathcal{N}(0,1).$$

В результате, 95%-ный доверительный интервал для ρ есть

$$CI_\rho = \left[\widehat{\rho} - 1.96\sqrt{\frac{1-\widehat{\rho}^2}{T}};\ \widehat{\rho} + 1.96\sqrt{\frac{1-\widehat{\rho}^2}{T}}\right].$$

Обратимся еще раз к ЦПТ для зависимых наблюдений. Вид вариационной матрицы в асимптотическом распределении оценки требует некоторого пояснения. Когда мы имеем дело с последовательностью мартингальных приращений Z_t, у нас $E[Z_t Z_{t-j}] = 0$ для $j > 0$, поэтому асимптотическая дисперсия для ПМП имеет простой вид: $\sigma^2 = E[Z_t^2]$. Однако, всё сложнее для более зависимых наблюдений:

$$\begin{aligned}
& Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\right] = \frac{1}{T} Var\left[\sum_{t=1}^{T} Z_t\right] = \\
= \;& \frac{1}{T}[T Var(Z_t) + (T-1)Cov(Z_t, Z_{t+1}) + (T-1)Cov(Z_t, Z_{t-1}) + \\
+ \;& (T-2)Cov(Z_t, Z_{t+2}) + (T-2)Cov(Z_t, Z_{t-2}) + \ldots + \\
+ \;& Cov(Z_1, Z_T) + Cov(Z_T, Z_1)] \underset{T\to\infty}{\to} \sum_{j=-\infty}^{+\infty} Cov(Z_t, Z_{t-j}).
\end{aligned}$$

Рассмотрим пример с зависимыми наблюдениями, когда асимптотическую дисперсионную матрицу приходится считать по указанной выше формуле. Ясно, что в этом случае ошибки должны быть скоррелироваными.

Пример. Рассмотрим процесс скользящего среднего первого порядка $MA(1)$:

$$z_t = \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim iid(0, \sigma^2).$$

Заметим, что

$$Var(z_t) = (1+\theta^2)\sigma^2,\ Cov(z_t, z_{t-1}) = \theta\sigma^2,\ Cov(z_t, z_{t-j}) = 0,\ j > 1.$$

В этом случае,

$$\sum_{j=-\infty}^{+\infty} Cov(z_t, z_{t-j}) = (1+\theta^2)\sigma^2 + 2\theta\sigma^2 = (1+\theta)^2\sigma^2.$$

Тогда, согласно ЦПТ для зависимых наблюдений,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t \xrightarrow{d} \mathcal{N}(0, (1+\theta)^2\sigma^2).$$

Обратим внимание на то, что в этом случае z_t не является ПМП относительно $I_t = \{z_{t-1}, z_{t-2}, z_{t-3}\ldots\}$, т.к. $E[z_t|z_{t-1}, z_{t-2}, \ldots] = \theta\varepsilon_{t-1} \neq 0$.

Для получения пивотальной статистики возникает необходимость состоятельного оценивания асимптотической дисперсионной матрицы. Вид искомой оценки может быть

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(Z_t - \overline{Z})(Z_t - \overline{Z})' + \sum_{j=1}^{T-1}\frac{1}{T}\sum_{t=j+1}^{T}\{(Z_t - \overline{Z})(Z_{t-j} - \overline{Z})' + (Z_t - \overline{Z})(Z_{t+j} - \overline{Z})'\}.$$

Увы, такая оценка не будет состоятельной, т.е. $\widehat{\Omega} \stackrel{p}{\nrightarrow} \Omega$. Дело в том, что из-за конечности выборки невозможно состоятельно оценить крайние члены ряда. Таким образом, используя эргодичность, необходимо "обрезать" ряд на слагаемом номер $m << T$, таком, чтобы при $T \to \infty$ мы имели $m \to \infty$ и $m/T \to 0$. Ньюи и Уэст предложили состоятельную оценку вариационной матрицы, которая по построению является *положительно определенной*:

$$\widehat{\Omega}^{NW} = \sum_{j=-m}^{m}\left(1 - \frac{|j|}{m+1}\right)\frac{1}{T}\sum_{t=\max(1,1+j)}^{\min(T,T+j)}(Z_t - \overline{Z})(Z_{t-j} - \overline{Z})'.$$

Для практического использования была предложена следующая формула выбора m:

$$m = \left[4\left(\frac{T}{100}\right)^{1/3}\right].$$

Такой выбор m дает хорошие результаты в смысле точности оценок, за исключением тех случаев, когда затухание возмущений в процессе происходит медленно, т.е. корни соответствующих полиномов лежат близко к единичному кругу.

Вернемся к уже рассмотренному примеру MA(1): $z_t = \varepsilon_t + \theta\varepsilon_{t-1}$. Вот результат, который мы получили:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t \xrightarrow{d} \mathcal{N}(0,(1+\theta)^2\sigma^2).$$

Допустим теперь, что мы хотим получить состоятельную оценку для асимптотической дисперсии. На практике у нас есть три возможных способа:

- Мы можем построить состоятельные оценки $\widehat{\theta} \xrightarrow{p} \theta$ и $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, а затем сконструировать состоятельную оценку асимптотической дисперсии: $\widehat{\sigma}_z^2 = (1+\widehat{\theta})^2\widehat{\sigma}^2$.

- Зная, что искомая дисперсия выражается как $\sigma_z^2 = Var(z_t) + 2Cov(z_t, z_{t-1})$, мы можем сконструировать состоятельную оценку в виде $\widehat{\sigma}_z^2 = \widehat{Var(z_t)} + 2\widehat{Cov(z_t, z_{t-1})}$, где

$$\widehat{Var(z_t)} = \frac{1}{T}\sum_{t=1}^{T} z_t^2,\ \widehat{Cov(z_t, z_{t-1})} = \frac{1}{T}\sum_{t=2}^{T} z_t z_{t-1}.$$

- Мы можем использовать приведенную выше оценку Ньюи–Уэста:

$$\widehat{\sigma}_z^2 = \sum_{j=-m}^{m}\left(1-\frac{|j|}{m+1}\right)\frac{1}{T}\sum_{t=\max(1,1+j)}^{\min(T,T+j)} z_t z_{t-j}.$$

8 Введение в асимптотику для нестационарных процессов

Если временной ряд не стационарен, а имеет стохастические тренды, построение статистических выводов значительно усложняется. Здесь мы рассмотрим простейший пример важного класса нестационарных процессов. Пусть X_t описывается уравнением случайного блуждания, т.е.:

$$X_t = X_{t-1} + \varepsilon_t,\ X_0 = 0,\ \varepsilon_t \sim iid(0,\sigma^2).$$

Тогда выборочное среднее выражается следующим образом:

$$\frac{1}{T}\sum_{t=1}^{T} X_t = \frac{\varepsilon_T}{T} + \frac{2}{T}\varepsilon_{T-1} + \cdots + \frac{T-t+1}{T}\varepsilon_t + \cdots + \varepsilon_1.$$

Следовательно,

$$Var\left(\frac{1}{T}\sum_{t=1}^{T} X_t\right) = \sigma^2\left(1 + \left(\frac{T-1}{T}\right)^2 + \cdots + \left(\frac{2}{T}\right)^2 + \left(\frac{1}{T}\right)^2\right),$$

т.е.

$$Var\left(\frac{1}{T}\sum_{t=1}^{T} X_t\right) = \sigma^2\frac{(T+1)(2T+1)}{6T} = O(T).$$

Рассуждая аналогично, в результате мы получим:

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} X_t \xrightarrow{p} V_1,\ \frac{1}{T}\sum_{t=1}^{T} X_{t-1}\varepsilon_t \xrightarrow{p} V_2,\ \frac{1}{T^2}\sum_{t=1}^{T} X_{t-1}^2 \xrightarrow{p} V_3,$$

где $V_1,\ V_2, V_3$ – некоторые случайные величины.

Если мы теперь используем МНК-оценку для ρ, которое равно единице, то асимптотические свойства этой оценки будут следующие:

$$T(\widehat{\rho}-1) \xrightarrow{p} \frac{V_2}{V_3}.$$

Во-первых, МНК-оценка в данном случае *суперсостоятельна*, т.е. скорость сходимости к асимптотическому распределению превышает $\sqrt{T}$. Во-вторых, асимптотическое распределение нестандартно: оно не является нормальным, зато обладает ненулевыми смещением и скошенностью. Оно носит название *распределения Дики–Фуллера*.

II Бутстраповский подход

1 Приближение истинного распределения бутстраповским

В основе бутстраповского подхода лежит идея, что истинное распределение данных можно хорощо приблизить эмпирическим. Таким образом может быть получено приближенное распределение интересующей нас статистики. Пусть из исходной популяции с распределением $F(x)$ была получена выборка размера n. Тогда эмпирическая функция распределения $F_n(x) = \frac{1}{n}\sum_{i=1}^{T} I(X_i \leqslant x)$ равномерно почти наверное стремится к $F(x)$ при $n \to \infty$. Это свойство мотивирует использование бутстрапа.

Чтобы более наглядно пояснить бутстраповский метод, рассмотрим простейший пример. Пусть у нас есть всего два наблюдения:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},\ \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Допустим, нас интересует коэффициент регрессии y на x, т.е. $y_i = \theta x_i + \varepsilon_i$. В этом случае МНК-оценка θ равна

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

Эмпирическая функция распределения данных есть

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

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

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

Это распределение и является бутстраповским. Соответственно, МНК-оценка распределена согласно ее бутстраповскому распределению

$$\widehat{\theta}_2^* = \begin{cases} 1/2 & \text{с вероятностью } 1/4 \\ 4/5 & \text{с вероятностью } 1/2 \\ 2 & \text{с вероятностью } 1/4 \end{cases}$$

Используя это бутстраповское распределение, можно строить доверительные интервалы или тестировать гипотезы обычным образом.

Пример, рассмотренный нами, был чрезвычайно прост: размер исходной выборки был равен 2. В общем случае, когда мы имеем n наблюдений, количество вариантов для бутстраповских статистик имеет порядок n^n. Таким образом, в вычислительном плане задача сильно усложняется по мере роста n.

2 Приближение с помощью симуляций

Как упоминалось, при увеличении размера выборки объем вычислений для получения бутстраповского распределения быстро возрастает. Поэтому, как правило, процедура бутстрапирования осуществляется с помощью симуляций. Здесь мы приведем описательный алгоритм построения бутстраповских доверительных интервалов.

Бутстраповский алгоритм.

1. Выбрать количество псевдовыборок B (обычно хватает 1000). Для $b = 1, 2, \ldots, B$ построить псевдовыборки $(z_1^*; z_2^*; \ldots; z_n^*)_b$, вытягивая элементы псевдовыборок случайным образом *с возвращением* из исходной выборки $(z_1; \ldots; z_n)$. Для каждой псевдовыборки вычислить псевдостатистику $\widehat{\theta}_b^* = \widehat{\theta}((z_1^*; \ldots; z_n^*)_b)$.

2. Полученные псевдостатистики $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*$ отсортировать в порядке возрастания. В качестве квантилей $q_{\alpha_1}^*$, $q_{1-\alpha_2}^*$ взять значения $\widehat{\theta}_{[B\alpha_1]}^*$, $\widehat{\theta}_{[B(1-\alpha_2)+1]}^*$, на основе которых построить доверительный интервал.

3 Какие статистики бутстрапить?

Ответ на вопрос, какие статистики лучше использовать при построении доверительных интервалов с помощью бутстрапа, кроется в двух простых соображениях. Во-первых, бутстраповское распределение центрировано не около истинного значения статистики, а около его выборочного аналога. Во-вторых, полагается бутстрапировть асимптотически пивотальные статистики.

Рассмотрим несколько вариантов бутстраповских статистик, используемых для построения доверительных интервалов и подчеркнем их положительные и отрицательные качества. Пусть нас интересует построение статистических выводов относительно параметра β из ее оценки $\widehat{\beta}$.

- **Эфроновский доверительный интервал.** В данном случае бутстрапируемой статистикой является сама оценка, т.е. $\widehat{\theta} = \widehat{\beta}$. Таким образом, мы получаем бутстраповское распределение $\{\widehat{\theta}_b^* = \widehat{\beta}_b^*\}_{b=1}^B$. Соответствующие квантили распределения – $q^*_{\alpha/2}$, $q^*_{1-\alpha/2}$, а доверительный интервал –

$$CI_E = [q^*_{\alpha/2},\ q^*_{1-\alpha/2}].$$

  Эфроновский доверительный интервал был популярен, когда бутстраповский подход только начинал использоваться. На самом деле, этот доверительный интервал дает плохую аппроксимацию для истинных уровней значимости, поскольку сохраняет смещение исходной выборки.

