ЭКОНОМЕТРИКА ДЛЯ ПОДГОТОВЛЕННЫХ

Курс лекций

Станислав Анатольев

Российская Экономическая Школа

КЛ/2003/008

Москва

2003

Содержание

Введение

1 Описание курса

Настоящие лекции – продолжение курса «Эконометрика для продолжающих». Здесь мы будем заниматься в основном нелинейными моделями, а также панельными структурами. Будет преобладать анализ оценок в неявной форме, являющихся решениями определённых оптимизационных задач. Естественно, анализ таких оценок сложнее, чем тех, которые задаются в явной форме. Нас по-прежнему будут интересовать асимптотические свойства, также иногда мы будем обращаться к бутстраповскому подходу. Заключительная часть курса посвящена анализу панельных данных, когда протяжённость данных двоякая, это и кросс-секции, и временные ряды. Вновь акцент делается скорее на концептуальную составляющую, нежели на математическую сложность, хотя последняя нередко неизбежна. Домашние задания по курсу содержат как теоретические задачи, так и практические задания, подразумевающие использование пакета GAUSS. Задания служат важной компонентой обучающего процесса, в которой часто будут встречаться теоретические и эмпирические примеры.

Выражаю благодарность Людмиле Солнцевой за техническую помощь. Дарья Нехороших осуществила подготовку черновой версии конспектов. Георгий Карташов исправил множество опечаток. 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, 2005

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. Hamilton, James *Time Series Analysis*, Princeton University Press, 1994

6. Baltagi, Badi *Econometric Analysis of Panel Data*, John Wiley & Sons, 1995, 2001

7. Newey, Whitney and McFadden, Daniel *Large sample estimation and hypothesis testing*, in: *Handbook of Econometrics*, volume 4, Elsevier Science, 1994

I Экстремальное оценивание и экстремальные оценки

Экстремальное оценивание – это в каком-то смысле обобщение знакомого нам нелинейного метода наименьших квадратов (НМНК). НМНК – это один из важнейших частных случаев; другим является метод максимального правдоподобия (ММП). Сначала мы рассмотрим общую теорию, а затем – эти два важных частных случая. Рассмотрим следующую постановку:

$$\theta = \arg\max_{q \in \Theta} E[h(z, q)],$$

где θ – истинное значение параметра q, вектора размерности $k \times 1$, Θ – множество допустимых значений параметра, h – известная функция, а z – представитель наблюдаемых данных.

Почему такая постановка интересна? Некоторые оценки действительно строятся на основе подобных соотношений. Например, мы нашли такое соотношение ранее, когда речь шла о нелинейном методе наименьших квадратов. А именно, если

$$y = g(x, \beta) + e, \quad E[e|x] = 0,$$

то определив $z = (x', y)'$, $q = b$, $\theta = \beta$, $\Theta = \mathbb{R}$ и $h(z, q) = -(y - g(x, b))^2$, мы имеем данную постановку.

Применим принцип аналогий (напоминание: если имеется какое-то популяционное соотношение, идентифицируюшее истинный параметр, то, создав аналогичное соотношение для выборки, можно получить оценку для этого параметра). Пусть имеется случайная выборка $z_1, \ldots, z_n$. Применение принципа аналогий дает оценку

$$\widehat{\theta} = \arg\max_{q \in \Theta} \frac{1}{n} \sum_{i=1}^{n} h(z_i, q).$$

Множитель $1/n$ можно удалить, т.к. он не влияет на максимизацию. Полученная оценка и называется *экстремальной оценкой* параметра θ. Название «экстремальная» происходит от того факта, что оценка решает задачу на экстремум.

Нас интересуют асимптотические свойства экстремальных оценок. Увы, мечтать о благоприятных свойствах в конечных выборках не приходится из-за нелинейности h. Ожидаемые асимптотические свойства, как обычно, – это состоятельность и асимптотическая нормальность. Действительно, при благоприятных условиях,

- $\widehat{\theta} \xrightarrow{p} \theta$, т.е. оценка состоятельна;
- $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V_{\widehat{\theta}})$, т.е. оценка асимптотически нормальна;

- Её асимптотическая дисперсионная матрица, как обычно, имеет «сэндвичную» структуру
$$V_{\widehat{\theta}} = (E[h_{\theta\theta}])^{-1}E[h_\theta h_\theta'](E[h_{\theta\theta}])^{-1},$$
где $h_\theta = \dfrac{\partial h(z,\theta)}{\partial q}$ – вектор размерности $k \times 1$, а $h_{\theta\theta} = \dfrac{\partial^2 h(z,\theta)}{\partial q\,\partial q'}$ – матрица размерности $k \times k$.

В частном случае НМНК,

$$\widehat{\beta} = \arg\max_{b\in\mathbb{R}} - \sum_{i=1}^{n}(y_i - g(x_i, b))^2.$$

Если избавиться от "$-$" и max поменять на min, получится знакомая НМНК-оценка. Далее,
$$h_\theta = \frac{\partial}{\partial b}(-(y - g(x,b))^2|_{b=\beta} = 2(y - g(x,\beta))g_\beta(x,\beta),$$
где $y-g(x,\beta)$ – регрессионная ошибка, а $g_\beta(x,\beta)$ – квазирегрессор. Матрица, стоящая в середине $V_{\widehat{\theta}}$, равна $E[h_\theta h_\theta'] = 4E[e^2 g_\beta g_\beta']$. Для вычисления боковых матриц нам необходимо

$$h_{\theta\theta} = \frac{\partial}{\partial b'}[2(y - g(x,b)g_b(x,b))]_{b=\beta} = -2g_\beta(x,\beta)g_\beta(x,\beta)' + 2(y - g(x,\beta))\frac{\partial^2 g(x,\beta)}{\partial b\,\partial b'}.$$

Возьмём математическое ожидание и учтём, что $E[e|x] = 0$, тогда

$$V_{\widehat{\theta}} = (E[g_\beta g_\beta'])^{-1}E[g_\beta g_\beta' e^2](E[g_\beta g_\beta'])^{-1}.$$

Здесь $E[g_\beta g_\beta']$ – аналог Q_{xx} для линейной модели, а $E[g_\beta g_\beta' e^2]$ – аналог Q_{e^2xx}.

Вернёмся к общему случаю. Выпишем условия первого порядка и линеаризуем их, т.е. разложим в ряд Тэйлора вокруг истинного θ:

$$0 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial h(z_i,\widehat{\theta})}{\partial q} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial h(z_i,\theta)}{\partial q} + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 h(z_i,\widetilde{\theta})}{\partial q\,\partial q'}\sqrt{n}(\widehat{\theta} - \theta),$$

где $\widetilde{\theta}$ лежит покомпонентно между θ и $\widehat{\theta}$. Тогда

$$\sqrt{n}(\widehat{\theta} - \theta) = -\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 h(z_i,\widetilde{\theta})}{\partial q\,\partial q'}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial h(z_i,\theta)}{\partial q}.$$

Далее, при наличии РЗБЧ, ЦПТ и состоятельности $\widetilde{\theta}$, имеем

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 h(z_i,\widetilde{\theta})}{\partial q\,\partial q'} \xrightarrow{p} E[h_{\theta\theta}], \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial h(z_i,\theta)}{\partial q} \xrightarrow{d} \mathcal{N}(0, E[h_\theta h_\theta']).$$

Подставив эти соотношения в предыдущее равенство, получим искомый вид дисперсионной матрицы $V_{\widehat{\theta}}$.

Что же такое *«благоприятные» условия*? Если строго подходить к этому вопросу, то это список условий, большинство из которых – технические, которые обеспечивают выполнение теорем, которые мы постоянно используем (ЗБЧ, РЗБЧ, существование различных моментов, гладкость h по параметру). Но для нас более интересны следующие «идеологические» условия:

1. *Условие идентификации.* Это главное «идеологическое» условие, связанное с единственностью решения задачи на экстремум. Его смысл в том, что нет других значений параметра, также максимизирующих целевую функцию, вообще (глобальная идентификация) или в окрестности истинного параметра (локальная идентификация).

   - Условие локальной идентификации выполнено, когда матрица ожиданий вторых производных $E[h_{\theta\theta}]$ имеет полный ранг;
   - Условие глобальной идентификации выполнено, если решение популяционной задачи на экстремум единственно.

2. Множество Θ компактно. Нам больше важна ограниченность Θ, т.к. неограниченность может привести к несуществованию максимума целевой функции в ряде случаев. Например, она может оказаться возрастающей до бесконечности. Вопрос о том, почему необходима замкнутость Θ, пока отложим.

3. Истинное θ находится не на границе Θ. Если θ окажется на границе Θ, вряд ли возможна асимптотическая нормальность оценки вокруг истинного параметра, ибо нормальность подразумевает возможность находиться как справа, так и слева от него.

Наряду с НМНК, важным частным случаем экстремального оценивания является *метод максимального правдоподобия (ММП).*

II Метод максимального правдоподобия

1 Оценка максимального правдоподобия и её мотивация

Постановка задачи. Пусть известно распределение данных с точностью до конечномерного параметра θ размерности $k \times 1$: $z \sim f(z|\theta)$, где $f(z|\theta)$ – плотность (для дискретных распределений – вероятности). Оценка максимального правдоподобия определяется как

$$\widehat{\theta} = \arg\max_{q \in \Theta} \mathcal{L}(z_1, \dots, z_n, q),$$

где $\mathcal{L}(z_1, \ldots, z_n, \theta)$ – *функция правдоподобия*, совместная плотность (в дискретном случае – совместная вероятностная масса) наблюдений.

В случае, когда выборка $z_1, \ldots, z_n$ случайная, $\mathcal{L}(z_1, \ldots, z_n, q)$ распадается на произведение плотностей:

$$\widehat{\theta} = \arg\max_{q\in\Theta} \prod_{i=1}^{n} f(z_i|q) = \arg\max_{q\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log f(z_i|q),$$

где $\sum_{i=1}^{n} \log f(z_i|q) = l(z_1, \ldots, z_n, q)$ – логарифмическая функция правдоподобия.

Для состоятельного оценивания θ необходимо, чтобы оцениваемый параметр решал аналогичную задачу в популяции:

$$\theta = \arg\max_{q\in\Theta} E[\log f(z|q)].$$

Убедимся, что это имеет место.

Лемма (Информационное неравенство). Предположим, что $f(z|q) \neq f(z|\theta)$ с вероятностью 1 $\forall\ q \neq \theta$. Тогда $E[\log f(z|q)]$ максимизируется единственным образом в истинном параметре θ.

Доказательство. Воспользуемся неравенствои Йенсена: если $g(x)$ строго выпукла вверх, то $E[g(X)] < g(E[X])$, где X – случайная величина, не являющаяся константой. Пусть q – произвольное значение параметра, а θ – его истинное значение, и рассмотрим следующую разность:

$$E[\log f(z|q)] - E[\log f(z|\theta)] = E\left[\log\frac{f(z|q)}{f(z|\theta)}\right] < \log E\left[\frac{f(z|q)}{f(z|\theta)}\right] =$$

$$= \log\int\left(\frac{f(z|q)}{f(z|\theta)} f(z|\theta)\,dz\right) = \log 1 = 0.$$

Определение. *Скор-функцией* называется

$$s(z, q) \equiv \frac{\partial \log f(z|q)}{\partial q}.$$

Лемма (Нулевое ожидаемое скор). Математическое ожидание скор-функции в истинном параметре равно 0:

$$E[s(z, \theta)] = 0.$$

Доказательство. Достаточно рассмотреть условие первого порядка для популяционной задачи $\theta = \arg\max_{q\in\Theta} E[\log f(z|q)]$.

Идея Голдбергера: мы можем применить принцип аналогий к правилу нулевого ожидаемого скор. Тогда у нас будет k уравнений для оценок k параметров

$$\frac{1}{n}\sum_{i=1}^{n} s(z_i, \widehat{\theta}_{ZES}) = 0.$$

Очевидно, оценка $\widehat{\theta}_{ZES}$ совпадает с ММП-оценкой, но, в отличие от последней, может не определяться единственным образом, даже если условие идентификации выполнено (упражнение: почему?).

Следующая лемма является ключевой в вопросе асимптотической эффективности оценок максимального правдоподобия.

Лемма (Информационное равенство). Введём понятие информационной матрицы:

$$\mathcal{I}(q) \equiv -E\left[\frac{\partial s(z,q)}{\partial q'}\right].$$

Тогда $\mathcal{I}(\theta) = E[s(z,\theta)s(z,\theta)']$.

Доказательство. Рассмотрим тождество

$$1 = \int f(z|q)\,dz \quad \forall\ q \in \Theta.$$

Здесь $f(z|q)$ – любая плотность, необязательно истинная. Дифференцируем по q:

$$0 = \int \frac{\partial f(z|q)}{\partial q}\,dz = \int \frac{\partial \log f(z|q)}{\partial q} f(z|q)\,dz.$$

Дифференцируем ещё раз по q':

$$0 = \int \frac{\partial^2 \log f(z|q)}{\partial q\,\partial q'} f(z|q)\,dz + \int \frac{\partial \log f(z|q)}{\partial q}\frac{\partial \log f(z|q)}{\partial q'} f(z|q)\,dz.$$

Оценим в $q = \theta$:

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

2 Асимптотические свойства оценок максимального правдоподобия

При благоприятных условиях,

- ММП-оценка состоятельна: $\widehat{\theta} \xrightarrow{p} \theta$;
- ММП-оценка асимптотически нормальна: $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V_{\widehat{\theta}})$;
- Асимптотическая дисперсионная матрица ММП-оценки равна

$$V_{\widehat{\theta}} = (E[h_{\theta\theta}])^{-1}E[h_\theta h_\theta'](E[h_{\theta\theta}])^{-1} = (-\mathcal{I}(\theta))^{-1}\mathcal{I}(\theta)(-\mathcal{I}(\theta))^{-1} = (\mathcal{I}(\theta))^{-1}.$$

  То, что «сэндвич» сократился, является показателем эффективности, что мы позже рассмотрим подробнее.

Итак, у нас есть два способа оценивания асимптотической дипересионной матрицы:

$$\widehat{V}_{\widehat{\theta}}^{(1)} = \left(-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial s(z_i, \widehat{\theta})}{\partial q'} \right)^{-1},$$

если пользоваться определением, и

$$\widehat{V}_{\widehat{\theta}}^{(2)} = \left(\frac{1}{n} \sum_{i=1}^{n} s(z_i, \widehat{\theta})\, s(z_i, \widehat{\theta})' \right)^{-1},$$

если пользоваться леммой об информационном равенстве. Обе оценки состоятельны, если выполнены все необходимые условия. Вторая оценка проще, ибо не нужно считать вторые производные. Нет даже проблем с положительной определенностью, т.к. вторая оценка положительно определена по построению, а в первой положительную определенность гарантируют условия второго порядка (при оптимизации мы получаем, что матрица вторых производных отрицательно определена в $\widehat{\theta}$). Можно построить еще одну оценку:

$$\widehat{V}_{\widehat{\theta}}^{(3)} = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial s(z_i, \widehat{\theta})}{\partial q'} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} s(z_i, \widehat{\theta})\, s(z_i, \widehat{\theta})' \right) \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial s(z_i, \widehat{\theta})}{\partial q'} \right)^{-1}.$$

Эта оценка более робастна к невыполнению некоторых условий. Например, бывают «невинные» ошибки спецификации плотности, при которых состоятельность ММП-оценок сохраняется. В такой ситуации полезно строить более робастную оценку асимптотической дипересионной матрицы, ибо только она одна из трёх оказывается состоятельной.

3 Когда теория максимального правдоподобия не работает?

Обсудим случаи нарушения некоторых предположений, которые мы неявно использовали, но явно о них не говорили.

1. Носитель случайной величины z не должен зависеть от значения параметра. Мы использовали это предположение, когда меняли порядок дифференцирования и интегрирования.

   Контрпример. Пусть $z \sim U[0, \theta]$, т.е.

$$f(z|\theta) = \begin{cases} 0, & z < 0, \\ 1/\theta, & 0 \leqslant z < \theta, \\ 0, & z \geqslant \theta, \end{cases}$$

где θ – истинный параметр, а $z_1, \ldots, z_n$ – выборка. Рассмотрим

$$\mathcal{L}(z_1, ...z_n, q) = \left(\frac{1}{q}\right)^n \prod_{i=1}^{n} \mathbb{1}[0 \leqslant z_i < q] \to \max_{0 \leqslant q}.$$

Тогда $\widehat{\theta} = \max\{z_1, \ldots, z_n\}$. Найдём кумулятивную функцию распределения для этой оценки и посмотрим, куда она сходится.

$$F_{\widehat{\theta}}(t) = Pr\{z_1 \leqslant t, \ldots, z_n \leqslant t\} = (F_z(t))^n =$$

$$= \begin{cases} 0, & t < 0 \\ (t/\theta)^n, & 0 \leqslant t < \theta \\ 1, & t \geqslant \theta \end{cases} \to \begin{cases} 0, & t < 0 \\ 1, & \theta \leqslant t \end{cases}$$

Следовательно, $\widehat{\theta} \xrightarrow{d} \theta$, а так как пределом является константа, то и $\widehat{\theta} \xrightarrow{p} \theta$, т.е. состоятельность имеется. А вот асимптотической нормальности нет:

$$F_{n(\widehat{\theta}-\theta)}(t) = Pr\{n(\widehat{\theta} - \theta) \leqslant t\} = Pr\left\{\widehat{\theta} \leqslant \theta + \frac{t}{n}\right\} = F_{\widehat{\theta}}(\theta + t/n) =$$

$$= \begin{cases} 0, & \theta + t/n < 0 \\ (1 + t/n\theta)^n, & 0 \leqslant \theta + t/n < \theta \\ 1, & \theta + t/n \geqslant \theta \end{cases} \to \begin{cases} e^{t/\theta}, & t < 0 \\ 1, & t \geqslant 0 \end{cases}$$

Итак, $n(\widehat{\theta} - \theta) \xrightarrow{d} -Exp(1/\theta)$. Следовательно, ММП-оценка асимптотически экспоненциально распределена, а скорость сходимости получилась больше обычной (оценка *сверхсостоятельна*).

2. Размерность k параметра θ должна быть фиксированной.

   Контрпример. Пусть

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_i \\ \mu_i \end{pmatrix}, \sigma^2 I_2\right), \quad i = 1, \ldots, n,$$

   т.е. у нас $2n$ наблюдений для оценивания $k = (n + 1)$-мерного параметра. Увы,

$$\widehat{\mu}_i = \frac{x_i + y_i}{2} \overset{p}{\nrightarrow} \mu_i, \quad \widehat{\sigma^2} = \frac{1}{4n} \sum_{i=1}^{n} (x_i - y_i)^2 \xrightarrow{p} \frac{\sigma^2}{2}.$$

3. Желательно, чтобы множество Θ было замкнутым.

   Контрпример. Рассмотрим смесь нормально распределённых величин:

$$z \sim \begin{cases} \mathcal{N}(\mu, \sigma^2) & \text{с вероятностью } \frac{1}{2}, \\ \mathcal{N}(0, 1) & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

Здесь
$$q = (m, s^2),\ \theta = (\mu, \sigma^2),\ \Theta = (-\infty, +\infty) \times (0, +\infty).$$

Проблема возникает из-за невключения нуля в допустимое множество:
$$f(z|m, s^2) = \frac{1}{2}\frac{1}{\sqrt{2\pi s^2}}e^{-(z-m)^2/2s^2} + \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-z^2/2},$$
тогда $\prod_{i=1}^n f(z_i|m, s^2) \to \max$ достигается выбором $\widehat{\mu} = z_1$, $\widehat{\sigma^2} \to 0$, если $f(z_1|\widehat{\mu}, \widehat{\sigma^2}) \geqslant f > 0$.

4 Асимптотическая эффективность ММП-оценок

Приведём иногда встречающееся утверждение об асимптотической эффективности ММП-оценок.

Неверный результат. Оценка максимального правдоподобия $\widehat{\theta}$ асимптотически эффективна в классе состоятельных асимптотически нормальных оценок.

Это неверно из-за существования так называемых *суперэффективных оценок.* Пусть оценка $\widehat{\theta}$, которую мы считаем эффективной, состоятельна и асимптотически нормальна, т.е.
$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V_{\widehat{\theta}}).$$
Построим следующую оценку $\widehat{\theta}^*$:
$$\widehat{\theta}^* = \begin{cases} 0, & \text{если } |\widehat{\theta}| < n^{-\frac{1}{4}}, \\ \widehat{\theta}, & \text{в противном случае.} \end{cases}$$
Тогда
$$\sqrt{n}(\widehat{\theta}^* - \theta) \xrightarrow{d} \begin{cases} 0, & \text{если} \quad \theta = 0 \\ \mathcal{N}(0, V_{\widehat{\theta}}), & \text{если} \quad \theta \neq 0. \end{cases}$$
Если истинный параметр равен 0, построенная оценка более эффективна, чем изначальная; в остальных случаях она настолько же эффективна. Вместе с тем такое улучшение эффективности нам неинтересно, т.к. улучшается оно только в частном, маловероятном и заранее непредсказуемом случае.

Верный результат. Пусть
$$\theta = \arg\max_{q\in\Theta} E[h(z, q)]$$
для некоторой $h(z, q) \neq \log f(z|q)$, и это верно, какое бы истинное $\theta \in \Theta$ ни было. Пусть оценки $\widehat{\theta}$ и
$$\widetilde{\theta} \equiv \arg\max_{q\in\Theta} \sum_{i=1}^n h(z_i, q)$$

состоятельны и асимптотически нормальны. Тогда оценка $\widehat{\theta}$ асимптотически по крайней мере не менее эффективна, чем $\widetilde{\theta}$.

Доказательство. Запишем условие первого порядка для θ:

$$E\left[\frac{\partial h(z,\theta)}{\partial q}\right] = 0 \quad \forall\, \theta \in \Theta.$$

Продифференцируем дважды по θ, переписав математическое ожидание в виде интеграла:

$$\int \frac{\partial h(z,\theta)}{\partial q} f(z|\theta)\, dz = 0,$$

$$\int \frac{\partial^2 h(z,\theta)}{\partial q\, \partial q'} f(z|\theta)\, dz + \int \frac{\partial h(z,\theta)}{\partial q} \frac{\partial \log f(z|\theta)}{\partial q'} f(z|\theta)\, dz = 0.$$

или в терминах математических ожиданий,

$$E[h_{\theta\theta}] + E[h_\theta s'] = 0,$$

где для краткости мы обозначили

$$h_\theta \equiv \frac{\partial h(z,\theta)}{\partial q}, \quad h_{\theta\theta} \equiv \frac{\partial^2 h(z,\theta)}{\partial q\, \partial q'}, \quad s \equiv \frac{\partial \log f(z,\theta)}{\partial q}.$$

Теперь рассмотрим разность матриц

$$V_{\widetilde{\theta}} - V_{\widehat{\theta}} = (E[h_{\theta\theta}])^{-1} E[h_\theta h_\theta'] (E[h_{\theta\theta}])^{-1} - (E[ss'])^{-1}.$$

Необходимо показать, что это – неотрицательно определенная матрица. Действительно,

$$V_{\widetilde{\theta}} - V_{\widehat{\theta}} = (E[h_\theta s'])^{-1} (E[h_\theta h_\theta'] - E[h_\theta s'](E[ss'])^{-1} E[s h_\theta'])(E[s h_\theta'])^{-1}.$$

Обозначим $u = h_\theta - E[h_\theta s'](E[ss'])^{-1} s$, что есть проекционная ошибка линейной проекции h_θ на s. Тогда

$$E[uu'] = E[h_\theta h_\theta'] - E[h_\theta s'](E[ss'])^{-1} E[s h_\theta'] - E[h_\theta s'](E[ss'])^{-1} E[s h_\theta'] +$$

$$+ E[h_\theta s'](E[ss'])^{-1} E[ss'] (E[ss'])^{-1} E[s h_\theta'] = E[h_\theta h_\theta'] - E[h_\theta s'](E[ss'])^{-1} E[s h_\theta'].$$

Получившаяся матрица положительно полуопределена по построению. Как следствие, оценка максимального правдоподобия в смысле асимптотической эффективности не может быть хуже любой другой экстремальной оценки.

Другими словами, ММП-оценка асимптотически эффективна в классе экстремальных оценок того же самого параметра.

5 Условный метод максимального правдоподобия

Обычно эконометрические модели формируются в терминах условных плотностей или вероятностей $y|x$. Рассмотренная же нами теория справедлива для совместного распределения всей пары. Обычно совместная плотность распределения $f(y,x|q)$ не специфицирована – мы не хотим предполагать форму плотности x, так как x, как правило, – экзогенная переменная, поведение которой мы не собираемся моделировать.

Пусть специфицирована только условная плотность $f(y|x,q)$, и пусть маржинальная плотность $f(x)$ независима от q (это серьезное предположение, но оно часто оправдано экзогенностью x).

Пример (Нормальная линейная регрессия).

$$y|x \sim \mathcal{N}(x'\beta, \sigma^2),$$

тогда $q = (b\ s^2)'$, $\theta = (\beta\ \sigma^2)'$, и

$$\log f(y|x,q) = \text{const} - \frac{1}{2}\log s^2 - \frac{(y-x'b)^2}{2s^2}.$$

Совместная плотность распадается на произведение

$$f(y,x|q) = f(y|x,q)f(x),$$

и, следовательно,

$$\log f(y,x|q) = \log f(y|x,q) + \log f(x).$$

Тогда

$$\widehat{\theta} = \arg\max_q \sum_{i=1}^n \log f(y,x|q) = \arg\max_q \sum_{i=1}^n \log f(y|x,q),$$

т.к. маржинальная лог-плотность $\log f(x)$ не оказывает влияния на максимизацию. Итак, при предположении о независимости $f(x)$ от q теория остаётся верна.

Продолжаем рассматривать начатый пример:

$$\begin{aligned}
s(y|x,q) &= \begin{pmatrix} \dfrac{y-x'b}{s^2}x \\ -\dfrac{1}{2s^2} + \dfrac{(y-x'b)^2}{2s^4} \end{pmatrix}, \\
\frac{\partial s(y|x,q)}{\partial q'} &= \begin{pmatrix} -\dfrac{xx'}{s^2} & -\dfrac{y-x'b}{s^4}x \\ -\dfrac{y-x'b}{s^4}x' & \dfrac{1}{2s^4} - \dfrac{(y-x'b)^2}{s^6} \end{pmatrix}, \\
\mathcal{I}(\theta) &= -E\left[\frac{\partial s(y|x,\theta)}{\partial q'}\right] = \frac{1}{\sigma^2}\begin{pmatrix} E[xx'] & 0 \\ 0' & \dfrac{1}{2\sigma^2} \end{pmatrix}.
\end{aligned}$$

Нули в $\mathcal{I}(\theta)$ означают асимптотическую нескоррелированность оценок компонент β и σ^2. Итак, мы получили асимптотические свойства условный-ММП-оценок:

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(E[xx'])^{-1}),$$

$$\sqrt{n}(\widehat{\sigma^2} - \sigma^2) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4).$$

Из асимптотической нескоррелированности и асимптотической нормальности следует асимптотическая независимость оценок компонент. Мы вывели асимптотические свойства, пропустив этап построения самой оценки. Через этот этап чаще всего приходится проходить, прибегая к численным методам, но в данном случае оценки выводятся в явном виде:

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\sigma^2} \end{pmatrix} = \arg\max_{b,s^2} \left\{ -\frac{n}{2}\log s^2 - \frac{1}{2s^2}\sum_{i=1}^{n}(y_i - x_i'b)^2 \right\},$$

откуда

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

т.е. ММП-оценка β совпала с МНК-оценкой, и

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

т.е. ММП-оценка σ^2 совпала с обычно используемой оценкой дисперсии без корректировки на степени свободы.

Пример (Модель бинарного выбора). Пусть

$$y = \begin{cases} 1, & x'\beta + e \geqslant 0, \\ 0, & x'\beta + e < 0, \end{cases} \qquad e|x \sim \mathcal{N}(0,1).$$

Тогда

$$y|x \sim \mathrm{Bernoulli}(\Phi(x'\beta)),$$

так как

$$Pr\{y = 1|x\} = Pr\{e \geqslant -x'\beta|x\} = 1 - \Phi(-x'\beta) = \Phi(x'\beta).$$

Условная плотность (или, точнее говоря, вероятностная масса) есть

$$f(y|x,b) = \Phi(x'b)^y(1 - \Phi(x'b))^{1-y}.$$

Параметр здесь только b, поскольку β и σ^2 одновременно не идентифицируются. Мы фиксируем σ^2 на 1. Далее,

$$\log f(y|x,b) = y\log\Phi(x'b) + (1-y)\log(1 - \Phi(x'b))$$

и

$$\widehat{\beta} = \arg\max_b \ \sum_{i=1}^n y_i \log \Phi(x'b) + (1-y)\log(1-\Phi(x'b)).$$

Нет никакой надежды получить в явном виде оценку для параметра. О состоятельности $\widehat{\beta} \xrightarrow{p} \beta$ и асимптотической нормальности $\sqrt{n}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, V_{\widehat{\beta}})$ мы уже знаем. Остается найти асимптотическую дисперсионную матрицу. Условная скор-функция есть

$$s(y|x,b) = y\frac{\phi(x'b)}{\Phi(x'b)} + (1-y)\frac{-\phi(x'b)}{1-\Phi(x'b)}x = \frac{y-\Phi(x'b)}{\Phi(x'b)(1-\Phi(x'b))}\phi(x'b)x.$$

Легче возвести в квадрат $s(y|x,b)$, чем брать ее производную по b, поэтому

$$\mathcal{I}(\beta) = E\left[\frac{(y-\Phi(x'\beta))^2}{\Phi(x'\beta)^2(1-\Phi(x'\beta))^2}\phi(x'\beta)^2xx'\right] = E\left[\frac{\phi(x'\beta)^2}{\Phi(x'\beta)(1-\Phi(x'\beta))}xx'\right].$$

Здесь использовался факт, что условная дисперсия $E[(y-E[y|x])^2|x]$ для условного распределения Бернулли - это условная вероятность успеха, помноженная на условную вероятность провала $\Phi(x'\beta)(1-\Phi(x'\beta))$. Асимптотическая дисперсионная матрица легко состоятельно оценивается:

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

Поскольку здесь суммируются с положительными весами квадраты векторов, оценка положительно определена по построению.