- **Холловский доверительный интервал.** Холл предложил использовать для построения доверительного интервала рецентрированную статистику $\widehat{\theta} = \widehat{\beta} - \beta$, что снимает проблему смещения, связанного с конечностью выборки. Таким образом, получается бутстраповское распределение $\{\widehat{\theta}_b^* = \widehat{\beta}_b^* - \widehat{\beta}\}_{b=1}^B$. Соответствующие квантили – $q^*_{\alpha/2}$, $q^*_{1-\alpha/2}$, а доверительный интервал –

$$CI_H = [\widehat{\beta} - q^*_{1-\alpha/2},\ \widehat{\beta} - q^*_{\alpha/2}].$$

  Холловский доверительный интервал дает лучшую, чем Эфроновский, аппроксимацию уровней значимости. Плюсом использования Холловского доверительного интервала является отсутствие необходимости оценивания стандартных ошибок, хотя, как увидим впоследствии, такая стратегия оборачивается серьезным минусом.

- **t-процентный доверительный интервал.** Такой интервал использует в качестве бутстрапируемой статистики t-статистику, т.е. $\dfrac{\widehat{\beta} - \beta}{se(\widehat{\beta})}$. Таким образом, находят бутстраповское распределение статистики $\left\{\dfrac{\widehat{\beta}_b^* - \widehat{\beta}}{se(\widehat{\beta}_b^*)}\right\}_{b=1}^{B}$ и соответствующие квантили $q^*_{\alpha/2},\ q^*_{1-\alpha/2}$, а сам t-процентный доверительный интервал строят как

$$CI_t = [\widehat{\beta} - se(\widehat{\beta})q^*_{1-\alpha/2},\ \widehat{\beta} - se(\widehat{\beta})q^*_{\alpha/2}].$$

t-процентный доверительный интервал еще лучше аппроксимирует истинные уровни значимости, чем Холловский доверительный интервал. Но использовать его рекомендуется только если стандартные ошибки можно построить качественно.

- **Симметричный t-процентный доверительный интервал.** Такой интервал использует в качестве бутстрапируемой "симметризованную t-статистику" $\dfrac{|\widehat{\beta} - \beta|}{se(\widehat{\beta})}$. Распределение бутстраповской статистики есть $\left\{\dfrac{|\widehat{\beta}_b^* - \widehat{\beta}|}{se(\widehat{\beta}_b^*)}\right\}_{b=1}^{B}$, а правый квантиль – q^*_α. Симметричный t-процентный доверительный интервал есть

$$CI_{|t|} = [\widehat{\beta} - se(\widehat{\beta})q^*_{1-\alpha},\ \widehat{\beta} + se(\widehat{\beta})q^*_{1-\alpha}].$$

Симметричный t-процентный доверительный интервал имеет в определенных случаях преимущество перед t-процентным доверительным интервалом. А именно, если асимптотическое распределение статистики $\widehat{\beta} - \beta$ симметрично (как раз как в случае асимптотической нормальности), то $CI_{|t|}$ дает лучшую аппроксимацию уровней значимости.

4 Корректировка смещения

Бутстрап позволяет скорректировать смещение, связанное с конечностью выборки. Пусть у нас есть смещённая, но состоятельная статистика $\widehat{\beta}$:

$$E[\widehat{\beta}] \neq \beta.$$

Тогда мы можем выразить смещение следующим образом:

$$Bias = E[\widehat{\beta}] - \beta.$$

Если у нас есть возможность качественно оценить смещение, то мы можем скорректировать исходную статистику:

$$\widetilde{\beta} = \widehat{\beta} - \widehat{Bias}.$$

Смещение же можно оценить с помощью бутстрапа:

$$Bias^* = E^*[\widehat{\beta}_b^*] - \widehat{\beta} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\beta}_b^* - \widehat{\beta}.$$

Таким образом, скорректированная статистика есть

$$\widetilde{\beta} = \widehat{\beta} - \left(\frac{1}{B}\sum_{b=1}^{B} \widehat{\beta}_b^* - \widehat{\beta}\right) = 2\widehat{\beta} - \overline{\widehat{\beta}^*}.$$

5 Тестирование гипотез при помощи бутстрапа

Одной из основных целей бутстрапа является тестирование гипотез. Рассмотрим, как с помощью бутстрапа тестируются простейшие статистические гипотезы. Пусть нулевая гипотеза имеет вид $H_0 : \beta = \beta^0$, где β – скаляр.

- **Альтернативная гипотеза односторонняя:** $H_a : \beta > \beta^0$. Бутстрапим t-процентную статистику

  $$\widehat{\theta} = \frac{\widehat{\beta} - \beta}{se(\widehat{\beta})}$$

  и получаем бутстраповское распределение этой статистики и соответствующий квантиль:

  $$\left\{\widehat{\theta}_b^* = \frac{\widehat{\beta}_b^* - \widehat{\beta}}{se(\widehat{\beta}_b^*)}\right\}_{b=1}^{B} \Rightarrow q_{1-\alpha}^*.$$

  Гипотеза H_0 отвергается, если $\dfrac{\widehat{\beta} - \beta^0}{se(\widehat{\beta})} > q_{1-\alpha}^*$.

- **Альтернативная гипотеза двусторонняя:** $H_a : \beta \neq \beta^0$. В этом случае мы бутстрапим симметричную t-процентную статистику

  $$\widehat{\theta} = \frac{|\widehat{\beta} - \beta|}{se(\widehat{\beta})}.$$

  Получаем бутстраповское распределение и квантиль:

  $$\left\{\widehat{\theta}_b^* = \frac{|\widehat{\beta}_b^* - \widehat{\beta}|}{se(\widehat{\beta}_b^*)}\right\}_{b=1}^{B} \Rightarrow q_{1-\alpha}^*.$$

  Гипотеза H_0 отвергается, если $\dfrac{|\widehat{\beta} - \beta^0|}{se(\widehat{\beta})} > q_{1-\alpha}^*$.

Пусть нулевая гипотеза имеет вид $H_0 : \beta = \beta^0$, где β – вектор. В этом случае мы бутстрапим Вальдовскую статистику (с точностью до коэффициента пропорциональности)

$$\widehat{\theta} = (\widehat{\beta} - \beta)' \widehat{V}_{\beta}^{-1} (\widehat{\beta} - \beta).$$

Соответственно, получаем бутстраповское распределение и квантиль:

$$\left\{ \widehat{\theta}_b^* = (\widehat{\beta}_b^* - \widehat{\beta})' \widehat{V}_{\beta}^{*-1} (\widehat{\beta}_b^* - \widehat{\beta}) \right\}_{b=1}^{B} \Rightarrow q_{1-\alpha}^*.$$

Гипотеза H_0 отвергается, если $\widehat{\theta}^0 = (\widehat{\beta} - \beta^0)' \widehat{V}_{\beta}^{-1} (\widehat{\beta} - \beta^0) > q_{1-\alpha}^*$.

Пусть теперь нулевая гипотеза имеет вид линейных ограничений на коэффициенты $H_0 : R\beta = r$, где R – матрица ограничений. В этом случае мы снова бутстрапим Вальдовскую статистику (с точностью до коэффициента пропорциональности)

$$\widehat{\theta} = (R\widehat{\beta} - r)' (R \widehat{V}_{\beta} R')^{-1} (R\widehat{\beta} - r).$$

Получаем бутстраповское распределение, из которого находим соответствующий квантиль:

$$\left\{ \widehat{\theta}_b^* = (\widehat{\beta}_b^* - \widehat{\beta})' R' (R \widehat{V}_{\beta}^* R')^{-1} R (\widehat{\beta}_b^* - \widehat{\beta}) \right\}_{b=1}^{B} \Rightarrow q_{1-\alpha}^*.$$

Заметим, что мы рецентрируем бутстраповскую статистику. Без этого бутстраповское распределение унаследовало бы смещение, свойственное первоначальной статистике. Гипотеза H_0 отвергается, если $\widehat{\theta} = (R\widehat{\beta} - r)' (R \widehat{V}_{\beta} R')^{-1} (R\widehat{\beta} - r) > q_{1-\alpha}^*$.

6 Асимптотическое рафинирование

Иногда говорят, что с помощью бутстрапа достигается асимптотическое рафинирование. В этой главе мы обсудим, что такое асимптотическое рафинирование и в каких случаях оно имеет место.

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

Приведем примеры, использующие разложение Эджворта функции распределения статистики вокруг предельного распределения.

Пример 1: асимптотически пивотальная t-статистика. Пусть бутстрапируемая нами статистика есть

$$\widehat{\theta} = \frac{\widehat{\beta} - \beta}{se(\widehat{\beta})}.$$

Ее асимптотическое распределение, как мы уже видели, является стандартным нормальным: $\widehat{\theta} \xrightarrow{d} \mathcal{N}(0,1)$ (т.е. статистика асимптотически пивотальная). Обозначим точное распределение статистики через $F_{\widehat{\theta}}(x)$, а бутстраповское – через $F^*_{\widehat{\theta}}(x)$. Для кумулятивной функции стандартного нормального распределения используем обычное обозначение $\Phi(x)$.

Итак, разложим истинное и бутстраповское распределения вокруг асимптотического:

$$F_{\widehat{\theta}}(x) = \Phi(x) + \frac{h_1(x,F)}{\sqrt{n}} + \frac{h_2(x,F)}{n} + O\left(\frac{1}{n^{3/2}}\right),$$

$$F^*_{\widehat{\theta}}(x) = \Phi(x) + \frac{h_1(x,\widehat{F})}{\sqrt{n}} + \frac{h_2(x,\widehat{F})}{n} + O\left(\frac{1}{n^{3/2}}\right).$$

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

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

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

Здесь мы воспользовались тем фактом, что разность $h_1(x,\widehat{F}) - h_1(x,F)$ имеет асимптотику $\frac{1}{\sqrt{n}}$, поскольку

$$\begin{aligned}\sqrt{n}\left(\widehat{F}(x) - F(x)\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} 1[x_i \leq x] - E\left[1[x_i \leq x]\right]\right) \\ &\xrightarrow{d} \mathcal{N}(0, P\{x_i \leq x\}P\{x_i > x\}).\end{aligned}$$

Таким образом, в данном примере использование бутстрапа приводит к асимптотическому рафинированию.

Пример 2: асимптотически непивотальная статистика. Рассмотрим статистику

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

Сохранив обозначение кумулятивных функций распределения для точного распределения и бутстраповского из предыдущего примера, обозначим асимптотическое распределение через $\Phi(x, V_\beta)$. Заметим, что теперь наша статистика асимптотически непивотальная, т.е. ее асимптотическое распределение зависит от неизвестного параметра, в данном случае V_β. Как в предыдущем примере, разложим точное и бутстраповское распределения вокруг асимптотического:

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

$$F^*_{\widehat{\theta}}(x) = \Phi(x, V^*_{\beta}) + \frac{h_1(x, \widehat{F})}{\sqrt{n}} + O\left(\frac{1}{n}\right).$$

Ошибки аппроксимации для асимптотического и бутстраповского распределений считаются аналогично предыдущему примеру:

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

$$F^*_{\widehat{\theta}}(x) - F_{\widehat{\theta}}(x) = \Phi(x, V^*_{\widehat{\beta}}) - \Phi(x, V_{\widehat{\beta}}) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

Как видно, в данном случае использование бутстрапа не приводит к асимптотическому рафинированию. Вообще, как правило, бутстрапирование *асимптотически непивотальных* статистик не дает асимптотического рафинирования.

Пример 3: асимптотически пивотальная симметричная t-статистика. Теперь рассмотрим в качестве примера симметричную t-статистику

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

Сохраняя обозначения предыдущих примеров, разложим точное и бутстраповское распределения:

$$F_{\widehat{\theta}}(x) = Pr\{-x \leq \frac{\widehat{\beta} - \beta}{se(\widehat{\beta})} \leq x\} = 2\Phi(x) - 1 + \frac{2h_2(x, F)}{n} + O\left(\frac{1}{n^{3/2}}\right),$$

$$F^*_{\widehat{\theta}}(x) = 2\Phi(x) - 1 + \frac{2h_2(x, \widehat{F})}{n} + O\left(\frac{1}{n^{3/2}}\right).$$

Таким образом, ошибки аппроксимации для асимптотики и бутстрапа имеют порядки

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

$$F^*_{\widehat{\theta}}(x) - F_{\widehat{\theta}}(x) = \frac{2}{n}\left(h_2(x, \widehat{F}) - h_2(x, F)\right) + O\left(\frac{1}{n^{3/2}}\right) = O\left(\frac{1}{n^{3/2}}\right).$$

Таким образом, мы получаем асимптотическое рафинирование. Заметим, что бутстрапирование симметричного двустороннего теста имеет ошибку более высокого порядка, чем бутстрапирование одностороннего теста.

7 Построение псевдовыборок при бутстрапе кросс-секций

Рассмотрим случай простейшей множественной регрессии с независимыми наблюдениями:

$$y = x'\beta + e, \;\; E[e|x] = 0, \;\; \{(x_i, y_i)\} \sim iid.$$

Существует несколько альтернативных способов построения псевдовыборки для этой регрессии:

1. **Непараметрическое построение псевдовыборки.** Из исходных наблюдений $\{(x_i, y_i)\}_{i=1}^n$ случайно с возвращением извлекаются n наблюдений (x_i^*, y_i^*).