Ранее для этого примера мы строили НМНК- и ВНМНК-оценки на основе регрессии

$$E[y|x] = \Phi(x'\beta),$$

и в случае ВНМНК, скедастичной функции

$$V[y|x] = \Phi(x'\beta)(1-\Phi(x'\beta)).$$

Из сравнения асимптотических дисперсионных матриц следует, что ВНМНК-оценка и ММП-оценка асимптотически эквивалентны, хотя только вторая явно использует спецификацию плотности.

В обоих примерах мы имели дело с условным распределением и обходились с ним как с совместным, так как x предполагался экзогенным в обоих примерах. Интересно также рассмотреть более общий случай, когда хотя бы один параметр совместного распределения сидит и условном, и в маржинальном плотностях. Такие случаи редки в экономической практике, но интересно, что же здесь следует делать.

Пусть $q = (q_1, q_2, q_3)$, $\theta = (\theta_1, \theta_2, \theta_3)$, $f(y, x|q_1, q_2, q_3) = f(y|x, q_1, q_2) f(x|q_2, q_3)$. Имея спецификацию совместной плотности, можно построить оценки разных компонент θ:

$$\text{Совместный-ММП-оценки:} \quad \begin{pmatrix} \widehat{\theta}_1^J \\ \widehat{\theta}_2^J \\ \widehat{\theta}_3^J \end{pmatrix} = \arg\max_{q_1, q_2, q_3} \sum_{i=1}^{n} \log f(y_i, x_i|q_1, q_2, q_3),$$

$$\text{Условный-ММП-оценки:} \quad \begin{pmatrix} \widehat{\theta}_1^C \\ \widehat{\theta}_2^C \end{pmatrix} = \arg\max_{q_1, q_2} \sum_{i=1}^{n} \log f(y_i|x_i, q_1, q_2),$$

$$\text{Маржинальный-ММП-оценки:} \quad \begin{pmatrix} \widehat{\theta}_2^M \\ \widehat{\theta}_3^M \end{pmatrix} = \arg\max_{q_2, q_3} \sum_{i=1}^{n} \log f(x_i|q_2, q_3).$$

В итоге мы имеем 2 способа оценивания θ_1, 3 способа оценивания θ_2 и 2 способа оценивания θ_3.

Результат. При благоприятных условиях все три оценки состоятельны (каждая для своих частей θ, естественно) и асимптотически нормальны. Более того, $\widehat{\theta}_1^C$, $\widehat{\theta}_2^C$, $\widehat{\theta}_2^M$ и $\widehat{\theta}_3^M$ не могут быть более асимптотически эффективными, чем $\widehat{\theta}_1^J$, $\widehat{\theta}_2^J$ и $\widehat{\theta}_3^J$.

6 Тестирование в контексте метода максимального правдоподобия

Существует триада асимптотически эквивалентных тестов: тест Вальда W, тест отношения правдоподобия LR и тест множителей Лагранжа LM.

Любая нулевая гипотеза имеет структуру $H_0 : \theta \in \Theta_0 \subset \Theta$. Сначала рассмотрим простую нулевую гипотезу $H_0 : \theta = \theta_0$, т.е. множество $\Theta_0 = \{\theta_0\}$ состоит из одного элемента. Тогда

1. Тестовая статистика Вальда есть

$$W = n(\widehat{\theta} - \theta_0)' \widehat{\mathcal{I}}(\theta_0)(\widehat{\theta} - \theta_0) \xrightarrow{d} \chi^2_k.$$

   Вместо θ_0 внутри $\widehat{\mathcal{I}}(\theta_0)$ можно взять $\widehat{\theta}$; число степеней свободы предельного распределения равно размерности параметра θ. Матрица $\widehat{\mathcal{I}}\,(\theta_0)$, обратная к асимптотической дисперсионной матрице, пивотизирует тестовую статистику.

2. Тест отношения правдоподобия основан на значениях логарифмической функции правдоподобия

$$l_n(q) = \sum_{i=1}^{n} \log f(z_i|q).$$

Разложим $l_n(\theta_0)$ по формуле Тэйлора до второго порядка вокруг $\widehat{\theta}$:

$$\frac{1}{n}\ l_n(\theta_0) - \frac{1}{n}\ l_n(\widehat{\theta}) = \frac{1}{n}\underbrace{\left(\frac{\partial l_n(\widehat{\theta})}{\partial q'}\right)}_{0}(\theta_0 - \widehat{\theta}) + \frac{1}{2}(\theta_0 - \widehat{\theta})'\frac{1}{n}\frac{\partial^2 l_n(\widetilde{\theta})}{\partial q\,\partial q'}(\theta_0 - \widehat{\theta}),$$

где

$$\frac{\partial l_n(\widehat{\theta})}{\partial q'} = 0$$

согласно условиям первого порядка для $\widehat{\theta}$. При нулевой гипотезе $\widehat{\theta} \xrightarrow{p} \theta_0$ и $\widetilde{\theta} \xrightarrow{p} \theta_0$, так что

$$\frac{1}{n}\frac{\partial^2 l_n(\widetilde{\theta})}{\partial q\,\partial q'} \xrightarrow{p} E\left[\frac{\partial s(z,\theta_0)}{\partial q'}\right] = -\mathcal{I}(\theta_0).$$

В итоге

$$LR = 2(l_n(\widehat{\theta}) - l_n(\theta_0) \xrightarrow{d} \chi^2_k.$$

Асимптотически эта статистика ведет себя как Вальдовская и, как следствие, имеет такое же асимптотическое распределение.

3. Тест множителей Лагранжа (скор-тест). Нам известен следующий факт о скор-функции: если нулевая гипотеза верна, то

$$E[s(z,\theta_0)] = 0.$$

Согласно ЦПТ, при нулевой гипотезе

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) \xrightarrow{d} \mathcal{N}(0,\mathcal{I}(\theta_0)).$$

Пивотизировав эту статистику, получим:

$$LM = \frac{1}{n}\left(\sum_{i=1}^{n} s(z_i,\theta_0)\right)'\widehat{\mathcal{I}}(\theta_0)^{-1}\left(\sum_{i=1}^{n} s(z_i,\theta_0)\right) \xrightarrow{d} \chi^2_k.$$

Название «тест множителей Лагранжа» – это следствие интерпретации истинного скор как множителя Лагранжа для следующей оптимизационной задачи:

$$\max_{q=\theta_0} \frac{1}{n}\ l_n(q).$$

В случае составной нулевой гипотезы, однако, такая интерпретация не работает.

Пример. Пусть $z_1,\dots,z_n \sim i.i.d.$ Bernoulli(θ) и $H_0 : \theta = \theta_0$, где $0 < \theta_0 < 1$. Тогда

$$l_n(q) = \sum_{i=1}^{n} z_i \log q + (1 - z_i)\log(1 - q).$$

Из условия первого порядка для ММП-оценки находим

$$0 = \sum_{i=1}^{n} \left(\frac{z_i}{\widehat{\theta}} - \frac{1 - z_i}{1 - \widehat{\theta}} \right) \quad \Rightarrow \quad \widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} z_i.$$

Скор и информационная матрица равны

$$s(z, q) = \frac{z}{q} - \frac{1 - z}{1 - q} = \frac{z - q}{q(1 - q)},$$

$$\mathcal{I}(\theta) = E \left[\frac{(z - \theta)^2}{\theta^2 (1 - \theta)^2} \right] = \frac{1}{\theta(1 - \theta)}.$$

Выпишем тестовые статистики:

$$\begin{aligned}
W &= \frac{n(\widehat{\theta} - \theta_0)^2}{\theta_0(1 - \theta_0)}, \\
LR &= 2 \sum_{i=1}^{n} z_i \log \frac{\widehat{\theta}}{\theta_0} + (1 - z_i) \log \frac{1 - \widehat{\theta}}{1 - \theta_0} = \\
&= 2n \left(\widehat{\theta} \log \frac{\widehat{\theta}}{\theta_0} + (1 - \widehat{\theta}) \log \frac{1 - \widehat{\theta}}{1 - \theta_0} \right), \\
LM &= \frac{1}{n} \left(\sum_{i=1}^{n} \frac{z_i - \theta_0}{\theta_0(1 - \theta_0)} \right)^2 \theta_0(1 - \theta_0) = W.
\end{aligned}$$

Теперь рассмотрим составную нулевую гипотезу

$$H_0 : \underset{r \times 1}{g}(\theta) = 0, \quad r \leqslant k,$$

т.е. $\Theta_0 = \{\theta \in \Theta,\ g(\theta) = 0\}$. В этом случае

$$\begin{aligned}
W &= ng(\widehat{\theta})'[\widehat{G}(\widehat{\mathcal{I}}(\widehat{\theta}))^{-1}\widehat{G}']^{-1} g(\widehat{\theta}) \xrightarrow{d} \chi^2_r, \quad \text{где } \underset{r \times k}{G} \equiv \frac{\partial g(\theta)}{\partial \theta'},\ \widehat{G} \equiv \frac{\partial g(\widehat{\theta})}{\partial \theta'}, \\
LR &= 2(l_n(\widehat{\theta}) - l_n(\widehat{\theta}^R)) \xrightarrow{d} \chi^2_r, \quad \text{где } \widehat{\theta}^R = \arg\max_{q \in \Theta_0} l_n(q), \\
LM &= \frac{1}{n} \left(\sum_{i=1}^{n} s(z_i, \widehat{\theta}^R) \right)' \left(\widehat{\mathcal{I}}(\widehat{\theta}^R) \right)^{-1} \left(\sum_{i=1}^{n} s(z_i, \widehat{\theta}^R) \right) \xrightarrow{d} \chi^2_r.
\end{aligned}$$

Какую статистику лучше использовать? Все три асимптотически эквивалентны, хотя численно разные (в частных случаях они могут совпадать, а иногда их можно даже ранжировать). Соображения, которые можно использовать при выборе тестовой статистики, следующие:

Инвариантиность к перепараметризации H_0. Например, одну и ту же нулевую гипотезу можно сформулировать как $\theta_1 = \theta_2$, либо как $\theta_1/\theta_2 = 1$. W-статистика не инвариантна к такой перепараметризации, LR – инвариантна, а LM – инвариантна или нет в зависимости от обстоятельств и способа её построения.

Удобство использования. Если относительно несложно оценить и ограниченную модель, и неограниченную, то лучше использовать LR-статистику (хотя бы из-за инвариантности относительно перепараметризации). Если сложно оценить ограниченную модель, можно отдать предпочтение W-статистике, если же сложно оценить неограниченную модель, стоит предпочесть LM-статистику.

7 ММП-оценивание в моделях временных рядов

Пусть теперь данные $z_1, \dots, z_T$ представляют собой стационарный временной ряд. Поскольку наблюдения не независимы, $\mathcal{L}(z_1, \dots, z_T, q) \neq \prod_{t=1}^{T} f(z_t|q)$, функция правдоподобия не распадается на произведение маржинальных плотностей. Оказывается, что функцию правдоподобия все равно можно разложить на произведение, но посложнее:

$$\mathcal{L}(z_1, \dots, z_T, q) = f(z_T|z_{T-1}, \dots, z_1)\mathcal{L}(z_1, \dots, z_{T-1}, q) = \dots.$$

Предположим, что z_t обладает марковским свойством, т.е.

$$f(z_t|z_{t-1}, \dots) = f(z_t|z_{t-1}, \dots, z_{t-p})$$

для некоторого конечного p. Тогда

$$\mathcal{L}(z_1, \dots, z_T, q) = \prod_{t=p+1}^{T} f(z_t|z_{t-1}, \dots, z_{t-p}, q) \times \prod_{t=1}^{p} f(z_t|z_{t-1}, \dots, z_1, q)$$

Точный метод максимального правдоподобия даёт точный-ММП-оценку

$$\widehat{\theta}^E = \arg\max_q \log \mathcal{L}(z_1, \dots, z_T, q).$$

Подсчитать такую оценку сложно, поэтому предпочитают использовать приближенный подход, дающий приближённый-ММП-оценку

$$\widehat{\theta}^A = \arg\max_q \sum_{t=1+p}^{T} \log f(z_t|z_{t-1}, \dots, z_{t-p}, q),$$

которая игнорирует зависимость $\sum_{t=1}^{p} \log f(z_t|z_{t-1}, \dots, z_1, q)$ от q.

Точный и приближенный методы максимального правдоподобия асимптотически эквивалентны, хотя численно оценки могут сильно отличаться. Рассмотрим примеры использования метода максимального правдоподобия в моделях временных рядов, имея в виду приближенный ММП.

Пример 1. Рассмотрим авторегрессию порядка p с нормальными инновациями

$$y_t|x_t \sim \mathcal{N}(x_t'\beta, \sigma^2),$$

где $x_t = (1 \; y_{t-1} \; \ldots \; y_{t-p})'$ и $\beta = (\mu \; \alpha_1 \; \ldots \; \alpha_p)'$. Обозначим $\theta = (\beta' \; \sigma^2)'$. Приближенная логарифмическая функция правдоподобия равна

$$l^A(y_1, \ldots, y_T, q) = \text{const} - \sum_{t=1+p}^{T} \left\{ \frac{1}{2} \log s^2 + \frac{(y_t - x_t' b)^2}{2s^2} \right\},$$

а ММП-оценки –

$$\begin{aligned} \widehat{\beta} &= \left(\sum_{t=1+p}^{T} x_t x_t' \right)^{-1} \sum_{t=1+p}^{T} x_t y_t, \\ \widehat{\sigma^2} &= \frac{1}{T-p} \sum_{t=1+p}^{T} (y_t - x_t' \widehat{\beta})^2. \end{aligned}$$

Если бы мы работали не с приближённым правдоподобием, а точным, у нас бы не вывелись оценки в явной форме. Нам бы пришлось еще учесть сумму с индексом, бегущим от 1 до p, и учесть плотности для начальных y-ков, условные на предыстории, а предыстория у них у всех разная. Поэтому нам бы пришлось это нормальное распределение интегрировать и находить для каждого из этих p наблюдений свою условную плотность.