2. **Построение псевдовыборки по остаткам.** Сначала оценивается модель и находится состоятельная оценка $\widehat{\beta}$. Затем вычисляются остатки: $\widehat{e}_i = y_i - x_i'\widehat{\beta}$. Из множества пар $\{(x_i, \widehat{e}_i)\}_{i=1}^n$ случайным образом с возвращением извлекаются n наблюдений $(x_i^*, \widehat{e}_i^*)$. Затем восстанавливается переменная левой части $y_i^* = x_i'^*\widehat{\beta} + \widehat{e}_i^*$. Заметим, что в данном контексте этот метод построения псевдовыборки идентичен непараметрическому методу, но идентичность пропадает в более сложных случаях.

3. **Построение псевдовыборки по остаткам (специальный случай).** Если исследователю заранее известно, что ошибки и регрессоры независимы, то эффективность бутстрапа можно увеличить, извлекая случайно с возвращением x_i^* из $\{x_i\}_{i=1}^n$ и $\widehat{e}_i^*$ из $\{\widehat{e}_i\}_{i=1}^n$ по отдельности.

4. **Построение псевдовыборки по остаткам (специальный случай).** Если исследователь знает, что ошибки и регрессоры независимы, и, кроме того, ошибки распределены нормально, т.е. $e_i \sim \mathcal{N}(0, \sigma^2)$, то эффективность бутстрапа можно увеличить (по сравнению с предыдущим случаем), выбирая регрессоры и остатки для псевдовыборки по отдельности, и, кроме того, остатки стоит извлекать из нормального распределения, т.е. x_i^* извлекаются случайно с возвращением из $\mathcal{N}(0, \widehat{\sigma}^2)$.

8 Построение псевдовыборок при бутстрапе временных рядов

Временной ряд отличается от кросс-секционной выборки тем, что наблюдения здесь зависимы, поэтому случайное перемешивание при непараметрическом бутстрапе разрушает эту зависимость, так что вероятностная структура псевдоданных уже не имитирует вероятностную структуру данных. Чтобы избежать этого, используется блочный бутстрап, в котором псевдовыборка строится из блоков исходной выборки. Аналогично случаю независимых наблюдений, во временых рядах возможно построение псевдовыборки по остаткам, однако такой способ применим только в тех редких случаях, когда ошибки или инновации серийно независимы. Рассмотрим несколько альтернативных способов построения блочной псевдовыборки.

1. **Построение псевдовыборки из перекрывающихся блоков.** Исходная выборка делится на некоторое количество перекрывающихся блоков. Длина блока выбирается исследователем исходя из временной структуры ряда. Пусть $\{y_t\}_{t=1}^T$

– исходная выборка, а l – длина блока. Тогда в первый блок войдут наблюдения $y_1, \ldots, y_l$, во второй – $y_2, \ldots, y_{l+1}$, в третий – $y_3, \ldots, y_{l+2}$, и наконец в $T-l+1$-ый – наблюдения $y_{T-l+1}, \ldots, y_T$. При построении псевдовыборки блоки извлекаются случайно с возвращением.

2. **Построение псевдовыборки из неперекрывающихся блоков.** В данном случае исходная выборка делится на некоторое количество неперекрывающихся блоков. Длина блока, так же как и в предыдущем случае, выбирается исследователем. Пусть исходная выборка состоит из наблюдений $\{y_t\}_{t=1}^{T}$. Тогда в первый блок войдут наблюдения $y_1, \ldots, y_l$, во второй – $y_{l+1}, \ldots, y_{2l}$, и наконец, в последний $\left[\frac{T}{l}\right]$-ый блок – наблюдения $y_{l\left[\frac{T}{l}\right]-l+1}, \ldots, y_{l\left[\frac{T}{l}\right]}$. При построении псевдовыборки блоки извлекаются случайным образом с возвращением.

3. **Построение стационарной псевдовыборки.** Предыдущие два варианта построения псевдовыборки, как правило, нарушают стационарность ряда, т.е. из стационарной исходной выборки получаются нестационарные псевдовыборки. Чтобы получить стационарную псевдовыборку, был предложен способ, основанный на нефиксированной длине блоков. А именно, задается вероятность окончания блока p. Первый элемент псевдовыборки выбирается случайно. Затем с вероятностью $(1-p)$ в текущий блок включается следующий элемент исходной выборки, а с вероятностью p начинается новый блок, первый элемент которого снова выбирается случайно из исходной выборки. Так продолжается, пока в псевдовыборку не будет набрано нужное количество элементов.

III Основные эконометрические понятия

Данный раздел кратко повторяет основные понятия, изученные в курсе статистики и теории вероятностей.

1 Условное математическое ожидание

Пусть (X, Y) – пара случайных величин. Функция совместной плотности распределения

$$f_{(X,Y)}(x, y) \geq 0$$

обладает свойством нормировки

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dx dy = 1.$$

Вероятность для (X,Y) попасть в прямоугольник $[a,b]\times[c,d]$ определяется как

$$Pr\{a \le X \le b,\ c \le Y \le d\} = \int_c^d \int_a^b f_{(X,Y)}(x,y)dxdy.$$

Маргинальная функция плотности распределения X связана с совместной плотностью (X,Y) как

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)dy.$$

Условная плотность распределения Y при $X=x$ есть

$$f_{Y|X=x}(x,y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}.$$

Условные вероятности попадания Y в отрезок $[c,d]$ определяются выражениями

$$Pr\{c \le Y \le d \mid X = x\} = \int_c^d f_{Y|X=x}(x,y)dy,$$

$$Pr\{c \le Y \le d \mid a \le X \le b\} = \frac{\int_c^d \int_a^b f_{X,Y}(x,y)dxdy}{\int_a^b f_X(x)dx}.$$

Условное математическое ожидание Y при условии $X=x$ есть

$$E[Y \mid X = x] = \int_{-\infty}^{+\infty} y f_{Y|X=x}(x,y)dy.$$

Заметим, что функция условного математического ожидания $E[Y \mid X]$ является случайной величиной (так как X случаен). Имеет место *закон последовательных математических ожиданий*:

$$E[h(X,Y)] = E[E[h(X,Y) \mid X]],$$

где $h(X,Y)$ – произвольная функция от (X,Y). В случае непрерывных распределений справедливость этого закона легко показать:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f_{(X,Y)}(x,y)dxdy = \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} h(x,y)f_{Y|X}(x,y)dy\right] f_X(x)dx.$$

2 Предсказывание

Часто в эконометрике встречается ситуация, когда исследователь хочет по переменным X (называемых *регрессорами*) предсказать значение Y (называемой *зависимой переменной*). Существует несколько полезных результатов, связанных с такой постановкой задачи.

Теорема. Оптимальным предиктором Y из X в смысле минимизации среднеквадратичной ошибки предсказания является условное математическое ожидание $E[Y \mid X]$.

Доказательство. Пусть $g(X)$ – произвольный предиктор. Тогда среднеквадратичная ошибка предсказания будет выражаться следующим образом:

$$\begin{aligned} MSPE &= E[(Y - g(X))^2] = E[(Y - E[Y|X] + E[Y|X] - g(X))^2] = \\ &= E[(Y - E[Y|X])^2] + E[(E[Y|X] - g(X))^2] \geq E[(Y - E[Y|X])^2]. \end{aligned}$$

Заметим, что равенство достигается при $g(X) = E[Y|X]$, т.е. условное математическое ожидание действительно минимизирует среднеквадратичную ошибку предсказания.

Определение. *Ошибкой оптимального предсказания* называется величина $e = Y - E[Y|X]$.

Ошибка оптимального предсказания обладает следующими свойствами:

$$E[e|X] = 0 \quad \Rightarrow E[e] = 0 \quad \Rightarrow E[eh(X)] = 0.$$

Отсюда в частности следует, что условное математическое ожидание является несмещённым предиктором.

Определение. *Линейным предиктором Y по X* называется любая линейная функция от X: $g(X) = a + bX$.

Теорема. Оптимальным линейным предиктором Y по X в смысле минимизации среднеквадратичной ошибки предсказания является *наилучший линейный предиктор*

$$BLP(Y|X) = \alpha + \beta X, \quad \beta = \frac{Cov(X,Y)}{Var(X)}, \quad \alpha = E[Y] - \beta E[X].$$

Доказательство. Задача минимизации

$$MSPE = E[(Y - a - bX)^2] \to \min_{a,b}$$

имеет условия первого порядка

$$-E[2(Y - a - bX)] = 0, \quad -E[2(Y - a - bX)X] = 0.$$

Отсюда можно получить α и β из постановки теоремы.

Теорема. Наилучшей линейной аппроксимацией для условного среднего $E[Y|X]$ в смысле минимизации среднеквадратичной ошибки аппроксимации является наилучший линейный предиктор $BLP(Y|X)$.

Доказательство. Аналогично доказательству предыдущей теоремы нужно решить оптимизационную задачу

$$MSAE = E[(E[Y|X] - a - bX)^2] \to \min_{a,b}.$$

Получим α и β из постановки предыдущей теоремы.

Определение. Ошибкой наилучшего линейного предсказания называется величина $u = Y - BLP(Y|X)$.

Ошибка наилучшего линейного предсказания обладает свойствами

$$E[u] = 0, \quad E[uX] = 0.$$

Теорема. Если условное среднее $E[Y|X]$ линейно по X, то $E[Y|X] = BLP(Y|X)$.

3 Свойства двумерного нормального распределения

Рассмотрим двумерную случайную величину, распределенную согласно нормальному закону:

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right).$$

Ее плотность распределения задается следующим выражением:

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}}{2(1-\rho^2)}\right].$$

Ниже перечислены свойства такого распределения.

1. Каждая из компонент двумерной нормальной величины распределена нормально:
$$X \sim \mathcal{N}(\mu_X, \sigma_X^2).$$

2. Условное распределение $Y|X = x$ нормально:
$$Y|X = x \sim \mathcal{N}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2)\right).$$
Из этого свойства также вытекает условная гомоскедастичность и $E[Y|X] = BLP[X|Y]$.

3. Если Y и X нескоррелированы (т.е. $\rho = 0$), то Y и X независимы.

4. Линейная функция от нормальной случайной величины является также нормальной случайной величиной:
$$A\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N}\left(A\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, A\begin{pmatrix} \sigma_Y^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}A'\right).$$
Здесь A – 2×2 матрица линейного преобразования.

4 Свойства многомерного нормального распределения

Пусть Y – многомерная величина, распределенная согласно нормальному закону, т.е.

$$Y \sim \mathcal{N}(\mu, \Sigma),$$

где μ – $k \times 1$ вектор средних, Σ – $k \times k$ дисперсионная матрица. Плотность распределения Y есть

$$f_Y(y) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left[-\frac{(y-\mu)'\Sigma^{-1}(y-\mu)}{2}\right].$$

Разобьем Y на две части:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) = \mathcal{N}(\mu, \Sigma).$$

Многомерное нормальное распределение обладает следующими свойствами:

1. $Y_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$.
2. $Y_2|Y_1 = y_1 \sim \mathcal{N}\left(\mu_2 + B'(y_1 - \mu_1), \Sigma_{22} - B'\Sigma_{11}B\right)$, где $B = \Sigma_{11}^{-1}\Sigma_{12}$.
3. Если $\Sigma_{12} = 0$, то Y_1 и Y_2 независимы.
4. $g + HY \sim \mathcal{N}(g + H\mu, H\Sigma H')$, где g – фиксированный вектор, а H – матрица линейного преобразования.

5 Принцип аналогий

При построении всевозможных оценок используют *принцип аналогий*, основная идея которого состоит в замене неизвестной истинной функции распределения известной эмпирической. Эту идею мы уже встречали при изучении бутстрапа. Пусть интересующий нас параметр θ известным образом зависит от функции распределения X, $F_X(x)$. Тогда, согласно принципу аналогий, оценку $\widehat{\theta}$ можно построить, заменив истинную функцию распределения $F_X(x)$ на ее выборочный аналог

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[x_i \leq x].$$

Приведем примеры.

Пример 1. Пусть интересующий нас параметр есть теоретическое среднее:

$$\theta = E[X] = \int_{-\infty}^{+\infty} x dF(x).$$

Тогда, по принципу аналогий, его аналоговая оценка будет равна

$$\widehat{\theta} = \int_{-\infty}^{+\infty} x dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Пример 2: МНК-оценка. Покажем, что МНК-оценка также является аналоговой оценкой. Пусть исходная модель будет

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

Тогда параметр β находится из условия: $E[(y - x'\beta)x] = 0$. Его вид:

$$\beta = (E(xx'))^{-1}E(xy).$$

Используя принцип аналогий, получим МНК-оценку:

$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right).$$

Пример 3: ещё раз МНК-оценка. МНК-оценку можно получить как аналоговую и из условия минимизации среднеквадратичной ошибки прогноза. Исходная регрессионная модель в этом случае есть

$$E[y|x] = x'\beta.$$

Параметр β находится из условия минимизации среднеквадратичной ошибки:

$$\beta = \arg\min_b E[(y - x'b)^2].$$

Соответствующее аналоговое условие записывается в виде

$$\widehat{\beta} = \arg\min_b \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'b)^2.$$

Очевидно, что результатом решения этой задачи на экстремум вновь является МНК-оценка.

6 Основные понятия, связанные с регрессией

Сейчас и в дальнейшем мы не будем различать прописные и строчные буквы для обозначения случайных величин и их реализаций, иначе можно с ума сойти. Пусть у нас есть пара (y, x), где y – скаляр, а x – вектор.

Определение. Регрессией (в широком смысле) называется какое-либо свойство условного распределения y при заданном x как функция от x.

Приведем несколько примеров регрессий:

Пример 1. Регрессия среднего $E[y|x]$.

Пример 2. Медианная регрессия $Med[y|x]$.

Пример 3. Квантильная регрессия $q_\alpha[y|x]$.

Пример 4. Модальная регрессия $Mode[y|x]$.

Рассмотрим подробнее *регрессию среднего*, которая наиболее часто используется в эконометрическом анализе. Ошибкой регрессии среднего называется величина $e = y - E[y|x]$. Эта ошибка обладает свойством $E[e|x] = 0$, и, как следствие, свойством $E[eh(x)] = 0$ для любой функции $h(x)$. В частности, $E[e] = 0$. Однако регрессоры x и ошибка e могут не быть независимыми. Регрессию среднего среднего можно записать в привычном виде как

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

Пока еще не сделано никаких предположений, кроме существования введенных объектов.

Обычно исследователь, обладая совокупностью наблюдений $\{(y_i, x_i)\}_{i=1}^n$, случайным образом выбранных из популяции (y, x), хочет оценить, используя эти данные, функцию $E[y|x]$. Существует несколько различных подходов к данной задаче:

1. **Непараметрическое оценивание:** При таком подходе делаются слабые предположения о гладкости функции $E[y|x]$ и, возможно, ее производной по x и плотности распределения x.

2. **Параметрическое оценивание:** При таком подходе предполагается известным вид функции $E[y|x] = g(x, \beta)$. Неизвестными являются конечное число параметров $\beta \in \mathbb{R}^k$. Эти параметры оцениваются, что даёт оценку и для $g(x, \beta)$. Таким образом, функция условного среднего параметризуется, отсюда и название метода. Параметрический метод оценивания является более эффективным, чем непараметрический, если спецификация модели правильная. Однако, если функция условного среднего параметризована неверно, то оценивание приводит к несостоятельным оценкам для $E[y|x]$.

3. **Полупараметрическое оценивание:** Полупараметрическое оценивание представляет собой нечто среднее между параметрическим и непараметрическим подходами. Можно сказать, что $E[y|x]$ параметризуется, но количество параметров бесконечно. Примером являются так называемые индексные модели, где вид функции условного среднего нам неизвестен, однако известно, что она зависит от конечномерной линейной комбинации регрессоров.

IV Линейная регрессия среднего

1 Метод наименьших квадратов

Пусть $E[y|x] = x'\beta$, тогда регрессия среднего записывается привычным образом как

$$y = x'\beta + e, \quad E[e|x] = 0, \quad \{(y_i, x_i)\}_{i=1}^n \sim iid.$$

Предположим, матрица $E[xx']$ невырожденная. Тогда параметр β, минимизирующий среднеквадратичную ошибку, будет единственным решением задачи

$$\beta = \arg\min_b E[(y - x'b)^2].$$

Пользуясь принципом аналогий, можно построить оценку для β:

$$\widehat{\beta} = \arg\min_b \frac{1}{n}\sum_{i=1}^n (y_i - x_i'b)^2 = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i.$$

Это и есть оценка метода наименьших квадратов, или МНК-оценка.

2 Асимптотические свойства МНК-оценки

Для выяснения асимптотических свойств МНК-оценки перепишем ее в следующем виде:

$$\widehat{\beta} = \beta + \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i e_i.$$

Как мы уже знаем, МНК-оценка состоятельна, т.е. $\widehat{\beta} \xrightarrow{p} \beta$. Кроме того, МНК-оценка асимптотически распределена по нормальному закону:

$$\sqrt{n}(\widehat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n x_i e_i \xrightarrow{d} \mathcal{N}\left(0, Q_{xx}^{-1} Q_{e^2xx} Q_{xx}^{-1}\right),$$

где мы использовали следующие обозначения:

$$Q_{xx} = E[xx'], \quad Q_{e^2xx} = E[e^2xx'] = Var[xe].$$

Приведенные асимптотические свойства следуют из ЗБЧ и ЦПТ. Из закона больших чисел следует, что

$$\frac{1}{n}\sum_{i=1}^n x_i x_i' \xrightarrow{p} E[xx'] = Q_{xx},$$

$$\frac{1}{n}\sum_{i=1}^n x_i e_i \xrightarrow{p} E[xe] = E[xE[e|x]] = 0,$$

что влечёт состоятельность МНК-оценки. Из центральной предельной теоремы для независимых одинаково распределенных величин следует, что

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i e_i \xrightarrow{d} \mathcal{N}\left(0, Var[xe]\right) = \mathcal{N}\left(0, Q_{e^2xx}\right),$$

что влечёт асимптотическую нормальность МНК-оценки.

Рассмотрим специальный случай, когда регрессионная ошибка условно гомоскедастична, т.е. $E[e^2|x] = \sigma^2 = \text{const}$. В этом случае $Q_{e^2xx} = \sigma^2 Q_{xx}$ и асимптотическое распределение МНК-оценки имеет дисперсионную матрицу в упрощенном виде:

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 Q_{xx}^{-1}\right).$$

Кроме того, легко построить состоятельную оценку этой дисперсионной матрицы:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta})^2 \xrightarrow{p} \sigma^2, \quad \widehat{Q}_{xx} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} Q_{xx}.$$

Состоятельность первой оценки следует из ЗБЧ. Состоятельность последней довольно легко показать:

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta})^2 = \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \frac{1}{n}\sum_{i=1}^{n}(x_i'\beta - x_i'\widehat{\beta})^2 + \frac{2}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)(x_i'\beta - x_i'\widehat{\beta}) = \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + (\beta - \widehat{\beta})'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)(\beta - \widehat{\beta}) + \frac{2}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)x_i'(\beta - \widehat{\beta}).
\end{aligned}$$

Далее, применяя ЗБЧ и теорему Слуцкого, получим:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 \xrightarrow{p} \sigma^2,$$

$$(\beta - \widehat{\beta})'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)(\beta - \widehat{\beta}) \xrightarrow{p} 0,$$

$$\frac{2}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)x_i'(\beta - \widehat{\beta}) \xrightarrow{p} 0.$$

Всё вместе влечёт состоятельность оценки условной дисперсии регрессионной ошибки:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\beta})^2 \xrightarrow{p} \sigma^2.$$

Теперь рассмотрим общий случай условной гетероскедастичности. В этом случае нам нужно состоятельно оценить матрицу Q_{e^2xx}. Можно показать, что состоятельной оценкой этой матрицы будет следующая:

$$\widehat{Q}_{e^2xx} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i'(y_i - x_i'\widehat{\beta})^2 \xrightarrow{p} Q_{e^2xx}.$$

Итак, состоятельная оценка дисперсионной матрицы МНК-оценки в случае условной гетероскедастичности запишется как

$$\widehat{V}_\beta = \widehat{Q}_{xx}^{-1}\widehat{Q}_{e^2xx}\widehat{Q}_{xx}^{-1}.$$

Будем называть *стандартной ошибкой* оценки $\widehat{\beta}_j$ величину

$$se(\widehat{\beta}_j) = \sqrt{\frac{1}{n}\left[\widehat{V}_\beta\right]_{jj}}.$$

Тогда соответствующая *t-статистика* будет асимптотически пивотальной оценкой, асимптотическое распределение которой является стандартным нормальным:

$$t_j = \frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \xrightarrow{d} \mathcal{N}(0,1).$$

Вальдовская статистика для ограничений общего вида $h(\beta) = 0$, где число ограничений $l \leq k$, имеет асимптотическое распределение хи-квадрат:

$$W = h(\widehat{\beta})'\left[\widehat{H}\widehat{V}_\beta\widehat{H}'\right]^{-1} h(\widehat{\beta}) \xrightarrow{d} \chi^2(l),$$

где использованы обозначения

$$H = \frac{\partial h(\beta)}{\partial \beta'}, \quad \widehat{H} = \frac{\partial h(\widehat{\beta})}{\partial \beta'}.$$

3 Свойства МНК-оценки в конечных выборках

Введем следующие обозначения:

$$X = (x_1, x_2, \ldots, x_n)', \quad y = (y_1, y_2, \ldots, y_n)', \quad \varepsilon = (e_1, e_2, \ldots, e_n)'.$$

Тогда уже знакомую нам регрессионную модель линейного условного среднего можно переписать в матричном виде

$$y = X\beta + \varepsilon, \quad E[\varepsilon|X] = 0.$$

МНК-оценка в таком случае запишется как

$$\widehat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon.$$

Эта оценка обладает следующими свойствами в конечных выборках:

- Условная несмещённость:

$$E[\widehat{\beta}|X] = \beta + (X'X)^{-1}X'E[\varepsilon|X] = \beta.$$

- Безусловная несмещённость следует из условной несмещённости.

- Условная дисперсия оценки есть

$$Var[\widehat{\beta}|X] = (X'X)^{-1}X'\Omega X(X'X)^{-1},$$

где $\Omega = Var[y|X] = E[\varepsilon\varepsilon'|X]$.

4 Обобщенный метод наименьших квадратов

Определение. Пусть $E[y|X] = X\beta$. Классом *линейных оценок* β называется класс, содержащий оценки вида $A(X)y$, где $A(X)$ – матрица $k \times n$, которая зависит только от X.

Пример. Для МНК-оценки $A(X) = (X'X)^{-1}X'$.

Определение. Пусть $E[y|X] = X\beta$. Классом *линейных несмещённых оценок* β называется класс, содержащий оценки вида $A(X)y$, где $A(X)$ – матрица $k\times n$, зависящая только от X и удовлетворяющая условию $A(X)X = I_k$.

Пример. Для МНК-оценки $A(X)X = (X'X)^{-1}X'X = I_k$.

Заметим, что $Var[A(X)y|X] = A(X)\Omega A(X)'$. Мы хотим найти линейную несмещённую оценку, которая минимизирует $Var[A(X)y|X]$.

Теорема (Гаусса-Маркова). Наилучшей линейной несмещённой оценкой линейной регрессии среднего является оценка $\widetilde{\beta} = A^*(X)y$, где

$$A^*(X) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}.$$

В этом случае дисперсионная матрица оценки имеет вид

$$Var[\widetilde{\beta}|X] = (X'\Omega^{-1}X)^{-1}.$$

Доказательство. Оценка $\widetilde{\beta}$ принадлежит классу линейных несмещённых оценок, ибо $A^*(X)X = I_k$. Возьмем произвольную матрицу $A(X)$, такую, что $A(X)X = I_k$. В этом случае имеют место следующие равенства:

$$(A(X) - A^*(X))X = 0,$$

$$(A(X) - A^*(X))\Omega A^*(X)' = (A(X) - A^*(X))\Omega\Omega^{-1}X(X'\Omega^{-1}X)^{-1} = 0.$$

Тогда

$$\begin{aligned}
& Var[A(X)Y|X] = A(X)\Omega A(X)' = \\
= \;& (A(X) - A^*(X) + A^*(X))\Omega(A(X) - A^*(X) + A^*(X))' = \\
= \;& (A(X) - A^*(X))\Omega(A(X) - A^*(X))' + Var[A^*(X)Y|X] \geq Var[\widetilde{\beta}|X].
\end{aligned}$$

Следовательно, оценка $\widetilde{\beta}$ является наилучшей в классе линейных несмещённых оценок.

Определение. Пусть $E[y|X] = X\beta$. Оценка $\widetilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ называется оценкой *обобщённого метода наименьших квадратов* (ОМНК).

Следствие 1. ОМНК-оценка $\widetilde{\beta}$ является эффективной в классе линейных несмещённых оценок.

Следствие 2. Если ошибка линейной регрессии среднего обладает свойством условной гомоскедастичности, то $\widetilde{\beta} = \widehat{\beta}$, т.е. МНК- и ОМНК-оценки совпадают.

Ниже приведена таблица, содержащая условные дисперсионные матрицы МНК- и ОМНК-оценок в конечных выборках для случаев условной гетеро- и гомоскедастичности.

	OLS	GLS
Гомоскедастичность	$\sigma^2(X'X)^{-1}$	$\sigma^2(X'X)^{-1}$
Гетероскедастичность	$(X'X)^{-1}X'\Omega X(X'X)^{-1}$	$(X'\Omega^{-1}X)^{-1}$

Замечание 1. ОМНК-оценка $\widetilde{\beta}$ является *недоступной*, поскольку матрица Ω неизвестна.

Замечание 2. ОМНК-оценка $\widetilde{\beta}$ является частным случаем оценки *взвешенного метода наименьших квадратов* (ВМНК)

$$\widehat{\beta}_{WLS} = (X'WX)^{-1}X'WY,$$

где W – положительно определенная матрица.