Пример 2. Рассмотрим модель бегущего среднего порядка p с нормальными инновациями:

$$y_t = \mu + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \ldots + \phi_p \varepsilon_{t-p} \quad \varepsilon_t | I_{t-1} \sim \mathcal{N}(0, \sigma^2).$$

Обозначим $q = (m \; f_1 \; \ldots \; f_p \; s^2)'$ и $\theta = (\mu \; \phi_1 \; \ldots \; \phi_p \; \sigma^2)'$. Приближенная логарифмическая функция правдоподобия равна

$$l^A(y_1, \ldots, y_T, q) = \text{const} - \sum_{t=1+p}^{T} \left\{ \frac{1}{2} \log s^2 + \frac{\varepsilon_t(q)^2}{2s^2} \right\},$$

где

$$\varepsilon_t(q) = y_t - m - f_1 \varepsilon_{t-1}(q) - \cdots - f_p \varepsilon_{t-p}(q).$$

Поскольку у нас рекурсивное соотношение, нужны стартовые значения. Для асимптотических свойств неважно откуда, поэтому проще всего из нулевых:

$$\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-p} = 0.$$

Максимизировать всё равно приходится численно. Из-за того, что есть произвол в выборе стартовых значений, разные исследователи, работающие с одними и теми же данными над одной и той же задачей, могут получить разные численные результаты. Это нормально.

III Обобщённый метод моментов

1 Условия на моменты

Пусть известно, что параметр θ, который мы хотим оценить, удовлетворяет так называемому *условию на моменты*, или *условию ортогональности*

$$E[m(z,\theta)] = 0,$$

где z – представитель наблюдаемых данных, θ – вектор параметров размерности $k\times 1$, $m(z,q)$ – *функция моментов*, вид которой известен с точностью до параметров, и которая – вектор размерности $l\times 1$, причём $l \geqslant k$. Если $l = k$, у нас случай *точной идентификации*, если же $l > k$ – случай *сверхидентификации*.

Далее приведены примеры, которые мы будем рассматривать на протяжении всей темы.

Пример 1. Пусть известно, что распределение z нескошенное, то есть $E[(z-\theta)^3] = 0$, и мы хотим оценить среднее z, то есть истинный параметр – $\theta = E[z]$. Тогда условие на моменты запишется так:

$$m(z,q) = \begin{pmatrix} z-q \\ (z-q)^3 \end{pmatrix},$$

и $k = 1$, $l = 2$.

Пример 2. Пусть известно, что x и y имеют общее среднее, то есть $E[x] = E[y] = \theta$. Наша цель – оценить это среднее. Условие на моменты запишется так:

$$m(x,y,q) = \begin{pmatrix} x-q \\ y-q \end{pmatrix},$$

и вновь $k = 1$, $l = 2$.

В обоих примерах мы имеем сверхидентификацию.

2 Оценка обобщённого метода моментов

Пусть имеется случайная выборка $z_1,\dots,z_n$. Рассмотрим сначала случай точной идентификации $l = k$. Применение принципа аналогий дает уравнение

$$\frac{1}{n}\sum_{i=1}^{n} m(z_i,\widehat{\theta}) = 0.$$

Неизвестных здесь столько же, сколько уравнений, и $\widehat{\theta}$ называется оценкой *классического метода моментов (КММ)*.

Пример 1 (продолжение). Возьмём только первое из условий, тогда КММ-оценка будет равна

$$\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} z_i.$$

Возьмём только второе из условий, тогда КММ-оценка будет иметь неявный вид

$$\sum_{i=1}^{n}(z_i - \widehat{\theta}_2)^3 = 0.$$

Пример 2 (продолжение). Возьмём только первое или второе из условий, тогда КММ-оценка будет, соответственно,

$$\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Понятно, что в общем случае $\widehat{\theta}_1 \neq \widehat{\theta}_2$.

Чаще всего на практике, правда, идентифицирующей информации больше, чем достаточно, то есть $l > k$. Разобравшись в этой более сложной ситуации, мы увидим, что КММ – это лишь ее частный случай.

Если $l > k$, вышеприведённый трюк не работает, ибо результирующая система из l уравнений с k неизвестными не имеет в общем случае решений. Выходом из ситуации служит преобразование условий на моменты в оптимизационную задачу:

$$E[m(z,\theta)] = 0 \quad \Rightarrow \quad \theta = \arg\min_{q\in\Theta} \underbrace{E[m(z,q)]'}_{1\times l} \underbrace{W}_{l\times l} \underbrace{E[m(z,q)]}_{l\times 1},$$

где W – взвешивающая матрица, любая симметричная положительно определенная матрица.

Теперь можно построить оценку, которая гарантировано существует, используя принцип аналогий:

$$\widehat{\theta} = \arg\min_{q\in\Theta} \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i,q)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i,q)\right).$$

Это – оценка *обобщенного метода моментов (ОММ)*. Здесь W_n – симметричная положительно определённая матрица, состоятельно оценивающая W:

$$W_n \xrightarrow{p} W.$$

В типичных случаях W_n зависит от данных. Конечно же, $\widehat{\theta}$ зависит от выбора W и W_n.

Замечание. КММ есть частный случай ОММ при $l = k$. При этом выбор W и W_n не играет никакой роли.

Пример 1 (продолжение). Возьмём оба условия на моменты, тогда ОММ-оценка запишется так:

$$\widehat{\theta} = \arg\min_q \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (z_i - q) & \frac{1}{n}\sum_{i=1}^n (z_i - q)^3 \end{pmatrix} W_n \begin{pmatrix} n^{-1}\sum_{i=1}^n (z_i - q) \\ n^{-1}\sum_{i=1}^n (z_i - q)^3 \end{pmatrix}$$

Пример 2 (продолжение). Возьмём оба условия на моменты, тогда ОММ-оценка запишется так:

$$\widehat{\theta} = \arg\min_q \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (x_i - q) & \frac{1}{n}\sum_{i=1}^n (y_i - q) \end{pmatrix} W_n \begin{pmatrix} n^{-1}\sum_{i=1}^n (x_i - q) \\ n^{-1}\sum_{i=1}^n (y_i - q) \end{pmatrix}$$

3 Асимптотические свойства ОММ-оценок

Как и ранее, важнейшим условием является *условие идентификации:* $E[m(z,q)] = 0$ тогда и только тогда, когда $q = \theta$. Обозначим

$$\underset{l\times k}{Q_{\partial m}} = E\left[\frac{\partial m(z,\theta)}{\partial q'}\right],$$

$$\underset{l\times l}{Q_{mm}} = E\left[m(z,\theta)m(z,\theta)'\right].$$

При благоприятных условиях,

- ОММ-оценка состоятельна: $\widehat{\theta} \xrightarrow{p} \theta$;
- ОММ-оценка асимптотически нормальна: $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V_{\widehat{\theta}})$;
- Асимптотическая дисперсионная матрица ОММ-оценки равна

$$V_{\widehat{\theta}} = (Q'_{\partial m} W Q_{\partial m})^{-1} Q'_{\partial m} W Q_{mm} W Q_{\partial m} (Q'_{\partial m} W Q_{\partial m})^{-1}.$$

Выведем асимптотическое распределение, предположив состоятельность. Запишем условия первого порядка для $\widehat{\theta}$:

$$2\left(\frac{1}{n}\sum_{i=1}^n \frac{\partial m(z_i,\widehat{\theta})'}{\partial q}\right) W_n \left(\frac{1}{n}\sum_{i=1}^n m(z_i,\widehat{\theta})\right) = 0,$$

а теперь разложим последний множитель в ряд Тэйлора вокруг θ до членов первого порядка:

$$\frac{1}{n}\sum_{i=1}^n m(z_i,\widehat{\theta}) = \frac{1}{n}\sum_{i=1}^n m(z_i,\theta) + \frac{1}{n}\sum_{i=1}^n \frac{\partial m(z_i,\widetilde{\theta})}{\partial q'}(\widehat{\theta} - \theta).$$

В итоге получаем:

$$\sqrt{n}(\widehat{\theta} - \theta) = -\left(\frac{1}{n}\sum_{i=1}^n \frac{\partial \widehat{m}'}{\partial q}\; W_n\; \frac{1}{n}\sum_{i=1}^n \frac{\partial \widetilde{m}'}{\partial q}\right)^{-1} \frac{1}{n}\sum_{i=1}^n \frac{\partial \widehat{m}'}{\partial q}\; W_n\; \frac{1}{\sqrt{n}}\sum_{i=1}^n m(z_i,\theta)$$

$$\xrightarrow[n\to\infty]{d} (Q'_{\partial m}WQ_{\partial m})^{-1}Q'_{\partial m}W\mathcal{N}(0, Q_{mm}),$$

откуда непосредственно следует вид матрицы $V_{\widehat{\theta}}$.

В частном случае точной идентификации выбор взвешивающей матрицы, как популяционной W, так и выборочной W_n, не имеет никакого значения для оценки, в том числе и для её асимптотических свойств. Действительно, матрица $Q_{\partial m}$ квадратная и её можно обратить:

$$\begin{aligned}
V_{\widehat{\theta}} &= (Q'_{\partial m}WQ_{\partial m})^{-1}Q'_{\partial m}WQ_{mm}WQ_{\partial m}(Q'_{\partial m}WQ_{\partial m})^{-1} \\
&= Q^{-1}_{\partial m}W^{-1}Q'^{-1}_{\partial m}Q'_{\partial m}WQ_{mm}WQ_{\partial m}Q^{-1}_{\partial m}W^{-1}Q'^{-1}_{\partial m} \\
&= Q^{-1}_{\partial m}W^{-1}WQ_{mm}WW^{-1}Q'^{-1}_{\partial m} \\
&= Q^{-1}_{\partial m}Q_{mm}Q'^{-1}_{\partial m}.
\end{aligned}$$

4 Эффективная оценка обобщённого метода моментов

Теперь мы хотим выбрать W таким образом, чтобы $\widehat{\theta}$ была настолько асимптотически эффективна, насколько это возможно.

Результат. Оптимальный выбор взвешивающей матрицы W в смысле минимизации $V_{\widehat{\theta}}$ есть Q^{-1}_{mm}, а минимально достижимое значение $V_{\widehat{\theta}}$ при этом выборе равно

$$(Q'_{\partial m}Q^{-1}_{mm}Q_{\partial m})^{-1}.$$

Интуиция. Мы взвешиваем функции на моменты обратно их стандартным отклонениям, выравнивая таким образом условия на моменты по степени доверия.

Доказательство. Покажем, что если взять разницу между асимптотическими матрицами ОММ-оценок $V^{W}_{\widehat{\theta}}$ и $V^{Q^{-1}_{mm}}_{\widehat{\theta}}$, первая из которых соответствует произвольной взвешивающей матрице, а вторая – оптимальной, то в результате получится неотрицательно определенная матрица:

$$\begin{aligned}
V^{W}_{\widehat{\theta}} - V^{Q^{-1}_{mm}}_{\widehat{\theta}} &= (Q'_{\partial m}WQ_{\partial m})^{-1}Q'_{\partial m}W \times \\
&\times \underbrace{(Q_{mm} - Q_{\partial m}(Q'_{\partial m}Q^{-1}_{mm}Q_{\partial m})^{-1}Q'_{\partial m})}_{(\star)} \times \\
&\times WQ_{\partial m}(Q'_{\partial m}WQ_{\partial m})^{-1}.
\end{aligned}$$

Эта матрица неотрицательно определена, если матрица $(\star)$ неотрицательно определена. Определим вектор u следующим образом:

$$u = m(z, \theta) - Q_{\partial m}(Q'_{\partial m}Q^{-1}_{mm}Q_{\partial m})^{-1}Q'_{\partial m}Q^{-1}_{mm}m(z, \theta).$$

Легко убедиться, что матрица $(\star)$ равна $E[uu']$, т.е. она неотрицательно определена.

Определение. ОММ-оценка, использующая оптимальное взвешивание $W = Q_{mm}^{-1}$, называется *эффективной* или *оптимальной ОММ-оценкой.*

Определение. *Доступной эффективной ОММ-оценкой* называется ОММ-оценка

$$\widehat{\theta} = \arg\min_{q\in\Theta} \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i, q)\right)' \widehat{Q}_{mm}^{-1} \left(\sum_{i=1}^{n} m(z_i, q)\right),$$

где $\widehat{Q}_{mm} \xrightarrow{p} Q_{mm}$, и $\widehat{Q}_{mm}$ симметрична и положительно определена.

Построим доступные эффективные ОММ-оценки для наших двух примеров.

Пример 1 (продолжение). Обозначим за $\mu_r = E[(z-\theta)^r]$ центрированный момент r-го порядка. Тогда

$$Q_{\partial m} = E\begin{bmatrix} -1 \\ -3(z-q)^2 \end{bmatrix}_{q=\theta} = -\begin{pmatrix} 1 \\ 3\mu_2 \end{pmatrix},$$

$$Q_{mm} = E\begin{bmatrix} (z-q)^2 & (z-q)^4 \\ (z-q)^4 & (z-q)^6 \end{bmatrix}_{q=\theta} = \begin{pmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{pmatrix}.$$

Оценку для μ_r легко построить из принципа аналогий, как среднее по выборке от r-х степеней отклонения z от истинного параметра, для которого у нас есть предварительная оценка $\widehat{\theta}_1$:

$$\widehat{\mu_r} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \widehat{\theta}_1)^r,$$

тогда

$$\widehat{Q}_{\partial m} = -\begin{pmatrix} 1 \\ 3\widehat{\mu}_2 \end{pmatrix}, \quad \widehat{Q}_{mm} = \begin{pmatrix} \widehat{\mu}_2 & \widehat{\mu}_4 \\ \widehat{\mu}_4 & \widehat{\mu}_6 \end{pmatrix},$$

и

$$\widehat{\theta} = \arg\min_{q\in\Theta} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}(z_i - q) & \frac{1}{n}\sum_{i=1}^{n}(z_i - q)^3 \end{pmatrix} \begin{pmatrix} \widehat{\mu}_2 & \widehat{\mu}_4 \\ \widehat{\mu}_4 & \widehat{\mu}_6 \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^{n}(z_i - q) \\ n^{-1}\sum_{i=1}^{n}(z_i - q)^3 \end{pmatrix}.$$

Обратив $\begin{pmatrix} \widehat{\mu}_2 & \widehat{\mu}_4 \\ \widehat{\mu}_4 & \widehat{\mu}_6 \end{pmatrix}$, можно игнорировать положительный определитель, который на максимизацию не влияет. Ищем $V_{\widehat{\theta}}$:

$$V_{\widehat{\theta}} = \left(\begin{pmatrix} -1 & -3\mu_2 \end{pmatrix} \begin{pmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ -3\mu_2 \end{pmatrix} \right)^{-1} = \frac{\mu_2\mu_6 - \mu_4^2}{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3},$$

и для неё – состоятельную оценку:

$$\widehat{V}_{\widehat{\theta}} = \frac{\widehat{\mu}_2\widehat{\mu}_6 - \widehat{\mu}_4^2}{\widehat{\mu}_6 - 6\widehat{\mu}_2\widehat{\mu}_4 + 9\widehat{\mu}_2^3} \xrightarrow{p} V_{\widehat{\theta}}.$$

Интересно убедиться в большей эффективности доступной ОММ-оценки по сравнению с $\widehat{\theta}_1$. В последнем случае $V_{\widehat{\theta}_1} = \mu_2$. Итак,

$$\mu_2 \geqslant \frac{\mu_2\mu_6 - \mu_4^2}{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3},$$
$$9\mu_2^4 - 6\mu_2^2\mu_4 + \mu_4^2 \geqslant 0,$$
$$(3\mu_2^2 - \mu_4)^2 \geqslant 0.$$

Если $\mu_4 \neq 3\mu_2^2$, то вновь построенная оценка более асимптотически эффективна. Если же $\mu_4 = 3\mu_2^2$ (например, в случае нормальности), то $\widehat{\theta}_1$ и $\widehat{\theta}$ асимптотически эквивалентны.