5 Асимптотические свойства ОМНК-оценок

Рассмотрим асимптотические свойства ОМНК-оценки. Для этого представим ее в следующем виде:

$$\widetilde{\beta} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{x_i x_i'}{\sigma^2(x_i)}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{x_i e_i}{\sigma^2(x_i)}.$$

Пользуясь законом больших чисел и центральной предельной теоремой, получим:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{x_i x_i'}{\sigma^2(x_i)} \xrightarrow{p} Q_{xx/\sigma^2} = E\left[\frac{xx'}{\sigma^2(x)}\right],$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{x_i e_i}{\sigma^2(x_i)} \xrightarrow{p} E\left[\frac{xe}{\sigma^2(x)}\right] = 0,$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_i e_i}{\sigma^2(x_i)} \xrightarrow{d} \mathcal{N}\left(0, Q_{xx/\sigma^2}\right).$$

Последнее выражение следует из

$$E\left[\left(\frac{xe}{\sigma^2(x)}\right)^2\right] = E\left[\frac{xx'}{\sigma^2(x)}E[e^2|x]\right] = Q_{xx/\sigma^2}.$$

Таким образом, ОМНК-оценка является состоятельной и асимптотически нормальной.

Ниже приведена таблица, содержащая асимптотические дисперсионные матрицы МНК- и ОМНК-оценок для случаев условной гетеро- и гомоскедастичности:

	OLS	GLS
Гомоскедастичность	$\sigma^2 Q_{xx}^{-1}$	$Q_{xx/\sigma^2}^{-1} = \sigma^2 Q_{xx}^{-1}$
Гетероскедастичность	$Q_{xx}^{-1} Q_{e^2xx} Q_{xx}^{-1}$	Q_{xx/σ^2}^{-1}

Теорема. ОМНК-оценка $\widetilde{\beta}$ асимптотически эффективна в классе оценок вида

$$\widehat{\beta}_{IV} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i y_i,$$

где $z_i = f(x_i)$ для любой функции $f : \mathbb{R}^k \to \mathbb{R}^k$.

Доказательство. Заметим, что МНК- и ОМНК-оценки принадлежат указанному классу, т.к. для МНК $z_i = x_i$, а для ОМНК $z_i = x_i/\sigma^2(x_i)$. Рассмотрим оценку

$$\widehat{\beta}_{IV} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i y_i.$$

Легко показать, что она состоятельна и асимптотически нормальна с асимптотической дисперсионной матрицей

$$V_{zz} = Q_{zx}^{-1} Q_{e^2zz} Q_{xz}^{-1},$$

где $Q_{zx} = E[zx']$, а $Q_{e^2zz} = E[zz'e^2] = E[zz'\sigma^2(x)]$. Зная, что асимптотическая дисперсия ОМНК-оценки равна Q^{-1}_{xx/σ^2}, рассмотрим разность

$$\begin{aligned} V_{zz} - Q^{-1}_{xx/\sigma^2} &= (E[zx'])^{-1} E[zz'\sigma^2(x)] (E[xz'])^{-1} - \left(E\left[\frac{xx'}{\sigma^2(x)}\right]\right)^{-1} = \\ &= (E[vu'])^{-1} E[vv'] (E[uv'])^{-1} - (E[uu'])^{-1} = \\ &= (E[vu'])^{-1} \left[E[vv'] - E[vu'] (E[uu'])^{-1} E[uv']\right] (E[uv'])^{-1} = \\ &= (E[vu'])^{-1} E[ww'] (E[uv'])^{-1} \geq 0. \end{aligned}$$

Здесь $v = z\sigma(x)$, $u = x/\sigma(x)$ и $w = v - E[vu'](E[uu'])^{-1}u$. Таким образом, мы показали, что ОМНК-оценка асимптотически эффективна в указанном классе.

Результат. ОМНК-оценка является аналоговой оценкой, т.е. может быть получена из принципа аналогий. А именно, ОМНК-оценка получается из условия

$$E\left[e\frac{x}{\sigma(x)}\right] = 0 \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widetilde{\beta})\frac{x_i}{\sigma(x_i)} = 0.$$

6 Доступная ОМНК-оценка

Как уже было замечено, чтобы получить ОМНК-оценку, нам необходимо знать дисперсионную матрицу ошибок Ω, главная диагональ которой напичкана величинами $\sigma^2(x_i)$, а на остальных местах стоят нули. Естественно полагать, что эти параметры являются неизвестными априори, поэтому они должны быть оценены. Плохо то, что для этого необходима модель для $\sigma^2(x)$. Эту функцию можно (и нужно!) оценивать непараметрически, но мы пока к этому не готовы.

Обычно предполагают, что дисперсия ошибок есть линейная функция от некоторой трансформации x:

$$\sigma^2(x) = E[e^2|x] = z'\gamma,$$

где z есть некоторая трансформация x, например $z = x^2$. Если предположение правильное, то можно оценить *скедастичную регрессию*

$$e^2 = z'\gamma + \varepsilon, \quad E[\varepsilon|z] = 0.$$

Оценив исходную регрессию с помощью МНК и скедастичную регрессию, используя квадраты МНК-остатков вместо квадратов ошибок, также с помощью МНК, имеем:

$$\widehat{e}_i = y_i - x_i'\widehat{\beta} = e_i + x_i'(\beta - \widehat{\beta}),$$

$$\widehat{\gamma} = \left(\sum_i z_i z_i'\right)^{-1} \left(\sum_i z_i e_i^2 + 2\sum_i z_i x_i'(\beta - \widehat{\beta})e_i + \sum_i z_i(x_i'(\beta - \widehat{\beta}))^2\right) \xrightarrow{p} \gamma.$$

Таким образом, мы получаем состоятельные оценки дисперсий ошибок:

$$\widehat{\sigma}^2(x_i) = z_i'\widehat{\gamma},$$

после чего, мы можем построить доступную оценку обобщенного метода наименьших квадратов (ДОМНК):

$$\widetilde{\beta}_F = \left(\sum_{i=1}^{n} \frac{x_i x_i'}{\widehat{\sigma}^2(x_i)}\right)^{-1} \sum_{i=1}^{n} \frac{x_i y_i}{\widehat{\sigma}^2(x_i)} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y.$$

Приведем алгоритм построения ДОМНК-оценки:

1. Используя МНК, оценить исходную регрессию и получить остатки $\widehat{e}_i$ для $i = 1, \dots, n$. Прогнать скедастичную регрессию, получить оценки $\widehat{\gamma}$ и построить оценки дисперсий ошибок $\widehat{\sigma}^2(x_i)$ (или $\widehat{\Omega}$).

2. Построить ДОМНК-оценку

$$\widetilde{\beta}_F = \left(\sum_{i=1}^{n} \frac{x_i x_i'}{\widehat{\sigma}^2(x_i)}\right)^{-1} \sum_{i=1}^{n} \frac{x_i y_i}{\widehat{\sigma}^2(x_i)} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y.$$

Вообще говоря, такой алгоритм построения оценок дисперсии ошибок не гарантирует их положительность. Ниже приведены способы избежать $\widehat{\sigma}^2(x_i) < 0$.

1. Выбрать некоторое малое $\delta > 0$. Положить $\widehat{\sigma}^2(x_i) = \max(z_i'\widehat{\gamma}, \delta)$.

2. Выбросить те наблюдения, для которых $\widehat{\sigma}^2(x_i) < 0$.

3. Положить $\widehat{\sigma}^2(x_i) = \frac{1}{n}\sum_{j=1}^{n} z_j'\widehat{\gamma}$ для тех наблюдений, для которых $\widehat{\sigma}^2(x_i) < 0$.

Результат. Если скедастичная функция правильно специфицирована, то ДОМНК-оценка $\widetilde{\beta}_F$ асимптотически эквивалентна ОМНК-оценке $\widetilde{\beta}$, т.е.

$$\sqrt{n}(\widetilde{\beta}_F - \beta) \xrightarrow{d} \mathcal{N}(0, Q_{xx/\sigma^2}^{-1}).$$

Состоятельная оценка асимптотической дисперсии в этом случае есть

$$\widehat{V}_\beta = n\left(\sum_{i=1}^{n} \frac{x_i x_i'}{\widehat{\sigma}^2(x_i)}\right)^{-1}.$$

Если скедастичная функция специфицирована неправильно, то оценка $\widetilde{\beta}_F$, тем не менее, остается состоятельной и асимптотически нормальной:

$$\sqrt{n}(\widetilde{\beta}_F - \beta) \xrightarrow{d} \mathcal{N}\left(0, Q_{xx/\sigma^2}^{-1} Q_{xx/\sigma^4 e^2} Q_{xx/\sigma^2}^{-1}\right),$$

где использованы следующие обозначения:

$$Q_{xx/\sigma^2} = E\left[\frac{xx'}{z'\gamma}\right], \quad Q_{xx/\sigma^4 e^2} = E\left[\frac{xx'}{(z'\gamma)^2}e^2\right].$$

Состоятельная оценка асимптотической дисперсии в этом случае равна

$$\widehat{V}_\beta = n\left(\sum_i \frac{x_i x_i'}{z_i'\widehat{\gamma}}\right)^{-1} \sum_i \frac{x_i x_i'}{(z_i'\widehat{\gamma})^2}\widehat{e}_i^2 \left(\sum_i \frac{x_i x_i'}{z_i'\widehat{\gamma}}\right)^{-1}.$$

7 Регрессия с неслучайной выборкой

В случае, когда наблюдения представляют собой неслучайную выборку, дисперсионная матрица ошибок $\Omega = Var[y|X]$ не является диагональной. Дисперсия МНК-оценки $\widehat{\beta} = (X'X)^{-1}X'y$ в этом случае есть

$$Var[\widehat{\beta}|X] = (X'X)^{-1}X'\Omega X(X'X)^{-1},$$

а дисперсия ОМНК-оценки $\widetilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ есть

$$Var[\widetilde{\beta}|X] = (X'\Omega^{-1}X)^{-1}.$$

Чтобы построить пивотальную статистику в случае неслучайных наблюдений, необходимо параметризовать дисперсионную матрицу ошибок Ω небольшим числом параметров.

8 МНК и ОМНК в регрессиях на временных рядах

Рассмотрим следующую регрессионную модель:

$$y_t = x_t'\beta + e_t, \quad E[e_t|I_{t-1}] = 0, \quad E[e_t^2|I_{t-1}] = \sigma^2(I_{t-1}),$$

где $\{(x_t, y_t)\}_{t=1}^T$ – стационарный и эргодичный процесс, а

$$I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots; x_t, x_{t-1}, \ldots\}.$$

Примерами таких моделей могут служить:

- Модель $AR(p)$, где $x_t = (y_{t-1}, y_{t-2}, \ldots, y_{t-p})'$.
- Линейное предсказание обменного курса:

$$s_{t+1} - s_t = \alpha + \beta(f_t - s_t) + e_t, \quad E[e_t|I_{t-1}] = 0,$$

  где f_t – цена форвардного контракта, а s_t – текущий обменный курс.

- Линейное предсказание инфляции:

$$\pi_{t+1} = \alpha + \beta i_t + e_t, \quad E[e_t|I_{t-1}] = 0,$$

где π_{t+1} – инфляция, а i_t – процентная ставка.

Заметим, что условное математическое ожидание $E[Y|X]$ не есть $X\beta$, поэтому не следует ожидать хороших свойств в конечных выборках. Рассмотрим асимптотические свойства МНК- и ОМНК-оценок.

МНК-оценка. Ясно, что МНК-оценка состоятельна:

$$\widehat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t \xrightarrow{p} \beta.$$

Это следует из того, что:

$$E[x_t e_t] = E[E[x_t e_t|I_{t-1}]] = E[x_t E[e_t|I_{t-1}]] = 0.$$

Кроме того, из центральной предельной теоремы для последовательностей мартингальных приращений следует асимптотическая нормальность МНК-оценки:

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, V_\beta), \quad V_\beta = Q_{xx}^{-1} Q_{e^2xx} Q_{xx}^{-1},$$

где $Q_{xx} = E[x_t x_t']$, $Q_{e^2xx} = E[x_t x_t' e_t^2]$. Недиагональные элементы матрицы Q_{e^2xx} равны 0, поскольку

$$E[x_t e_t x_{t-j}' e_{t-j}] = E[E[x_t e_t x_{t-j}' e_{t-j}|I_{t-1}]] = E[x_t E[e_t|I_{t-1}] x_{t-j}' e_{t-j}] = 0.$$

ОМНК-оценка. ОМНК-оценка в моделях без серийной корреляциии в ошибках, очевидно, тоже будет состоятельна и асимптотически нормальна. Выглядит ОМНК-оценка следующим образом:

$$\widetilde{\beta} = \left(\sum_{t=1}^{T} \frac{x_t}{\sigma^2(I_{t-1})} x_t'\right)^{-1} \sum_{t=1}^{T} \frac{x_t}{\sigma^2(I_{t-1})} y_t.$$

На практике ОМНК-оценка редко используется во временных рядах, поскольку требует знания или состоятельного оценивания скедастичной функции $\sigma^2(I_{t-1})$, которая в принципе может зависеть от бесконечной предыстории I_{t-1}.

Теперь рассмотрим регрессионную модель на временных рядах более общего вида, с возможностью серийной корреляции в ошибках:

$$y_t = x_t'\beta + e_t, \quad E[e_t|I_{t-q}] = 0, \quad E[e_t^2|I_{t-q}] = \sigma^2(I_{t-q}),$$

где $I_{t-q} = \{y_{t-q}, y_{t-q-1}, \ldots; x_t, x_{t-1}, \ldots\}$.

Примерами таких моделей могут служить:

- Модель $ARMA(p,q)$, где $x_t = (y_{t-q}, y_{t-q-1}, \ldots, y_{t-q-p+1})'$.

- Линейное предсказание обменного курса:

$$s_{t+q} - s_t = \alpha + \beta(f_{t,q} - s_t) + e_t, \quad E[e_t|I_{t-q}] = 0,$$

  где $f_{t,q}$ – цена форвардного контракта на q периодов вперед, а s_t – текущий обменный курс.

- Линейное предсказание инфляции:

$$\pi_{t+q} = \alpha + \beta i_{t,q} + e_t, \quad E[e_t|I_{t-q}] = 0,$$

  где π_{t+q} – инфляция, а $i_{t,q}$ – процентная ставка на q периодов вперед.

МНК-оценка.

$$\widehat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t.$$

Оценка остается состоятельной и асимптотически нормальной, т.е.

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, V_\beta), \quad V_\beta = Q_{xx}^{-1} Q_{e^2xx} Q_{xx}^{-1},$$

но матрица Q_{e^2xx} считается по формуле

$$Q_{e^2xx} = E[x_t x_t' e_t^2] + \sum_{j=1}^{q-1} (E[x_t x_{t-j}' e_t e_{t-j}] + E[x_t x_{t+j}' e_t e_{t+j}]).$$

Сумма содержит конечное число компонент $2q-1$, ибо для $j > q-1$

$$E[x_t x_{t-j}' e_t e_{t-j}] = E[E[x_t X_{t-j}' e_t e_{t-j}|I_{t-q}]] = E[x_t x_{t-j}' E[e_t|I_{t-q}] e_{t-j}] = 0.$$

Чтобы состоятельно оценить асимптотическую дисперсию МНК-оценки V_β в случае серийной корреляции ошибок, можно воспользоваться, например, формулой Ньюи-Уэста.

ОМНК-оценка. ОМНК-оценка не используется в моделях с серийной корреляцией ошибок. Заметим,что в этом случае

$$\widetilde{\beta} \neq \left(\sum_{t=1}^{T} \frac{x_t}{\sigma^2(I_{t-1})} x_t'\right)^{-1} \sum_{t=1}^{T} \frac{x_t}{\sigma^2(I_{t-1})} y_t.$$

V Линейные модели с инструментальными переменными

1 Эндогенные переменные

Бывают случаи, когда условное среднее $E[y|x]$ не является интересующим нас объектом. Приведем несколько примеров.

Пример 1. Пусть $E[y|x^*] = (x^*)'\beta$, однако переменные x^* являются ненаблюдаемыми. Вместо них исследователь наблюдает переменные $x = x^* + u$, где u – независима от x^* и y. В этом случае МНК-оценка будет несостоятельной:

$$y = (x^*)'\beta + e = (x-u)'\beta + e = x'\beta + v, \quad v = e - u'\beta,$$

$$E[xv] = E[(x^* + u)(e - u'\beta)] = -E[uu'] \neq 0 \quad \Rightarrow \quad \widehat{\beta} \overset{p}{\nrightarrow} \beta.$$

В такой ситуации асимптотическое смещение оценки связано с ошибкой измерения.

Пример 2. Пусть у нас есть система линейных уравнений

$$\begin{aligned} \text{Спрос} \quad &: \quad Q = -\beta_1 P + e_1 \\ \text{Предложение} \quad &: \quad Q = \beta_2 P + e_2 \end{aligned}$$

$$\text{где } \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim iid(0, I_2).$$

Очевидно, что в этом случае цены коррелируют с ошибками:

$$E[e_1 P] \neq 0, \quad E[e_2 P] \neq 0.$$

Поэтому, используя МНК-оценку в регрессии Q на P, мы получим

$$\widehat{\beta} \xrightarrow{p} \frac{E[QP]}{E[P^2]}.$$

Исходную систему легко разрешить относительно выпуска и цен:

$$\begin{pmatrix} Q \\ P \end{pmatrix} = \frac{1}{\beta_1 + \beta_2} \begin{pmatrix} \beta_2 & \beta_1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix},$$

откуда сразу же следует, что

$$E[QP] = \frac{\beta_2 - \beta_1}{\beta_1 + \beta_2}, \quad E[P^2] = \frac{2}{\beta_1 + \beta_2}.$$

Таким образом, МНК-оценка не состоятельна ни для β_1, ни для β_2, а состоятельна для чего-то среднего между ними:

$$\widehat{\beta} \xrightarrow{p} \frac{E[QP]}{E[P^2]} = \frac{\beta_2 - \beta_1}{2}.$$

В обоих примерах переменные, коррелирующие с ошибками, являются эндогеными, и МНК-оценивание несостоятельно.

Определение. Переменная x в правой части структурного уравнения $y = x'\beta + e$ называется *эндогенной*, если $E[e|x] \neq 0$.

Определение. Переменная z называется экзогенной для структурного уравнения $y = x'\beta + e$, если $E[e|z] = 0$. В этом случае z может использоваться как *инструмент* для оценивания параметров уравнения.

Заметим, что в обычной регрессии условного среднего регрессоры являются экзогенными переменными, т.е. для модели

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

$z = x$ является экзогенной переменной. МНК как раз и использует её как инструмент.

2 Точная идентификация

Рассмотрим случай, когда количество инструментов l совпадает с количеством "регрессоров" k (случай *точной идентификации*). Пусть матрица $Q_{zz} = E[zz']$ невырождена. Из определения инструментов следует, что $E[ze] = 0$. Это условие называется *условием валидности* инструментов. Применяя к нему принцип аналогий, получим

$$\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\widehat{\beta}_{IV}) = 0,$$

откуда получаем *инструментальную оценку*

$$\widehat{\beta}_{IV} = \left(\sum_{i=1}^{n} z_i x_i'\right)^{-1} \sum_{i=1}^{n} z_i y_i.$$

В матричном виде инструментальная оценка выглядит следующим образом:

$$\widehat{\beta}_{IV} = (Z'X)^{-1}Z'Y, \quad Z = (z_1', z_2', \ldots, z_n')'.$$

В конечных выборках инструментальная оценка смещена, однако является состоятельной и асимптотически нормальной:

$$E[\widehat{\beta}_{IV}|Z, X] = (Z'X)^{-1}Z'E[Y|Z, X] \neq \beta,$$

$$\sqrt{n}(\widehat{\beta}_{IV} - \beta) \xrightarrow{d} \mathcal{N}(0, V_\beta), \quad V_\beta = Q_{zx}^{-1}Q_{e^2zz}Q_{xz}^{-1},$$

где $Q_{zx} = E[zx']$, $Q_{e^2zz} = E[e^2zz']$. Заметим, что вышеприведённая асимптотика справедлива только при выполнении *условия релевантности*: матрица Q_{zx} должна быть невырожденной. Состоятельные оценки матриц Q_{zx}, Q_{e^2zz} строятся очевидным образом:

$$\widehat{Q}_{zx} = \frac{1}{n}\sum_{i=1}^{n} z_i x_i', \quad \widehat{Q}_{e^2zz} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \widehat{e}_i^2, \quad \widehat{e}_i = y_i - x_i'\widehat{\beta}_{IV}.$$

Заметим, что домножение инструментов слева на любую соразмерную невырожденную матрицу констант C не меняет вида инструментальной оценки $\widehat{\beta}_{IV}$, и, естественно, её статистических свойств.

3 Сверхидентификация

Рассмотрим случай, когда $l > k$ (случай *сверхидентификации*). Пусть матрица Q_{zz} невырождена. Идея построения инструментальной оценки в этом случае состоит в следующем. Сначала найдём линейный по z предиктор x:

$$x = \Gamma z + u, \quad E[zu'] = 0.$$

Из последнего условия находим Γ:

$$E[z(x - \Gamma z)'] = 0 \quad \Rightarrow \quad \Gamma' = (E[zz'])^{-1}E[zx'] = Q_{zz}^{-1}Q_{zx}.$$

Теперь, возвращаясь к исходной структурной форме, мы можем написать:

$$y = (\Gamma z + u)'\beta + e = (\Gamma z)'\beta + v, \quad v = e + u'\beta.$$

Очевидно, выполняется соотношение

$$E[\Gamma z v] = \Gamma E[z(e + u'\beta)] = 0,$$

поэтому параметр структурного уравнения можно записать как

$$\beta = (E[\Gamma z(\Gamma z)'])^{-1}E[\Gamma z y] = (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}Q_{zy}.$$

Применяя принцип аналогий, получим инструментальную оценку (называемую оценкой *двухшагового метода наименьших квадратов*) для случая сверхидентификации:

$$\widehat{\beta}_{2SLS} = \left(\sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i y_i,$$

или в матричной форме,

$$\widehat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y.$$

Инструментальная оценка состоятельна и асимптотически нормальна, однако в общем случае не является асимптотически эффективной:

$$\widehat{\beta}_{2SLS} \xrightarrow{p} \beta, \quad \sqrt{n}(\widehat{\beta}_{2SLS} - \beta) \xrightarrow{d} \mathcal{N}(0, V_{2SLS}),$$

где

$$V_{2SLS} = (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}Q_{e^2zz}Q_{zz}^{-1}Q_{zx}(Q_{xz}Q_{zz}^{-1}Q_{xz})^{-1}.$$

Особым случаем является случай условной гомоскедастичности, когда $E[e^2|z] = \sigma^2 = \text{const}$. В этом случае асимптотическая дисперсионная матрица упрощается, поскольку $Q_{e^2zz} = \sigma^2 Q_{zz}$:

$$V_{2SLS} = \sigma^2(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}.$$

Заметим, что несмотря на присутствие двух шагов в названии, оценку можно подсчитать сразу по одной формуле. В то же время понимание её двухшаговой природы помогает в некоторых случаях лучше ориентироваться в свойствах оценки.

Аналогично случаю точной идентификации при сверхидентификации необходимо, чтобы ранг матрицы Q_{xz} был равен количеству регрессоров k:

$$rank(Q_{xz}) = k.$$

Замечание. Оценку $\widehat{\beta}_{2SLS}$ можно выразить как инструментальную оценку с точной идентификацией:

$$\widehat{\beta}_{2SLS} = \left(\sum_{i=1}^{n} \xi_i x_i'\right)^{-1} \sum_{i=1}^{n} \xi_i y_i, \quad \xi_i = \sum_{j=1}^{n} x_j z_j' \left(\sum_{j=1}^{n} z_j z_j'\right)^{-1} z_i.$$

Здесь ξ_i – трансформированные инструменты. Но с точки зрения асимптотической эффективности такая трансформация неоптимальна.

4 Неполная идентификация

Если матрица Q_{zx} имеет ранг меньше k, то условие идентификации не выполнено. В инструментах не хватает информации, способной однозначно идентифицировать параметр. В этом случае инструментальные оценки будут иметь неприятные асимптотические свойства.

Пример. Рассмотрим следующую модель:

$$l = k = 1, \quad y = \beta x + e, \quad E[e|z] = 0, \quad E[xz] = 0.$$

Последнее равенство означает, что ранг Q_{xz} равен 0, т.е. не выполнено условие релевантности инструментов. Согласно центральной предельной теореме имеет место совместная сходимость:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i e_i \xrightarrow{d} \mathcal{N}(0, Q_{z^2e^2}), \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i x_i \xrightarrow{d} \mathcal{N}(0, Q_{z^2x^2}).$$

Таким образом, инструментальная оценка уже не будет асимптотически нормальной:

$$\widehat{\beta}_{IV} = \frac{\sum z_i y_i}{\sum z_i x_i} = \beta + \frac{\frac{1}{\sqrt{n}}\sum_i z_i e_i}{\frac{1}{\sqrt{n}}\sum_i z_i x_i} \xrightarrow{d} \beta + \mathcal{D},$$

где $\mathcal{D}$ – некое негауссовское распределение с тяжёлыми хвостами. Таким образом, асимптотическое распределение оценки не является асимптотически нормальным и даже не имеет конечного среднего.

Если у нас есть подозрение, что инструменты нерелевантные, стоит вначале протестировать гипотезу: $H_0 : E[xz] = 0$, и если эта гипотеза отвергается, то инструмент надёжный. На практике чаще всего руководствуются значением F-статистики при линейной регрессии x на z.

5 Бутстрапирование инструментальных оценок

Процедура бутстрапа для инструментальных оценок практически ничем не отличается от построения бутстраповского распределения для обычной статистики от независимых наблюдений. Есть, правда, тонкий момент в случае сверхидентификации.