Определение. Условие на моменты называется *излишним при наличии других условий на моменты*, если эффективные ОММ-оценки на основе всех условий и на основе только тех других условий асимптотически эквивалентны.

Понятно, что излишнее условие лучше опустить, при этом асимптотическая эффективность не ухудшается. При этом оценки получаются проще, что обычно благоприятно сказывается на свойствах оценок в конечных выборках. В последние годы эконометристы серьёзно заинтересовались понятием излишних условий на моменты, и, в частности, вопросом тестирования на излишество.

Пример 2 (продолжение).

$$Q_{\partial m} = -\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad Q_{mm} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}, \quad \widehat{Q}_{mm} = \begin{pmatrix} \widehat{\sigma_x^2} & \widehat{\sigma_{xy}} \\ \widehat{\sigma_{xy}} & \widehat{\sigma_y^2} \end{pmatrix},$$

где

$$\begin{aligned} \widehat{\sigma_x^2} &= \frac{1}{n}\sum_{i=1}^n (x_i - \widehat{\theta}_1)^2, \\ \widehat{\sigma_{xy}} &= \frac{1}{n}\sum_{i=1}^n (x_i - \widehat{\theta}_1)(y_i - \widehat{\theta}_2), \\ \widehat{\sigma_y^2} &= \frac{1}{n}\sum_{i=1}^n (y_i - \widehat{\theta}_1)^2. \end{aligned}$$

где $\widehat{\theta}_1$ и $\widehat{\theta}_2$ – предварительные оценки на основе подвыборок по x и по y соответственно. Доступная эффективная ОММ-оценка равна

$$\widehat{\theta} = \arg\min_{q\in\Theta} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (x_i - q) & \frac{1}{n}\sum_{i=1}^n (y_i - q) \end{pmatrix} \begin{pmatrix} \widehat{\sigma_x^2} & \widehat{\sigma_{xy}} \\ \widehat{\sigma_{xy}} & \widehat{\sigma_y^2} \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n (x_i - q) \\ n^{-1}\sum_{i=1}^n (y_i - q) \end{pmatrix}.$$

Условие первого порядка:

$$2(-1 \ \ -1) \begin{pmatrix} \widehat{\sigma_x^2} & \widehat{\sigma_{xy}} \\ \widehat{\sigma_{xy}} & \widehat{\sigma_y^2} \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n (x_i - \widehat{\theta}) \\ n^{-1}\sum_{i=1}^n (y_i - \widehat{\theta}) \end{pmatrix} = 0.$$

Выразить $\widehat{\theta}$ можно в явном виде благодаря тому, что условия на моменты линейны по параметру:

$$\widehat{\theta} = \frac{\widehat{\sigma_y^2} - \widehat{\sigma_{xy}}}{\widehat{\sigma_x^2} + \widehat{\sigma_y^2} - 2\widehat{\sigma_{xy}}}\widehat{\theta}_1 + \frac{\widehat{\sigma_x^2} - \widehat{\sigma_{xy}}}{\widehat{\sigma_x^2} + \widehat{\sigma_y^2} - 2\widehat{\sigma_{xy}}}\widehat{\theta}_2.$$

Асимптотическая дисперсионная матрица равна

$$V_{\widehat{\theta}} = \left(\begin{pmatrix} -1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right)^{-1} = \frac{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}},$$

в то время как

$$V_{\widehat{\theta}_1} = \sigma_x^2, \quad V_{\widehat{\theta}_2} = \sigma_y^2.$$

Второе условие излишнее при наличии первого тогда и только тогда, когда дисперсия x равна ковариации x и y: $\sigma_x^2 = \sigma_{xy}$. Аналогично, первое условие излишнее при наличии второго тогда и только тогда, когда дисперсия y равна ковариации x и y: $\sigma_x^2 = \sigma_{xy}$.

В обоих примерах мы получили доступную ОММ-оценку за два шага. В конечных выборках могут быть неприятные последствия этой двухшаговости. То, что оба шага проводятся по одной и той же выборке, приводит к неточности асимптотического приближения.

Общий алгоритм двухшагового эффективного обобщённого метода моментов.

- **Шаг 1:** Взять предварительную состоятельную оценку $\widehat{\theta}_0$ параметра θ (например, неэффективную ОММ-оценку с $W_n = I_n$). С помощью неё построить состоятельную оценку $\widehat{Q}_{mm}$.

- **Шаг 2:** Вычислить эффективную ОММ-оценку, используя $W_n = \widehat{Q}_{mm}^{-1}$.

5 Тест на сверхидентифицирующие ограничения

Мы знаем, как строить эффективную ОММ-оценку, знаем, какие у нее асимптотические свойства, знаем, как выглядит асимптотическая дисперсия и как ее состоятельно оценивать. Всё готово для проведения обычной инференции, в частности, для тестирования обычных гипотез – ограничений на параметры. Оказывается, что благодаря особому типу задачи, а точнее, благодаря сверхидентифицирующей информации, имеется ещё возможность проверить верность модели в целом, то есть произвести тест на спецификацию. Следующее, что мы рассмотрим – это *тест на сверхидентифицирующие ограничения*.

Ранее в основном мы имели дело с точно идентифицированными задачами, где информации в постановке было ровно столько, сколько нужно для оценивания параметров, и она использовалась без остатка (иногда, правда, мы сами её ограничивали, как в регрессии среднего, используя нескоррелированность ошибки только с одной из функций от регрессоров). Теперь же из-за наличия сверхидентифицирующих ограничений количество идентифицирующей информации строго больше, чем достаточно для оценивания, и излишек мы можем использовать для тестирования самосовместимости модели, то есть условий на моменты.

Итак, нулевая гипотеза выглядит как

$$H_0: \quad E[m(z,\theta)] = 0,$$

а альтернативная – как

$$H_A: \quad \nexists\, \theta: \; E[m(z,\theta)] = 0.$$

Результат. Пусть $\widehat{\theta}$ – эффективная ОММ-оценка параметра θ на основе условий на моменты $E[m(z,\theta)] = 0$. Тогда *J-статистика*

$$J = n\left(\frac{1}{n}\sum_{i=1}^{n} m(z_i,\widehat{\theta})\right)' \widehat{Q}_{mm}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} m(z_i,\widehat{\theta})\right)$$

асимптотически распределена как χ^2_{l-k}.

Доказательство. Рассмотрим вектор

$$v = Q_{mm}^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\widehat{\theta}).$$

Тогда $J \overset{A}{=} v'v$, где $\overset{A}{=}$ означает асимптотическую эквивалентность. Разложим этот вектор в ряд Тэйлора вокруг истинного параметра θ:

$$v = Q_{mm}^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\theta) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\widetilde{\theta})}{\partial q'}\sqrt{n}(\widehat{\theta}-\theta).$$

Но

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\widehat{\theta}) \overset{d}{\to} \mathcal{N}(0,Q_{mm}),$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\widetilde{\theta})}{\partial q'} \overset{p}{\to} Q_{\partial m}.$$

Из нашего вывода асимптотического распределения ОММ-оценок следует, что

$$\sqrt{n}(\widehat{\theta}-\theta) \overset{A}{=} -(Q'_{\partial m}Q_{mm}^{-1}Q_{\partial m})^{-1}Q'_{\partial m}Q_{mm}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\theta).$$

Таким образом,

$$v \overset{A}{=} Q_{mm}^{-\frac{1}{2}}(I_l - Q_{\partial m}(Q'_{\partial m}Q_{mm}^{-1}Q_{\partial m})^{-1}Q'_{\partial m}Q_{mm}^{-1})\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\theta)$$

$$\xrightarrow{d} \mathcal{N}(0, I_l - Q_{mm}^{-\frac{1}{2}}Q_{\partial m}(Q'_{\partial m}Q_{mm}^{-1}Q_{\partial m})^{-1}Q'_{\partial m}Q_{mm}^{-\frac{1}{2}}).$$

Дисперсионная матрица этого асимптотического распределения является симметричной идемпотентной матрицей[1] ранга $l-k$, поэтому результат следует.

Разберем в этом свете пример 2, который касается оценки общего среднего выборок $x_1,\dots,x_n$ и $y_1,\dots,y_n$.

Пример 2 (продолжение). Ранее было получено, что

$$\widehat{\theta} = \alpha\overline{x} + (1-\alpha)\overline{y},$$

где

$$\alpha = \frac{\widehat{\sigma_y^2} - \widehat{\sigma_{xy}}}{\widehat{\sigma_x^2} + \widehat{\sigma_y^2} - 2\widehat{\sigma_{xy}}}.$$

Строим

$$J = n\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} x_i - \widehat{\theta} & \frac{1}{n}\sum_{i=1}^{n} y_i - \widehat{\theta}\end{pmatrix}\begin{pmatrix}\widehat{\sigma_x^2} & \widehat{\sigma_{xy}} \\ \widehat{\sigma_{xy}} & \widehat{\sigma_y^2}\end{pmatrix}^{-1}\begin{pmatrix}n^{-1}\sum_{i=1}^{n} x_i - \widehat{\theta} \\ n^{-1}\sum_{i=1}^{n} y_i - \widehat{\theta}\end{pmatrix} =$$

$$= \frac{n(\overline{x}-\overline{y})^2}{\widehat{\sigma_x^2} + \widehat{\sigma_y^2} - 2\widehat{\sigma_{xy}}} \xrightarrow{d} \chi_1^2.$$

Видим, что J базируется на разнице между $\overline{x}$ и $\overline{y}$, что логично, так как у нас нулевой гипотезой является равенство математических ожиданий x и y. Этот тест можно получить и без ОММ-теории, просто эвристически. Действительно,

$$H_0: \quad E[x-y] = 0.$$

Согласно принципу аналогий, при нулевой гипотезе

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \xrightarrow{p} 0, \quad t = \frac{\sqrt{n}(\overline{x}-\overline{y})}{\sqrt{\widehat{\sigma_x^2} + \widehat{\sigma_y^2} - 2\widehat{\sigma_{xy}}}} \xrightarrow{d} \mathcal{N}(0,1).$$

Получили тест, эквивалентный J-тесту.

[1] Напоминание: матрица M называется *симметричной и идемпотентной*, если $M' = M$ и $M^2 = M$. Матрица $I - M$ в этом случае, очевидно, тоже симметрична и идемпотентна. Результат: если вектор распределён нормально с такой дисперсионной матрицей, то квадрат его Евклидовой нормы распределён по χ^2 с количеством степеней свободы, равным рангу дисперсионной матрицы.

При ОММ-оценивании моделей со сверхидентифицирующими ограничениями J-тест нужно обязательно провести и желательно получить в нем положительный результат, то есть показать, что построенная модель совместна.

Техника проверки гипотезы стандартна: если $J \leqslant q_{1-\alpha}$, данные не противоречат спецификации модели, если же $J > q_{1-\alpha}$, то спецификация модели бракуется. В последнем случае остаётся неясным источник неверной спецификации. Отрицательный результат теста говорит только о том, что не все в порядке с какими-то условиями на моменты, но о том, с какими именно, ничего не говорит.

6 Инструментальные переменные и ОММ

Всё, что мы узнали про ОММ к настоящему моменту – это чисто статистическая теория, поскольку эконометрические модели мы не рассматривали; наши два примера также были чисто статистическими. Теперь же появляется привычная эконометрическая модель – линейное уравнение с инструментальными переменными, к которому мы и применим ОММ-теорию.

$$y = x' \underset{k\times 1}{\beta} + e, \quad E[e|\underset{l\times 1}{z}] = 0, \quad l \geqslant k.$$

Объясняющие переменные в x потенциально эндогенные. Имеются экзогенные инструменты z, которых l штук.

Мы уже рассматривали такую модель. В случае точной идентификации ($l = k$) мы получали простую оценку инструментальных переменных. В случае сверхидентификации мы из каких-то проекционных соображений строили оценку двухшагового (двухступенчатого) метода наименьших квадратов (2ШМНК). С помощью ОММ-теории мы можем показать, что обе эти оценки – это ОММ-оценки, и, кроме того, можно создать и другие экземпляры, превосходящие их в асимптотической эффективности. Продолжим описание модели:

- Условие нескоррелированности ошибок и инструментов, которое следует из посылки об экзогенности инструментов, называется «валидностью» последних:

$$E[ze] = 0.$$

- Условие скоррелированности инструментов и регрессоров называется «релевантностью» инструментов:

$$\underset{k\times l}{Q_{xz}} = E[xz'] \text{ – матрица полного ранга } k.$$

- Условие линейной неповторяемости инструментов:

$$\underset{l\times l}{Q_{zz}} = E[zz'] \text{ – невырожденная матрица.}$$

При точной идентификации ($l = k$) ОММ вырождается в классический метод моментов (КММ):

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

Гораздо интереснее случай сверхидентификации ($l > k$). Имеем условия на моменты

$$E[\underbrace{z(y - x'\beta)}_{m(x,y,z,\beta)}] = \underset{l\times 1}{0}.$$

Наши сопровождающие матрицы (которые мы переобозначим для удобства) равны

$$Q_{\partial m} = -E[zx'] = -Q_{xz}', \quad Q_{mm} = E[zz'(y - x'\beta)^2] = -Q_{zze^2}.$$

Последняя в случае условной гомоскедастичности упрощается, превращаясь в $\sigma^2 Q_{zz}$. Попробуем использовать неэффективный ОММ со следующим взвешиванием:

$$W_n = \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \xrightarrow{p} W = Q_{zz}^{-1} \not\propto Q_{mm}^{-1}.$$

По построению, это симметричная положительно определенная матрица. ОММ действительно неэффективный, кроме случая условной гомоскедастичности, когда $Q_{zz}^{-1} \propto Q_{mm}^{-1}$ и взвешивание оказывается оптимальным. Соответствующая ОММ-оценка есть

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

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

Получилось знакомое выражение, которое совпадает с оценкой по двухшаговому методу наименьших квадратов (2ШМНК-оценкой). Отсюда следуют следующие выводы:

- 2ШМНК-оценка является ОММ-оценкой;

- 2ШМНК-оценка асимптотически неэффективна в классе ОММ-оценок, которые используют условие на моменты $E[ze] = 0$, за исключением случая условной гомоскедастичности.