Случай $l = k$.

$$\widehat{\beta}_{IV} = \left(\sum z_i x_i'\right)^{-1} \sum z_i y_i, \quad \widehat{\beta}^*_{IV} = \left(\sum z_i^* x_i^{*\prime}\right)^{-1} \sum z_i^* y_i^*.$$

Если оценка асимптотической дисперсионной матрица выглядит как

$$\widehat{V}_{IV} = n\left(\sum z_i x_i'\right)^{-1} \sum z_i z_i' \widehat{e}_i^2 \left(\sum x_i z_i'\right)^{-1},$$

то её бутстраповский аналог –

$$\widehat{V}^*_{IV} = n\left(\sum z_i^* x_i^{*\prime}\right)^{-1} \sum z_i^* z_i^{*\prime} \widehat{e}_i^{*2} \left(\sum x_i^* z_i^{*\prime}\right)^{-1}.$$

Случай $l > k$. Заметим, что хотя в популяции выполнено условие $E[ze] = 0$, в выборке оно нарушается, ибо

$$\frac{1}{n}\sum_{i=1}^{n} z_i \widehat{e}_i \neq 0.$$

Поэтому при бутстрапе нужно помнить про рецентрирование. Итак, инструментальная оценка есть

$$\widehat{\beta}_{2SLS} = (\ldots)^{-1} \sum x_i z_i' \left(\sum z_i z_i'\right)^{-1} \sum z_i y_i,$$

а её бутстраповский аналог –

$$\widehat{\beta}^*_{2SLS} = (\ldots^*)^{-1} \sum x_i^* z_i^{*\prime} \left(\sum z_i^* z_i^{*\prime}\right)^{-1} \left(\sum z_i^* y_i^* - \sum z_i \widehat{e}_i\right).$$

Соответственно,

$$\widehat{V}_{2SLS} = n(\ldots)^{-1} \sum x_i z_i' \left(\sum z_i z_i'\right)^{-1} \sum z_i z_i' \widehat{e}_i^2 \left(\sum z_i z_i'\right)^{-1} \sum z_i x_i' (\ldots)^{-1},$$

$$\widehat{V}^*_{2SLS} = n(\ldots^*)^{-1} \sum x_i^* z_i^{*\prime} \left(\sum z_i^* z_i^{*\prime}\right)^{-1} \sum u_i^* u_i^{*\prime} \left(\sum z_i^* z_i^{*\prime}\right)^{-1} \sum z_i^* x_i^{*\prime} (\ldots^*)^{-1}.$$

Здесь $u_i^* = z_i^* \widehat{e}_i^* - \frac{1}{n} \sum_{j=1}^n z_j \widehat{e}_j$.

6 Инструментальные переменные во временных рядах

Рассмотрим следующую модель временного ряда:

$$y_t = x_t' \beta + e_t, \quad E[e_t | I_{t-1}] = 0, \quad I_{t-1} = \{y_{t-1}, y_{t-2} \ldots ; x_t, x_{t-1}, \ldots\}.$$

Возьмём вектор инструментов

$$z_t = (y_{t-1}, y_{t-2}, \ldots, y_{t-l_y}, x_t', x_{t-1}', \ldots, x_{t-l_x}')'.$$

Он валидный, так как все элементы принадлежат I_{t-1}, и

$$E[e_t | z_t] = E[E[e_t | I_{t-1}] | z_t] = 0.$$

При таким образом выбранном инструменте инструментальная оценка $\widehat{\beta}_{2SLS}$ совпадает с МНК-оценкой (упражнение: почему?), и, соответственно, обладает теми же свойствами. Поэтому в данной задаче обычно используют расширение инструментальных оценок – оценки обобщенного метода моментов. То же самое справедливо и в более общей модели, допускающей автокорреляцию ошибок:

$$y_t = x_t' \beta + e_t, \quad E[e_t | I_{t-q}] = 0, \quad z_t = \{y_{t-q}, \ldots, y_{t-l_y}, x_t', \ldots, x_{t-l_x}'\}'.$$

VI Оценивание нелинейной регрессии среднего

1 Нелинейность по отношению к регрессорам

Пусть условное среднее $E[y|x] = g(x, \beta)$ для некоторой нелинейной функции $g(\cdot, \cdot)$. В этом случае мы имеем дело с нелинейной моделью. Тем не менее, существуют случаи нелинейностей, сводящиеся к линейному случаю с помощью трансформаций.

Пусть $g(x, \beta)$ нелинейна по регрессорам x и линейна по параметрам β, тогда можно выполнить такую трансформацию $x \to z$, что $E[y|z] = z'\beta$.

Пример 1. Пусть условное среднее выражается нелинейной функцией от регрессоров следующего вида:

$$g(x, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2.$$

Тогда подходящей трансформацией будет:

$$z = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)'.$$

Пример 2. Условное среднее выражается нелинейной функцией регрессоров следующего вида:

$$g(x, \beta) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p.$$

Соответствующая трансформация регрессоров:

$$z = (1, x, \ldots, x^p).$$

Необходимо отметить о сложности в интерпретации коэффициентов. Маржинальное влияние регрессора x есть

$$\frac{\partial g(x, \beta)}{\partial x} = \beta_1 + 2\beta_2 x + \cdots + p\beta_p x^{p-1}.$$

Неясно, какое x подставлять в данную формулу, чтобы получить численное значение. Можно оценить в каком-то конкретном x, которое определяется из контекста задачи, или использовать среднее значение $\overline{x}$, или же оценить в средних значения трансформированных регрессоров $\overline{x}, \overline{x^2}, \ldots, \overline{x^{p-1}}$. В любом случае коэффициенты $\beta_1, \beta_2, \ldots, \beta_p$ сами по себе не имеют экономического смысла. Имеет смысл только их определённые комбинации.

"Линейными по-существу" моделями называются такие, которые, несмотря на обманчивую нелинейность, можно преобразовать к линейному виду. Рассмотрим такой пример:

$$y = AK^\alpha L^{1-\alpha} \exp(e), \quad E[e|A, K, L] = 0.$$

Здесь логарифмическая трансформация модели сводит ее к линейному случаю:

$$E[\log Y \,|\, \log A, \log K, \log L] = \log A + \alpha \log K + (1 - \alpha) \log L.$$

2 Нелинейные регрессионные модели

В данной главе мы рассмотрим нелинейные модели, которые не приводятся к линейным, т.е.

$$E[y|x] = g(x'\beta) \neq z'\beta$$

для любой функции $z(x)$.

Примеры.

- $g(x,\beta) = \beta_1 + \beta_2 \dfrac{x}{1+\beta_3 x}$
- $g(x,\beta) = \beta_1 + \beta_2 e^{\beta_3 x}$
- $g(x,\beta) = (\beta_1 + \beta_2 x_1)1[x_2 \leq \beta_3] + (\beta_4 + \beta_5 x_1)1[x_2 > \beta_3]$

Пусть функция $g(x,\beta)$ дифференцируема по обоим аргументам.

Определение. Величина $\dfrac{\partial g(x,\beta)}{\partial \beta'} = g_\beta(x,\beta)$ называется *квазирегрессором.*

В обычной линейной регрессии

$$g(x,\beta) = x'\beta \quad \Rightarrow \quad g_\beta(x) = x,$$

т.е. квазирегрессор не зависит от параметра β. Однако в общем случае квазирегрессоры зависят от параметров модели. Этот факт усложняет определённые этапы оценивания и инференции.

3 Оценивание нелинейным методом наименьших квадратов

Нам известно, что параметр β есть решение минимизационной задачи

$$\beta = \arg\min_b E[(y - g(x,b))^2].$$

Используя принцип аналогий, получим оценку *нелинейного метода наименьших квадратов (НМНК)*:

$$\beta = \arg\min_b \frac{1}{n}\sum_{i=1}^n (y_i - g(x_i,b))^2.$$

Условие первого порядка есть

$$\frac{1}{n}\sum_{i=1}^n (y_i - g(x_i,\widehat{\beta}))g_\beta(x_i,\widehat{\beta}) = 0.$$

Ясно, что явное аналитическое выражение для $\widehat{\beta}$ получить в общем случае невозможно, поэтому для нахождения НМНК-оценок пользуются численными методами.

Получение НМНК-оценки методом концентрации. Одним из численных методов получения НМНК-оценки является метод концентрации. Разделим параметры задачи на две группы, удовлетворяющие условиям

$$\beta = (\gamma_1', \gamma_2')', \quad g(x, \beta) = \gamma_1' x(\gamma_2).$$

Грубо говоря, условное среднее линейно по параметрам γ_1 и нелинейно по параметрам γ_2. Кроме того, предполагается, что число параметров γ_2 невелико, чтобы можно было быстро бегать по их сетке.

Пример. В качестве примера приведем следующую модель:

$$g(x, \beta) = \beta_1 + \beta_2 e^{\beta_3 x}.$$

Тогда соответствующее разделение параметров следующее:

$$\gamma_1 = (\beta_1, \beta_2)', \quad \gamma_2 = \beta_3 \quad \Rightarrow \quad x(\gamma_2) = (1, e^{\beta_3 x})'.$$

В подобных случаях используется двухуровневая процедура оценивания параметров:

$$\widehat{\beta} = \arg\min_{\gamma_2} \left[\min_{\gamma_1} \frac{1}{n} \sum_{i=1}^{n} (y_i - \gamma_1' x_i(\gamma_2))^2 \right].$$

1. При фиксированном параметре γ_2 параметр γ_1 оценивается с помощью МНК:

$$\widehat{\gamma}_1(\gamma_2) = (X'(\gamma_2) X(\gamma_2))^{-1} X'(\gamma_2) Y, \quad X(\gamma_2) = (x_1(\gamma_2), \ldots, x_2(\gamma_2))'.$$

2. Численно решается оптимизационная задача

$$\widehat{\gamma}_2 = \arg\min_{\gamma_2} \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{\gamma}_1' x_i(\gamma_2))^2 \right].$$

   Поскольку размерность γ_2 маленькая, то оптимум легко находится на сетке.

Приведем алгоритм метода концентрации.

- Для параметра γ_2 на некотором интервале $[\underline{\gamma}_2, \overline{\gamma}_2]$ строится сетка.
- Для каждого γ_2 на этой сетке оценивается $\widehat{\gamma}_1(\gamma_2)$ с помощью МНК и вычисляется целевое значение $\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{\gamma}_1(\gamma_2)' x_i(\gamma_2))^2$.
- Из всех значений γ_2 на сетке выбирается то, для которого целевое значение наименьшее.

- Если необходимо, в окрестности полученного значения γ_2 строится более мелкая сетка, и процедура повторяется.

Получение НМНК-оценки методом линеаризации. Другим численным методом получения НМНК-оценки является линеаризация условия первого порядка. Допустим, что $\widehat{\beta}_1$ – начальное предположение о численном значении оцениваемых параметров. Тогда с помощью линеаризации предлагается итеративная процедура перехода $\widehat{\beta}_j \to \widehat{\beta}_{j+1}$. Эта процедура продолжается до тех пор, пока для достаточно малого ε не будет выполнено условие $|\widehat{\beta}_{j+1}-\widehat{\beta}_j| < \varepsilon$. Более детально: линеаризованное условие первого порядка для НМНК-оценки есть

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i,\widehat{\beta}_j) - g_\beta(x_i,\widehat{\beta}_j)(\widehat{\beta}_{j+1}-\widehat{\beta}_j))g_\beta(x_i,\widehat{\beta}_j) \approx 0.$$

Вводя обозначение

$$d_j = \left(\sum_{i=1}^{n} g_\beta(x_i,\widehat{\beta}_j)g_\beta(x_i,\widehat{\beta}_j)'\right)^{-1}\sum_{i=1}^{n} g_\beta(x_i,\widehat{\beta}_j)(y_i - g(x_i,\widehat{\beta}_j)),$$

имеем итеративную процедуру в виде

$$\widehat{\beta}_{j+1} = \widehat{\beta}_j + d_j.$$

Если d_j слишком велико (процедура не сходится), то выбирается некоторое $\lambda_j \in [0,1]$, такое, чтобы целевая функция была минимальной, а процедура модифицируется как

$$\widehat{\beta}_{j+1} = \widehat{\beta}_j + \lambda_j d_j.$$

4 Асимптотические свойства НМНК-оценки

Определение. Говорят, что задача удовлетворяет условию *идентификации*, если $b=\beta$ тогда и только тогда, когда $g(x,\beta)=g(x,b)$ с вероятностью 1.

Если это условие выполнено, то ввиду тожества

$$E[(y-g(x,b))^2] = E[(y-g(x,\beta))^2] + E[(g(x,\beta)-g(x,b))^2]$$

минимизатор левой части равен истинному значению параметра β и определён однозначно.

Примеры.

- Рассмотрим линейную модель. Пусть матрица $Q_{xx}=E[xx']$ невырождена. Тогда при $\beta \neq b$ выполняется соотношение:

  $$E[(x'\beta - x'b)^2] = (\beta-b)'Q_{xx}(\beta-b) > 0.$$

  Следовательно, $x'\beta \neq x'b$ с вероятностью 1.