Построим теперь эффективную ОММ-оценку. Оптимальное взвешивание будет

$$W_n = \widehat{Q}_{mm}^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \widehat{e}_i^2\right)^{-1},$$

где $\widehat{e}_i$ – некоторые состоятельные (например, от применения 2ШМНК) остатки. Матрица, пропорциональная этой, тоже годится.

Строим саму оценку. По сравнению с предыдущей нужно поменять только взвешивание:

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

$$= \left(\sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i' \widehat{e}_i^2\right)^{-1} \sum_{i=1}^{n} z_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i' \widehat{e}_i^2\right)^{-1} \sum_{i=1}^{n} z_i y_i.$$

Это и есть наиболее эффективная ОММ-оценка в классе оценок, использующих условие на моменты $E[ze] = 0$.

Замечание. По сей день от дилетантов можно слышать восторженные замечания об ОММ, вроде «ОММ робастен к условной гетероскедастичности». Эта фраза лишь означает, что ОММ асимптотически эффективен в этих условиях, а 2ШМНК – нет, и отражает факт использования говорящим 2ШМНК в течение десятилетий и внезапно обнаружившим нечто лучшее.

7 ОММ как оптимальное инструментальное оценивание

Взглянем на уже известные вещи с другой точки зрения для выявления интересных закономерностей. Рассмотрим

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

$$= \left(\sum_{i=1}^{n} \widehat{\zeta}_i x_i'\right)^{-1} \sum_{i=1}^{n} \widehat{\zeta}_i y_i',$$

что выглядит как инструментальная оценка при точной идентификации с инструментом $\widehat{\zeta}_i$, который является линейной комбинацией первоначального инструмента:

$$\underset{k\times 1}{\widehat{\zeta}} = \frac{1}{n}\sum_{j=1}^{n} x_j z_j' \left(\frac{1}{n}\sum_{j=1}^{n} z_j z_j' \widehat{e}_j^2\right)^{-1} z = \underbrace{\widehat{Q}_{xz}}_{k\times l}\, \underbrace{\widehat{Q}_{zze^2}^{-1}}_{l\times l}\, \underbrace{z}_{l\times 1}.$$

Отсюда следует следующая интерпретация: эффективная ОММ-оценка берёт первоначальные инструменты z (их l штук), умножает их на выписанную выше последовательность матриц, которые в совокупности имеют размерность$k \times l$, слева. Таким образом берутся, пусть и сложные, линейные комбинации от первоначальных инструментов так, чтобы получился точно идентифицирующий инструмент. Такое взвешивание является оптимальным, ибо мы начали с эффективного ОММ. Таким образом, эффективный ОММ неявным образом «ужимает» информацию до необходимого размера оптимальным образом. Приведённый выше $\widehat{\zeta}$ называется *оптимальным инструментом.* Используя популяционные аналоги, можно получить *идеальный оптимальный инструмент*

$$\zeta = Q_{xz} Q_{zze^2}^{-1} z,$$

который асимптотически эквивалентен $\widehat{\zeta}$.

Даже в самой общей постановке, когда условия на моменты

$$E[m(z, \theta)] = 0$$

нелинейны по параметру, возможна похожая интерпретация. Эффективная ОММ-оценка в этом случае равна

$$\widehat{\theta} = \arg\min_{q \in \Theta} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, q) \right)' \widehat{Q}_{mm}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, q) \right),$$

а условия первого порядка –

$$\underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(z_i, \widehat{\theta})}{\partial q'} \right)'}_{\widehat{Q}'_{\partial m}} \widehat{Q}_{mm}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, \widehat{\theta}) \right) = 0,$$

или

$$\frac{1}{n} \sum_{i=1}^{n} \widehat{Q}'_{\partial m} \widehat{Q}_{mm}^{-1} m(z_i, \widehat{\theta}) = 0.$$

Эта система точно идентифицирована, несмотря на то, что первоначальная задача была сверхидентифицирована. Объект, стояший под знаком суммы, можно считать преобразованной функцией моментов:

$$\frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}(z_i, \widehat{\theta}) = 0,$$

где $\widehat{\mu}(z, \theta) = \widehat{Q}'_{\partial m} \widehat{Q}_{mm}^{-1} m(z, \theta)$ называется *оптимальнй функцией моментов.* Поскольку у нас использовался эффективный ОММ (эффективная взвешивающая матрица $\widehat{Q}_{mm}^{-1}$), такое преобразование оптимально.

Итак, с условиями на моменты происходит примерно то же, что и с инструментами в линейном случае: их более чем достаточно, l штук, но информацию в них можно сжать оптимальным образом до k элементов. Здесь тоже немного мешают «крышечки», и можно ввести понятие *идеальной оптимальной функции моментов*:

$$\mu(z,\theta) = Q'_{\partial m} Q_{mm}^{-1} m(z,\theta).$$

К сожалению, популяционные аналоги неизвестны, но именно такое соотношение лучше всего характеризует оптимальное сжатие информации из l элементов в k оптимальным образом.

8 МНК, ОМНК и ММП как ОММ-оценивание

Каждая из изученных нами оценок так или иначе может рассматриваться как одна из класса ОММ-оценок, если соответствующим образом сформировать условия на моменты.

В случае линейной регрессии мы создавали оценку из условия нескоррелированности регрессоров и ошибок

$$E[xe] = 0.$$

Применив принцип аналогий, получим

$$\frac{1}{n}\sum_{i=1}^{n} x_i \widehat{e}_i = 0,$$

т.е. условие для МНК-оценки. Эту же цепочку можно рассматривать как применение ОММ, где параметров в β столько же, сколько x-ов, или же КММ из-за точной идентификации.

Рассмотрим следующее условие на моменты, вытекающее из регрессионного предположения:

$$E\left[\frac{xe}{\sigma^2(x)}\right] = 0.$$

Применение ОММ к этому условию на моменты, а у нас опять точная идентификация, есть применение КММ. Принцип аналогий дает систему уравнений, из которой находится (недоступная) ОМНК-оценка:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{x_i \widehat{e}_i}{\sigma^2(x_i)} = 0.$$

Несколько сложнее с ММП. Здесь в качестве условия на моменты нужно взять правило нулевого ожидаемого скор (математическое ожидание скор-функции в истинном параметре равно 0):

$$E\left[\frac{\partial \log f(z|\theta)}{\partial q}\right] = 0.$$

Заметим, что опять имеем точно идентифицирующую параметр систему. Это означает, что применение ОММ опять вырождается в применение КММ, т.е. достаточно применить принцип аналогий:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(z|\widehat{\theta})}{\partial q}=0.$$

Поскольку полученные уравнения – это условия первого порядка для оценки максимального правдоподобия, их решение есть ММП-оценка.

Таким образом, класс ОММ-оценок очень широк. Даже если параметры идентифицируются не с помощью среднего, а с помощью, например, медианы, эти условия можно преобразовать в привычные условия на моменты. Например, в случае медианы, это математическое ожидание знака отклонения от параметра равно 0. Правда, для медианы нужно модифицировать ОММ-теорию, ибо мы использовали предположение дифференцируемой функции моментов. Оценки получатся состоятельными и асимптотически нормальными даже в таких случаях.

9 ОММ и модели с рациональными ожиданиями

Оказывается, что ОММ-теория остается верной и для моделей стационарных эргодических временных рядов. Правда, стационарность может быть немного ослаблена (например, дисперсия может слегка расти или флуктуировать во времени), но присутствия единичных корней допускать нельзя. Мы для определенности предполагаем стационарность и эргодичность.

Отличия от кросс-секционных моделей примерно те же самые, что встречались и в линейных моделях. В первую очередь присутствовала проблема оценивания бесконечной суммы в «сердцевине» выражения для асимптотической дисперсии. Похожая проблема имеется и здесь.

Под обозначением Q_{mm} теперь мы будем иметь в виду *долгосрочную* дисперсию

$$Q_{mm}=\sum_{j=-\infty}^{+\infty}E[m(z_t,\theta)m(z_{t-j},\theta)'].$$

Теперь эта матрица нужна нам не только для построения стандартных ошибок, но и для оптимального взвешивания. В общем случае, не имея информации о структуре серийной корреляции, эта матрица может быть оценена по формулам Ньюи-Веста или Андрюса. Напомним суть этих методов: оценивается конечное количество слагаемых, а границы растут с увеличением выборки, но медленно, а проблема положительной определенности матрицы Q_{mm} (предельно важная для ОММ-оценивания) решается дополнительным взвешиванием.

Может ли Q_{mm} состоять из одного «центрального» слагаемого? Да, если $m(z,\theta)$ есть последовательность мартингальных приращений (ПМП) по отношению к своему прошлому, то есть

$$E[m(z_t,\theta)|I_{t-1}] = 0.$$

Тогда $\forall\ j > 0$ (для отрицательных j логика та же),

$$E[m(z_t,\theta)m(z_{t-j},\theta)'] = E[\underbrace{E[m(z_t,\theta)|I_{t-1}]}_{0}m(z_{t-j},\theta)'] = 0.$$

Как следствие, Q_{mm} состоит всего из одного слагаемого, как в кросс-секционных моделях:

$$Q_{mm} = E[m(z_t,\theta)m(z_t,\theta)'].$$

Как следствие, ничего не меняется в ОММ-теории.

Если $m(z_t,\theta)$ серийно скоррелированно до известного порядка q, то

$$Q_{mm} = \sum_{j=-q}^{q} E[m(z_t,\theta)m(z_{t-j},\theta)'].$$

С этим объектом легче иметь дело, чем с бесконечной суммой. Каждое слагаемое можно состоятельно оценить и взять их сумму. Такой способ приводит к оценке Хансена-Ходрика. Но такая матрица может не оказаться положительно определенной. Если возникла такая проблема, то можно поменять стратегию и забыть об ограниченности серийной корреляции и вернуться к общему случаю, т.е. формулам Ньюи-Веста или Андрюса, что, конечно, неприятно из-за частичной потери информации.

Рассмотрим типичную нелинейную макроэкономическую модель

$$\begin{aligned} &\max E_0 \sum_{t=0}^{\infty} \beta^t u(c_t,\gamma) \\ &\text{s.t.} \quad c_t + b_t \leqslant y_t + R_t b_{t-1} \end{aligned}$$

где процессы для y_t, R_t считаются заданными, β – фактор дисконтирования, а γ – параметр предпочтений. Условие первого порядка (уравнение Эйлера) выглядит так:

$$E\left[\beta \frac{u'(c_{t+1},\gamma)}{u'(c_t,\gamma)} R_{t+1}|I_t\right] = 1.$$

До изобретения ОММ, то есть примерно до 80-х годов, поступали следующим образом. Делалось предположение, что $R_{t+1} = R$, а функция полезности квадратична:

$$u(c,\gamma) = \gamma_1 c + \frac{\gamma_2}{2}c^2.$$

Такая модель называется линейно-квадратичной. Тогда

$$u'(c,\gamma) = \gamma_1 + \gamma_2 c,$$

$$E\left[\beta\frac{\gamma_1+\gamma_2 c_{t+1}}{\gamma_1+\gamma_2 c_t}R|I_t\right]=1,$$

откуда вытекает линейная регрессия

$$c_{t+1}=\alpha_1+\alpha_2 c_t+e_{t+1},\quad E[e_{t+1}|I_t]=0.$$

Оценивание модели с использованием ОММ началось со статьи Хансена и Синглетона 1982-го года. Делается предположение о функции полезности CRRA-типа:

$$u(c,\gamma)=\frac{c^{1-\gamma}}{1-\gamma},\quad \gamma>0.$$

Маржинальная полезность равна $u'(c,\gamma)=c^{-\gamma}$. Тогда уравнение Эйлера переписывается как

$$E\left[\beta\left(\frac{c_t}{c_{t+1}}\right)^{\gamma}R_{t+1}-1|I_t\right]=0,$$

где $\theta=(\beta,\gamma)$ – оцениваемый параметр. Разность, которая стоит под знаком условного математического ожидания, можно интерпретировать, как ошибку планирования e_{t+1}. Получилось нечто очень похожее на нелинейную регрессию, хотя нет явного разделения на «зависимую» и «независимые» переменные.

Чтобы перейти от условных к безусловным математическим ожиданиям, выберем инструментальные переменные. Инструментов должно быть как минимум 2, поскольку размерность параметра $k=2$. Например, годится

$$z_t=\begin{pmatrix}1 & R_t & R_{t-1} & \frac{c_t}{c_{t-1}} & \frac{c_{t-1}}{c_{t-2}}\end{pmatrix}'.$$

На практике обычно используют именно такие инструменты. То, что в качестве инструментов взят рост потребления $\frac{c_t}{c_{t-1}}$ и т.п., а не, скажем, само потребление c_t, связано с требованием стационарности: чистое потребление обычно растёт во времени. В принципе, в качестве инструментов могут быть использованы рост доходов $\frac{y_t}{y_{t-1}}$ и рост инвестиций $\frac{b_t}{b_{t-1}}$, фигурирующие в бюджетном ограничении, ибо они тоже релевантны. Могут быть использованы любые стационарные комбинации, типа $\frac{c_{t-2}}{c_t}$, R_t^2, R_tR_{t-1}, и т.п. Пока алгоритма по их оптимальному выбору, увы, не существует. Однозначно следует включить константу. Увлекаться количеством инструментов не стоит, ибо замечено, что при их накоплении поведение ОММ-оценок всё менее точно описывается их асимптотическим распределением.

После того, как мы выбрали инструменты, можно сформулировать условия на моменты в привычном виде

$$E\left[z_t\underbrace{(\beta\left(\frac{c_t}{c_{t+1}}\right)^{\gamma}R_{t+1}-1)}_{e_{t+1}}\right]=0$$

и прогнать эффективный ОММ (здесь $l = 5$, $k = 2$), что подразумевает следующие действия:

1. Оценивание параметров γ и β.

2. Оценивание асимптотической дисперсионной матрицы и построение стандартных ошибок для обеих оценок.

3. Проведение теста на спецификацию (J-теста) с обязательным принятием гипотезы. Если она отвергается, следует изменить модель, приведя её ближе к действительности.

При этом нам необходимо организовать эффективное взвешивание, для чего нужна матрица Q_{mm}. Так как $z_t e_{t+1}$ являются ПМП по отношению к своему прошлому, матрица Q_{mm} имеет простую структуру, т.е. долгосрочная дисперсия состоит из одного лидирующего слагаемого – чистой дисперсии. Можно сказать, что нам повезло, но так «везет» в большинстве подобных задач. Это следствие того, что большинство подобных задач являются однопериодными, где решения принимаются в каждый момент времени касательно следующего момента времени.

Посмотрим теперь, когда такие осложнения могут возникнуть. Поменяем одно из условий первоначальной модели. Пусть теперь вложения реализуются через $p > 1$ периодов времени, т.е. решения принимаются касательно момента времени через p периодов. Бюджетное ограничение записывается как

$$c_t + b_t \leqslant y_t + R_t b_{t-p},$$

а уравнение Эйлера – как

$$E\left[\beta^p \frac{u'(c_{t+p},\gamma)}{u'(c_t,\gamma)} R_{t+p} | I_t\right] = 1.$$

Выбрав инструменты, имеем

$$E\left[z_t \underbrace{(\beta^p \left(\frac{c_t}{c_{t+p}}\right)^{\gamma} R_{t+p} - 1)}_{e_{t+p}}\right] = 0,$$

и прогоняем эффективный ОММ. Теперь Q_{mm} имеет более сложную структуру, т.к. $z_t e_{t+p}$ не является ПМП по отношению к своему прошлому, так что многопериодность модели приводит к наличию серийной корреляции. Из-за конечности горизонта планирования Q_{mm} все-таки будет конечной суммой:

$$Q_{mm} = \sum_{j=-(p-1)}^{p-1} E[z_t z_{t-j} e_t e_{t-j}].$$

Для оценивания мы можем применить формулу Хансена–Ходрика, состоятельно оценив все слагаемые. В случае, если не получилась положительно определённая матрица, можно её забраковать и использовать более общие методы (формулу Ньюи–Веста и т.п.).

Рассмотренные примеры типичны для применения ОММ в макроэкономических нелинейных моделях. Некоторые же модели линейны; они следуют не из задачи с оптимизирующими агентами, а из других соображений. Примерами являются модель Мишкина предсказывания будущей инфляции по текущим процентным ставкам и модель Хансена–Ходрика предсказывания изменений текущих обменных курсов по форвардным премиям. Для линейных моделей ОММ даст явную формулу для оценки.

10 Бутстрапирование ОММ-оценок

Напомним, что бутстрап – это альтернативный асимптотике способ построения статистических выводов. Бутстрап используется, когда нас не устраивает асимптотическое приближение из-за его неточности. А распределения ОММ-оценки и особенно J-статистики как раз плохо приближаются их асимптотическими распределениями.

Итак, генерируем бутстраповские псевдовыборки

$$\{z_1, \ldots, z_n\} \begin{array}{l} \nearrow \quad \{z_1^*, \ldots, z_n^*\}_1 \\ \longrightarrow \qquad \ldots \\ \searrow \quad \{z_1^*, \ldots, z_n^*\}_B \end{array}$$

где B – большое число. Из исходных данных получаем статистики $\widehat{\theta}$, W, $J, \ldots$. Для каждой бутстраповской псевдовыборки также получаем аналоги

$$\begin{array}{c} \widehat{\theta}_1^*,\ W_1^*,\ J_1^*, \ldots \\ \vdots \\ \widehat{\theta}_B^*,\ W_B^*,\ J_B^*, \ldots \end{array}$$

То, как легли значения этих статистик, мы воспринимаем как их бутстраповское распределение.

Напомним про необходимость рецентрирования бутстраповских статистик. Например для t-статистики

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

бутстраповский аналог будет выглядеть как

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

Построим бутстраповский аналог для ОММ-оценки

$$\widehat{\theta} = \arg\min_{q \in \Theta} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, q) \right)' \widehat{Q}_{mm}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, q) \right).$$

Просто заменить данные на псевдоданные мало, так как из-за сверхидентификации

$$E^*[m(z^*, \theta^*)] = \frac{1}{n} \sum_{i=1}^{n} m(z_i, \widehat{\theta}) \neq 0$$

в выборке, в то время как в популяции

$$E[m(z, \theta)] = 0.$$

Очевидно, необходимо дополнительное рецентрирование. Бутстраповский аналог оценки есть

$$\widehat{\theta}^* = \arg\min_{q \in \Theta} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i^*, q) - \frac{1}{n} \sum_{i=1}^{n} m(z_i, \widehat{\theta}) \right)' \widehat{Q}_{mm}^{*-1} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i^*, q) - \frac{1}{n} \sum_{i=1}^{n} m(z_i, \widehat{\theta}) \right).$$

Сделаем то же самое для J-статистики. Сама J-статистика есть

$$J = n \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, \widehat{\theta}) \right)' \widehat{Q}_{mm}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} m(z_i, \widehat{\theta}) \right),$$

а её бутстраповский аналог –

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

При рецентрировании бутстрап предоставляет *асимптотическое рафинирование* по сравнению с асимптотическим распределением, то есть то самое свойство, из-за которого бутстрап и используется: большая точность приближения, чем с помощью асимптотического распределения. При этом под точностью имеется в виду скорость сходимости. Без рецентрирования целевой функции бутстраповская процедура остается состоятельной, но асимптотическое рафинирование теряется.

11 Поведение ОММ-оценок в конечных выборках

Замечено, что двухступенчатая ОММ-оценка обладает плохими свойствами в маленьких выборках, иногда бывая даже менее эффективной, чем асимптотически неэффективные ОММ-оценки (например, с равномерным взвешиванием $W = I$). Причины этого – это, во первых, то что на первом шаге мы оцениваем порядка l^2 элементов матрицы Q_{mm}, что сложно сделать точно, особенно когда имеется много условий на моменты, и во-вторых, на обоих шагах используется одна и та же выборка.

Как решать эту проблему? В литературе встречались следующие предложения:

- Разбить выборку: $n = n_1 + n_2$, и использовать подвыборку n_i на i-м шаге. Недостаток такого предложения в том, что часть выборки уходит на предварительную процедуру. Как следствие, такая оценка асимптотически неэффективна.

- Суть *итеративного ОММ (ИОММ)* – расширение двухшаговой процедуры до многошаговой: повторить 2-й шаг со взвешивающей матрицей, которая использует не предварительную, а эффективную ОММ-оценку. Можно повторять итерации до сходимости (если процедура сойдется). Этот метод характеризуется следующими условиями первого порядка для самой последней итерации:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\widehat{\theta}_I)}{\partial q}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}m(z_i,\widehat{\theta}_I)m(z_i,\widehat{\theta}_I)'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}m(z_i,\widehat{\theta}_I)=0.$$

  Оценка $\widehat{\theta}_I$ менее смещена, чем $\widehat{\theta}$. Асимптотические свойства ИОММ, конечно же, те же самые, что и ОММ.

- *Постоянно обновляющийся ОММ* решает задачу

$$\widehat{\theta}_{CUP} = \arg\min_{q\in\Theta}\left(\frac{1}{n}\sum_{i=1}^{n}m(z_i,q)\right)'\left(\frac{1}{n}\sum_{i=1}^{n}m(z_i,q)m(z_i,q)'\right)\left(\frac{1}{n}\sum_{i=1}^{n}m(z_i,q)\right).$$

  Этот метод одношаговый, мы не оцениваем Q_{mm} заранее. В результате получается менее смещённая, чем ОММ, в конечных выборках оценка. Заметим, что из трех перечисленных эффективных оценок наилучшей в смысле наименьшего смещения является последняя.

- *Метод эмпирического правдоподобия.* Оценки метода эмпирического правдоподобия имеют те же асимптотические свойства, что и эффективные ОММ-оценки, но имеют совершенно другую мотивацию. Из-за своей одношаговой природы они обладают лучшими свойствами в конечных выборках, чем ОММ-оценки.

12 Тест Хаусмана на спецификацию модели

Пример 1. Положим, мы специфицировали регрессию, оценили её и интерпретировали результаты. Можно ли протестировать, что оцененное нами уравнение – действительно регрессия? Например, пусть мы имеем линейное уравнение

$$y = x'\beta + e,$$

и мы хотим узнать, является ли $x'\beta$ регрессионной функцией, или же это всего лишь линейная проекция y на пространство x-ов (вспомним, что если это не более чем проекция, её коэффициенты нельзя содержательно интерпретировать). За нулевую мы принимаем более ограничивающую гипотезу, то есть

$$H_0: \quad E[e|x] = 0, \quad H_A: \quad E[ex] = 0.$$

Для тестирования необходимо сконструировать подходящую статистику.

Проиллюстрируем технологию в более абстрактной постановке, а затем применим полученный результат к нашему частному примеру. Пусть специфицированы H_0 и H_A, а θ есть значение истинного параметра при предположении, что нулевая гипотеза верна. Пусть $\widehat{\theta}_0$ и $\widehat{\theta}_1$ – оценки, состоятельные для θ при нулевой гипотезе H_0, но $p\lim \widehat{\theta}_0 \neq p\lim \widehat{\theta}_1$ при альтернативной гипотезе H_A. Этих пределов может вообще не существовать, главное, чтобы разница $p\lim \widehat{\theta}_1 - p\lim \widehat{\theta}_0$ была ненулевой. На основе этой разницы $\widehat{q} = \widehat{\theta}_1 - \widehat{\theta}_0$ мы и строим статистику. При H_0, $\widehat{q} \xrightarrow{p} 0$, а при H_A, $\widehat{q} \xrightarrow{p} p\lim \widehat{\theta}_1 - p\lim \widehat{\theta}_0 \neq 0$.

Пусть при H_0, $\widehat{\theta}_0$ и $\widehat{\theta}_1$ совместно асимптотически нормальны:

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\theta}_0\\ \widehat{\theta}_1\end{pmatrix} - \begin{pmatrix}\theta\\ \theta\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}V_0 & C\\ C' & V_1\end{pmatrix}\right).$$

Тогда по теореме Манна–Вальда имеем

$$\sqrt{n}\widehat{q} \xrightarrow{d} \mathcal{N}(0, V_{\widehat{q}}),$$

где $V_{\widehat{q}} = V_0 + V_1 - C - C'$. Теперь необходимо пивотизировать нашу статистику. Пусть $\widehat{V}_{\widehat{q}}$ – состоятельная оценка матрицы $V_{\widehat{q}}$: $\widehat{V}_{\widehat{q}} \xrightarrow{p} V_{\widehat{q}}$. Отсюда строим статистику для теста Хаусмана:

$$\mathcal{H} = n\widehat{q}'\widehat{V}_{\widehat{q}}^{-1}\widehat{q} \xrightarrow{d} \chi^2_{dim(\theta)}.$$

Понятно, что при H_A значение этой статистики велико, т.к. велико значение самой $\widehat{q}$. Если $\mathcal{H} > q_{1-\alpha}^{\chi^2_{dim(\theta)}}$, мы отвергаем гипотезу о верной спецификации модели; если же $\mathcal{H} < q_{1-\alpha}^{\chi^2_{dim(\theta)}}$, то данные ей не противоречат.

Как построить оценку $\widehat{V}_{\widehat{q}}$? Наибольшие сложности возникают с оцениванием матрицы C; матрицы V_0 и V_1, асимптотические дисперсионные матрицы оценок $\widehat{\theta}_0$ и $\widehat{\theta}_1$, как правило, легко оцениваются. Оказывается, что C необязательно оценивать, если мудро выбрать оценки.

Лемма. Если при нулевой гипотезе $\widehat{\theta}_0$ асимптотически эффективна в некотором линейном классе оценок, которому $\widehat{\theta}_1$ также принадлежит, то $V_{\widehat{q}} = V_1 - V_0$, и поэтому $V_{\widehat{q}}$ может быть оценена как $\widehat{V}_{\widehat{q}} = \widehat{V}_1 - \widehat{V}_0$.

Замечание 1. Класс должен быть линейным, ему должно принадлежать всё, что лежит на отрезке между $\widehat{\theta}_0$ и $\widehat{\theta}_1$.

Интуитивное доказательство. Асимптотическая ковариация $ACov(\widehat{\theta}_0, \widehat{q})$ равно нулю, иначе мы смогли бы улучшить оценку $\widehat{\theta}_0$, а это противоречит её асимптотической эффективности. Тогда

$$V_1 = AVar(\widehat{\theta}_1) = AVar(\widehat{\theta}_0 + \widehat{q}) = V_0 + V_{\widehat{q}} \quad \Rightarrow \quad V_{\widehat{q}} = V_1 - V_0.$$

Замечание 2. Построенная таким образом $\widehat{V_{\widehat{q}}}$ может не быть неотрицательно определенной.

Замечание 3. Матрица $V_{\widehat{q}}$ может быть вырожденной. Тогда нужно использовать «обобщённую обратную» матрицу $\widehat{V_{\widehat{q}}}^{-}$ (A^- называется обобщённой обратной для матрицы A, если $AA^-A = A$).

Итак, подводём итог. Статистика Хаусмана имеет вид и асимптотическое распределение

$$\mathcal{H} = n\widehat{q}'\widehat{V}_{\widehat{q}}^{-}\widehat{q} \xrightarrow{d} \chi^2_{rank(V_{\widehat{q}})}.$$

Благоприятная ситуация для тестирования следующая:

	H_0	H_A
$\widehat{\theta}_0$	состоятельна, асимптотически нормальна и эффективна	несостоятельна
$\widehat{\theta}_1$	состоятельна, асимптотически нормальна, но неэффективна	состоятельна

Неформально говоря, $\widehat{\theta}_0$ правильно работает только при проверяемой спецификации, а при отклонениях должна быть «плохой»; $\widehat{\theta}_1$ является состоятельной при любых обстоятельствах, но она неэффективна (обычно она выбирается робастной к неправильной спецификации модели).

Пример 1 (продолжение). Вернёмся к проверке регрессионности уравнения:

$$H_0: \quad E[e|x] = 0, \quad H_A: \quad E[ex] = 0.$$

Необходимые компоненты для построения теста – вот они:

$$\begin{aligned}
\widehat{\beta}_1 &= \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \sum_{i=1}^n x_i y_i, \\
\widehat{V}_1 &= n \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \sum_{i=1}^n x_i x_i' \widehat{e}_i^2 \left(\sum_{i=1}^n x_i x_i'\right)^{-1}, \\
\widehat{\beta}_0 &= \left(\sum_{i=1}^n \frac{x_i x_i'}{\sigma^2(x_i)}\right)^{-1} \sum_{i=1}^n \frac{x_i y_i}{\sigma^2(x_i)}, \\
\widehat{V}_0 &= n \left(\sum_{i=1}^n \frac{x_i x_i'}{\sigma^2(x_i)}\right)^{-1}.
\end{aligned}$$

Здесь предполагается, что скедастичная функция $\sigma^2(x)$ известна. Тест $\mathcal{H}$ не будет работать (подумайте, почему) при условной гомоскедастичности или если $E\left[\dfrac{xe}{\sigma^2(x)}\right] = 0$ без справедливости регрессионного соотношения, что маловероятно.