- Рассмотрим теперь пример, где нет идентификации:

$$g(x,\beta) = \beta_1 + \beta_2 e^{\beta_4+\beta_3 x} = \beta_1 + e^{\log\beta_2+\beta_4+\beta_3 x}.$$

Идентифицировать параметры β_2 и β_4 одновременно невозможно.

Определение. Последовательность случайных функций $\{z_i(\theta)\}_{i=1}^n$ удовлетворяет равномерному закону больших чисел (РЗБЧ), если

$$\sup_\theta \|\frac{1}{n}\sum_{i=1}^n z_i(\theta) - p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\| \xrightarrow{p} 0.$$

Лемма. Если последовательность $\{z_i(\theta)\}_{i=1}^n$ удовлетворяет РЗБЧ и $\widehat{\theta}_n \xrightarrow{p} \theta$, то

$$\frac{1}{n}\sum_{i=1}^n z_i(\widehat{\theta}_n) \xrightarrow{p} p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta).$$

Доказательство. Запишем последовательность неравенств, воспользовавшись РЗБЧ и теоремой Манна-Вальда:

$$\begin{aligned}
& \left\|\frac{1}{n}\sum_{i=1}^n z_i(\widehat{\theta}_n) - p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\right\| \le \\
\le\; & \left\|\frac{1}{n}\sum_{i=1}^n z_i(\widehat{\theta}_n) - p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\Big|_{\widehat{\theta}_n}\right\| + \left\|p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\Big|_{\widehat{\theta}_n} - p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\right\| \le \\
\le\; & \sup_\theta\left\|\frac{1}{n}\sum_{i=1}^n z_i(\theta) - p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\right\| + \left\|p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\Big|_{\widehat{\theta}_n} - p\lim\frac{1}{n}\sum_{i=1}^n z_i(\theta)\right\| \\
& \xrightarrow{p} 0.
\end{aligned}$$

Как следствие, при выполнении РЗБЧ для соответствующих слагаемых имеем состоятельность следующих оценок:

$$\widehat{Q}_{e^2xx} = \frac{1}{n}\sum_{i=1}^n x_i x_i' \widehat{e}_i^2 \xrightarrow{p} Q_{e^2xx}, \quad \widehat{Q}_{gg} = \frac{1}{n}\sum_{i=1}^n g_\beta(x_i,\widehat{\beta}) g_\beta(x_i,\widehat{\beta})' \xrightarrow{p} Q_{gg},$$

где $Q_{gg} = E[g_\beta(x,\beta)g_\beta(x,\beta)']$.

Теорема. Пусть выполнены следующие требования:

1. Выполнено условие идентификации;

2. Функция $g(x,\beta)$ дважды непрерывно дифференцируема по b;

3. Для следующих последовательностей выполняется РЗБЧ:

$$(y_i - g(x_i,\beta))^2, \quad g_\beta(x_i,\beta)g_\beta(x_i,\beta)', \quad (y_i - g(x_i,\beta))\frac{\partial g_\beta(x_i,\beta)}{\partial\beta'};$$

4. Матрица $Q_{gg} = E[g_\beta(x,\beta)g_\beta(x,\beta)']$ невырождена;

5. Матрица $Q_{e^2gg} = E[g_\beta(x,\beta)g_\beta(x,\beta)'e^2]$ существует.

Тогда НМНК-оценка состоятельна и асимптотически нормальна:

$$\widehat{\beta} \xrightarrow{p} \beta, \quad \sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, Q_{gg}^{-1}Q_{e^2gg}Q_{gg}^{-1}).$$

Доказательство.

1. **Состоятельность:** Для любого $\varepsilon > 0$ с вероятностью, стремящейся к 1 при $n \to \infty$, имеем:

   $$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i, \widehat{\beta}))^2 < \frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i, \beta))^2 + \frac{\varepsilon}{3},$$

   так как оценка $\widehat{\beta}$ минимизирует $\dfrac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i, b))^2$. Поскольку РЗБЧ выполняется для $(y_i - g(x_i, \beta))^2$,

   $$E[(y_i - g(x_i, \widehat{\beta}))^2] < \frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i, \widehat{\beta}))^2 + \frac{\varepsilon}{3}.$$

   Аналогично,
   $$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i, \beta))^2 < E[(y_i - g(x_i, \beta))^2] + \frac{\varepsilon}{3}.$$

   Суммируя эти три неравенства, получаем:

   $$E[(y - g(x_i, \widehat{\beta}))^2] < E[(y - g(x_i, \beta))^2] + \varepsilon.$$

   Теперь определим ε. Для этого выберем открытую окрестность β, $N(\beta)$. Поскольку β решает задачу на минимум, должно быть выполнено следующее соотношение:
   $$\inf_{b \in N(\beta)^c} E[(y - g(x, b))^2] > E[(y - g(x, \beta))].$$

   Выберем следующее ε:

   $$\varepsilon = \inf_{b \in N(\beta)^c} E[(y - g(x, b))^2] - E[(y - g(x, \beta))],$$

   тогда выполнено следующее соотношение:

   $$E[(y - g(x, \widehat{\beta}))^2] < \inf_{b \in N(\beta)^c} E[(y - g(x, b))^2],$$

   что означает, что $\widehat{\beta} \in N(\beta)$. Следовательно, $\widehat{\beta} \xrightarrow{p} \beta$.

2. **Асимптотическая нормальность:** Разложим условие первого порядка в ряд Тэйлора вокруг β:

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i,\beta))g_\beta(x_i,\beta) + \\
+ \quad & \frac{1}{n}\sum_{i=1}^{n}\left[(y_i - g(x_i,\widehat{\beta}))\frac{\partial g_\beta(x_i,\widehat{\beta})}{\partial\beta'} - g_\beta(x_i,\widetilde{\beta})g_\beta(x_i,\widetilde{\beta})'\right](\widehat{\beta}-\beta) = 0,
\end{aligned}
$$

где β лежит между β и $\widetilde{\beta}$ покомпонентно. Следовательно,

$$
\begin{aligned}
\sqrt{n}(\widehat{\beta}-\beta) \quad &= \quad \left\{\frac{1}{n}\sum_{i=1}^{n}\left[(y_i - g(x_i,\widehat{\beta}))\frac{\partial g_\beta(x_i,\widehat{\beta})}{\partial\beta'} - g_\beta(x_i,\widetilde{\beta})g_\beta(x_i,\widetilde{\beta})'\right]\right\}^{-1} \times \\
&\times \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - g(x_i,\beta))g_\beta(x_i,\beta) \xrightarrow{d} \\
&\xrightarrow{d} \quad -\left\{E\left[(y_i - g(x_i,\beta))\frac{\partial g_\beta(x,\beta)}{\partial\beta'} - g_\beta(x,\beta)g_\beta(x,\beta)'\right]\right\}^{-1}\mathcal{N}\left(0, Q_{e^2gg}\right) \\
&= \quad \mathcal{N}\left(Q_{gg}^{-1}Q_{e^2qq}Q_{gg}^{-1}\right).
\end{aligned}
$$

Теорема доказана.

Рассмотрим специальный случай условной гомоскедастичности:

$$E[e^2|x] = \sigma^2 = \text{const}.$$

Как и для линейной модели, имеет место упрощение:

$$Q_{e^2gg} = \sigma^2 Q_{gg} \quad \Rightarrow \quad \sqrt{n}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q_{gg}^{-1}).$$

5 Асимптотическая эффективность и ВНМНК-оценка

НМНК-оценку можно рассматривать как аналоговую оценку, полученную из условия $E[eg_\beta(x,\beta)] = 0$. Можно построить другую аналоговую оценку, несколько изменив условие нескоррелированности:

$$E\left[e\frac{g_\beta(x,\beta)}{\sigma^2(x)}\right] = 0.$$

Это условие следует из регрессионного предположения. Согласно принципу аналогий

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i,\widetilde{\beta}))\frac{g_\beta(x_i,\widetilde{\beta})}{\sigma^2(x_i)} = 0.$$

Решение $\widetilde{\beta}$, полученное из этого уравнения, является оценкой взвешенного нелинейного метода наименьших квадратов (ВНМНК-оценкой). Она решает минимизационную задачу

$$\widetilde{\beta} = \arg\min_{b} \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - g(x_i, b))}{\sigma^2(x_i)},$$

состоятельна и асимптотически нормальна:

$$\widetilde{\beta} \xrightarrow{p} \beta, \quad \sqrt{n}(\widetilde{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, Q_{\frac{gg}{\sigma^2}}^{-1}),$$

$$Q_{\frac{gg}{\sigma^2}} = E\left[\frac{g_\beta(x, \beta) g_\beta(x, \beta)'}{\sigma^2(x)}\right].$$

В условиях условной гетероскедастичности ВНМНК-оценка более асимптотически эффективна по сравнению с НМНК, точно так же как ОМНК по сравнению с МНК для линейной регрессии. Можно ещё утверждать, что ВНМНК-оценка $\widetilde{\beta}$ является асимптотически эффективной в классе оценок, удовлетворяющих условию

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - g(x_i, \widehat{\beta}_{IV})) z_i = 0,$$

где z_i – произвольная функция от x_i, имеющая ту же размерность $k \times 1$.

6 Приложение: модель бинарного выбора

Рассмотрим следующую нелинейную модель:

$$y_i = \begin{cases} 1 & x_i'\beta + e_i \geq 0, \\ 0 & \text{иначе}, \end{cases} \qquad e_i | x_i \sim \mathcal{N}(0, 1).$$

Найдём форму регрессии:

$$E[y|x] = P\{x'\beta + e \geq 0 | x\} = P\{e \geq -x'\beta | x\} = \Phi(x'\beta).$$

Видно, что регрессия нелинейная. НМНК-оценка в этом случае есть

$$\widehat{\beta} = \arg\min_{b} \frac{1}{n} \sum_{i=1}^{n} (y_i - \Phi(x_i' b))^2$$

с асимптотическими свойствами

$$\widehat{\beta} \xrightarrow{p} \beta, \quad \sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, Q_{gg}^{-1} Q_{e^2 gg} Q_{gg}^{-1}),$$

где

$$g_\beta(x, \beta) = f(x'\beta)x, \quad Q_{gg} = E[f(x'\beta)^2 x x'], \quad Q_{e^2 gg} = E[f(x'\beta)^2 (y - \Phi(x'\beta))^2 x x'].$$

Асимптотически эффективная ВНМНК-оценка есть

$$\widetilde{\beta} = \arg\min_b \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \Phi(x_i'b))^2}{\Phi(x_i'\widehat{\beta})(1 - \Phi(x_i'\widehat{\beta}))},$$

ибо

$$\sigma^2(x) = Var[y|x] = \Phi(x'\beta)(1 - \Phi(x'\beta)) \neq \text{const}.$$

Выведем её асимптотические свойства:

$$\widetilde{\beta} \xrightarrow{p} \beta, \quad \sqrt{n}(\widetilde{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \left(E\left[\frac{f(x'\beta)^2 xx'}{\Phi(x'\beta)(1 - \Phi(x'\beta))}\right]\right)^{-1}\right).$$

7 Инференция при неидентифицированности некоторых параметров при нулевой гипотезе

В нелинейных моделях может сложиться особая ситуация, когда тестирование статистических гипотез нестандартно. Рассмотрим два примера.

Пример 1. Регрессия с гладкими порогами:

$$y = (\beta_1 + \beta_2 x) + (\beta_3 + \beta_4 x)\frac{1}{1 + e^{x - \beta_5}} + e, \quad E[e|x] = 0.$$

Если нулевая гипотеза состоит в том, что $\beta_3 = \beta_4 = 0$ (т.е. тестируется линейность модели), то при этой нулевой гипотезе параметр β_5 неидентифицируем.

Пример 2. Рассмотрим следующую разновидность ARCH-M модели:

$$y_t = \beta_0 + x_t'\beta_1 + \gamma\sigma_t^2 + e_t, \quad E[e_t|I_{t-1}] = 0, \quad E[e_t^2|I_{t-1}] = \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2.$$

Если нулевая гипотеза состоит в отсутствии ARCH эффекта, т.е. $H_0 : \alpha_1 = 0$, то при нулевой гипотезе параметр γ неидентифицируем, так как условная дисперсия постоянна и её влияние поглощается свободным членом β_0.

В таких ситуациях стандартная тестовая статистика (например, t или Вальдовская) асимптотически распределена не так, как мы привыкли, т.е. не как стандартно нормальная или хи-квадрат случайная величина. Вот как обычно решается подобная проблема. Пусть $\beta = (\beta_1', \beta_2')'$, где β_1 идентифицируется при нулевой гипотезе, а β_2 – нет. Построим Вальдовскую статистику $W(\beta_2)$ для всех возможных значений β_2. Тогда "суп-Вальдовская" статистика

$$\sup W = \sup_{\beta_2} W(\beta_2)$$

сходится к некоторому нестандартному распределению, которое получают с помощью симуляций.

Помимо приведённых выше, примерами являются тестирование на линейность самовозбуждающихся пороговых авторегрессий и тестирование на отсутствия структурных сдвигов.