Пример 2. Построим тест Хаусмана на экзогенность. Постановка задачи следующая:

$$y = x'\gamma_1 + z'\gamma_2 + e, \quad E[e|z_1, z_2] = 0, \quad E[e^2|z_1, z_2] = \sigma^2,$$

где $\underset{l_1\times 1}{z_1}$ и $\underset{l_2\times 1}{z_2}$ – валидные инструменты, и $l_2 \geqslant k$. Мы хотим протестировать $\underset{k\times 1}{x}$ на экзогенность. Сформируем $H_0 : x$ – экзогенный инструмент, т.е. $E[e|x, z_1, z_2] = 0$, и $H_A : x$ – эндогенный инструмент, т.е. $E[e|x, z_1, z_2] \neq 0$. Подходящие оценки: $(\widehat{\gamma}_1\ \widehat{\gamma}_2)'_0$ – 2ШМНК-оценка, использующая $(x\ z_1\ z_2)'$, и $(\widehat{\gamma}_1\ \widehat{\gamma}_2)'_1$ – 2ШМНК-оценка, использующая $(z_1\ z_2)'$. Можно, однако, придумать ситуации, когда тест Хаусмана не будет работать с таким выбором оценок (например, коллинеарность x и одного из z, и т.п.).

IV Анализ панельных данных

1 Некоторые полезные факты

- Идемпотентная матрица A обладает свойством $A^2 = A$. Симметричная идемпотентная матрица A вдобавок симметрична: $A' = A$. Все подобные матрицы, кроме единичной I_n, имеют редуцированный ранг, то есть вырождены.

- Произведение Кронекера:

$$\underset{n\times k}{A} \otimes \underset{m\times l}{B} = \underset{nm\times kl}{\begin{pmatrix} a_{11}B & \dots & a_{1k}B \\ \vdots & \vdots & \vdots \\ a_{n1}B & \dots & a_{nk}B \end{pmatrix}}.$$

  Свойства произведения Кронекера:

$$\begin{aligned} (A\otimes B)' &= A'\otimes B' \\ (A\otimes B)^{-1} &= A^{-1}\otimes B^{-1} \\ (A\otimes B)(C\otimes D) &= AC\otimes BD \end{aligned}$$

- Проекционная матрица (проектор): Пусть $n\times k$ $(n > k)$ матрица X имеет ранг k. Ассоциированными проекционными матрицами являются

$$\begin{aligned} P &= X(X'X)^{-1}X' \quad (r(P) = k), \\ M &= I_n - P \quad (r(M) = n-k). \end{aligned}$$

Проекторы являются симметричными идемпотентными матрицами. Если у нас есть вектор зависимых переменных y, а X является матрицей независимых переменных, действие проекторов приводит к разложению y на две ортогональные компоненты:

$$Py = X(X'X)^{-1}X'y = X\widehat{\beta} = \widehat{y},$$

$$My = y - \widehat{y} = \widehat{e},$$

а если подействовать этими проекторами на X, получим

$$PX = X, \quad MX = 0.$$

- Сумма квадратов проекционных остатков:

$$RSS = \widehat{e}'\widehat{e} = y'My = y'y - y'X(X'X)^{-1}X'y.$$

- Длинные и короткие регрессии. «Длина» регрессии зависит от количества регрессоров, много их или мало. Можно ли свести длинную регрессию к серии коротких? Ответ на этот вопрос был важен в эпоху слабости вычислительной техники, мы же используем этот результат для других целей. Пусть длинная регрессия будет

$$y = x_1'\beta_1 + x_2'\beta_2 + e,$$

где

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{matrix} \leftarrow & k_1 \times 1 \\ \leftarrow & k_2 \times 2 \end{matrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Считаем, что k_1 и k_2 маленькие, а $k_1 + k_2$ достаточно большое. Пусть $\widehat{\beta}$ – МНК-оценка длинной регрессии, а $\widehat{e}_i = y_i - x_i'\widehat{\beta}$ – её остатки. Мы хотим найти эти величины с помощью коротких регрессий.

Теорема Фриша–Во–Ловелла. Рассмотрим три регрессии:

$$\begin{aligned} y &= x_1'\gamma_1 + u_1, && (1) \\ x_2 &= x_1'\gamma_2 + u_2, && (2) \\ \widehat{u}_1 &= \widehat{u}_2{}'\beta_2 + e. && (3) \end{aligned}$$

Тогда

1. $\widehat{\beta}_2$ и $\widehat{e}_i$ можно получить как МНК-оценки и МНК-остатки регрессии (3) МНК-остатков регрессии (1) y на x_1 на МНК-остатках регрессии (2) x_2 на x_1.

2. Кроме того, $\widehat{\beta}_1$ можно получить как МНК-оценки регрессии $y - x_2'\widehat{\beta}_2$ на x_1.

Следствие. Если x_1 – константа, то $\widehat{\beta}_2$ можно получить от регрессии $y - \overline{y}$ на $x_2 - \overline{x}_2$, при этом $\widehat{\beta}_1 = \overline{y} - \overline{x}_2'\widehat{\beta}_2$.

Заметим, что в теореме и следствии из неё имеется в виду численная идентичность оценок.

2 Структура панельных данных

Пусть имеется выборка $\{y_{it}, x_{it}\}$, где $i = 1, \ldots, n$ – индекс объектов, а $t = 1, \ldots, T$ – индекс моментов времени. Видим, что в структуре панельных данных есть признаки и кросс-секций, и временных рядов. В то же время анализ панельных данных больше похож на анализ первых, нежели последних.

Если под i подразумеваются одни и те же объекты для всех t, то у нас *панель*. Если под i подразумеваются различные объекты для разных t, то у нас *псевдопанель*. Если имеются наблюдения по каждому $i = 1, \ldots, n$ и каждому $t = 1, \ldots, T$, то у нас *сбалансированная панель*. Если для некоторых i и/или t отсутствуют данные, то у нас *несбалансированная панель*. Для несбалансированных панелей анализ практически такой же, как и для сбалансированных, нужно лишь внимательно следить за пропусками данных. Мы сосредоточимся на сбалансированных панелях.

Вот преимущества использования панельных данных перед кросс-секциями и временными рядами:

- Достигается большая эффективность оценок (размер выборки nT);
- Использование панелей позволяет контролировать и тестировать индивидуальную неоднородность (неоднородность по объектам);
- Использование панелей позволяет идентифицировать эффекты, не обнаруживаемые в кросс-секциях.

Вот два классических примера, подчёркивающих последний пункт:
Пример 1. Женская занятость. Пусть 50% женщин работают, 50% не работают. При наличии кросс-секции, неясно, что верно: (1) половина женщин всегда работает, а другая половина всегда не работают, или же (2) женщины со временем меняют свой статус, каждая может некоторое время работать, некоторое время не работать. Наличие панели позволяет идентифицировать, какой из двух вариантов имеет место.
Пример 2. Эффект членства в профсоюзе на зарплату. Без панели неясно, зарплата больше там, где есть профсоюз, или же зарплата зависит от того, входит человек в

профсоюз или нет. В панели же можно отследить вход в профсоюз и выход из него и идентифицировать эффект членства в нём.

Асимптотика подразумевается (иногда, правда, мы будем пользоваться точной инференцией, но это вынужденная мера) как $n \to \infty$, а T фиксировано. Фиксированность T мотивирована тем, что большинство панелей короткие, где n большое, а T маленькое (типичная ситуация: в течение ряда лет исследуется множество фирм, при этом $T \sim 3-5$ лет). Как следствие, например, тестирование на единичные корни происходит не по Дики-Фуллеру.

Базовая линейная панельная модель имеет вид

$$y_{it} = \alpha + x_{it}'\beta + u_{it},$$

где $i = 1, \ldots, n$, $t = 1, \ldots, T$, $x_{it} \sim i.i.d.$ по объектам, для любого фиксированного t. Структура ошибки u_{it} подчиняется *модели составной ошибки (МСО)*, что означает, что ошибка состоит из нескольких компонент:

$$u_{it} = \mu_i + v_{it} \quad \text{– однонаправленная МСО,}$$

$$u_{it} = \mu_i + \lambda_t + v_{it} \quad \text{– двунаправленная МСО,}$$

где *индивидуальные эффекты* μ_i, *временные эффекты* λ_t и *идиосинкратические компоненты* v_{it} взаимно независимы. Предполагается, что

$$\mu_i \sim i.i.d.(0, \sigma_\mu^2), \quad \lambda_t \sim i.i.d.(0, \sigma_\lambda^2), \quad v_{it} \sim i.i.d.(0, \sigma_v^2).$$

Заметим, что v_{it} независимы по обоим индексам.

Индивидуальные (иногда и временные) эффекты можно трактовать как *фиксированные* и *случайные*. В первом случае мы их обрабатываем как случайные величины, во втором – как реализации случайных переменных.

3 Однонаправленная МСО с фиксированными эффектами

При фиксированных эффектах мы трактуем компоненты ошибки μ_i как фиксированные неизвестные параметры. А раз это параметры, они из компонент ошибки перемещаются в регрессионную функцию, в ошибке остаётся только идиосинкратическая компонента, и модель преобразуется в

$$y_{it} = \alpha + \mu_i + x_{it}'\beta + v_{it},$$

где $i = 1, \ldots, n$, $t = 1, \ldots, T$, и $v_{it} \sim i.i.d.(0, \sigma_v^2)$ независимы от x_{it}. Независимость v_{it} от регрессоров – это необязательное предположение, его можно ослабить. Такое сильное предположение сделано исключительно из соображений простоты анализа.

Итак, регрессионная функция получилась достаточно сложной – константа и новые фиктивные переменные

$$\begin{cases} 1, & \text{для } i\text{-го объекта}, \\ 0, & \text{для остальных}, \end{cases}$$

коэффициентами у которых служат μ_i. Сразу можно сказать, что после того, как мы присвоили индивидуальным эффектам статус регрессоров, у нас появилась мультиколлинеарность, поскольку в сумме все фиктивные переменные повторяют константу. В данной модели принято избавляться от константы, хотя возможны другие варианты. Можно посмотреть на этот вопрос по-другому. Индивидуальные эффекты μ_i пришли из ошибки, где от них требовалась центрированность вокруг нуля. Сейчас же мы этого не требуем, и α поглощается ими. Избавившись таким образом от одного регрессора, запишем систему в векторной и матричной формах.

В векторной форме всё связанное с объектом будет содержать один вектор:

$$y_i = i_T\mu_i + X_i\beta + v_i, \quad i = 1, \ldots, n,$$

где

$$\underset{T\times 1}{y_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad \underset{T\times 1}{i_T} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \underset{T\times k}{X_i} = \begin{pmatrix} X'_{i1} \\ X'_{i2} \\ \vdots \\ X'_{iT} \end{pmatrix}, \quad \underset{T\times 1}{v_i} = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{pmatrix}.$$

От индекса t мы избавились, теперь избавимся от индекса i, и получится матричная форма. Она нужна нам для того, чтобы приступить к знакомому нам МНК-оцениванию, хотя и на скалярную форму МНК-оценки нам также придётся переключаться.

Итак, матричная форма имеет вид:

$$y = \mathcal{D}\mu + X\beta + v,$$

где

$$\underset{nT\times 1}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{n\times 1}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \underset{nT\times 1}{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

$$\underset{nT\times k}{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \qquad \underset{nT\times n}{\mathcal{D}} = \begin{pmatrix} i_T & 0 & 0 & \ldots & 0 \\ 0 & i_T & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & i_T \end{pmatrix}.$$

Матрица $\mathcal{D}$ называется *матрицей индивидуальных фиктивных переменных.* Матрица $\mathcal{D}$ слишком большая, поэтому воспользуемся операцией произведения Кронекера чтобы записать её компактнее:

$$\mathcal{D} = I_n \otimes i_T.$$

Первое, что напрашивается для оценивания – это МНК, и в данном случае он действительно является хорошим методом: в ошибке остались только идиосинкратические компоненты, дисперсионная матрица которых диагональна с одним и тем же элементом σ_v^2 на главной диагонали. Поэтому МНК-оценки μ и β будут эффективными. Но для этого (в частности, даже для существования оценки) необходимо, чтобы $nT \times (n+k)$-матрица всех регрессоров $(\mathcal{D}\ X)$ имела полный ранг, равный количеству столбцов. Из этого, в частности, следует, что:

1. T не должно быть меньше 2, иначе индивидуальные эффекты полностью неидентифицируемы;

2. X не должен содержать переменные, не изменяющиеся во времени (например, образование, если в выборке только взрослые люди).

Итак, если всё в порядке с условием на ранг, можно успешно применить МНК. Возникают, правда, другие «но» (к приведённой мотивации в наше время хорошо развитой компьютерной техники можно придраться, но такова предыстория современной теории):

- Мы не хотим хранить *огромные* (размерности $\sim nT \times n$) матрицы;

- Мы не хотим обращать *большие* (размерности $\sim n \times n$) матрицы;

- Мы больше заинтересованы в оценивании β, нежели μ.

Поэтому, вместо того чтобы посоветовать на практике использовать «в лоб» МНК, мы ищем обходные манёвры. На самом деле, найдя эти обходные манёвры, мы не только избавимся от необходимости работать с большого размера матрицами, но и приобретём массу интуитивных соображений. У нас уже есть инструментарий – теорема Фриша–Во–Ловелла.

Построим проекционную (на фиктивные переменные) матрицу Q:

$$\underset{nT\times nT}{Q} = I_{nT} - \mathcal{D}(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'.$$

Мы можем прогнать две короткие регрессии, X и y, на эти фиктивные переменные, и получить

$$\begin{aligned} Qy \quad &- \quad \text{остатки от регрессии } y \text{ на } \mathcal{D}, \\ QX \quad &- \quad \text{остатки от регрессии } X \text{ на } \mathcal{D}. \end{aligned}$$

Тогда, согласно теореме Фриша–Во–Ловелла, мы можем получить оценку β с помощью МНК в регрессии Qy на QX:

$$\widehat{\beta} = ((QX)'QX)^{-1}(QX)'Qy = (\underbrace{X'QX}_{k\times k})^{-1}X'Qy.$$

Мы избавились от необходимости обращать большие матрицы: вместо матрицы $(n+k)\times(n+k)$ нам теперь нужно обращать матрицу $k\times k$. Кроме того, мы оценили именно то β, которое нас интересовало.

Построенная таким образом оценка $\widehat{\beta}$ имеет несколько названий: оценка наименьших квадратов с фиксированными эффектами (ФЭНК) и оценка наименьших квадратов с фиктивными переменными (НКФП). Самое главное название же выяснится, когда мы взглянем на оценку с другой перспективы.

Найдём вид матрицы Q:

$$\begin{aligned}\mathcal{D}'\mathcal{D} &= (I_n\otimes i_T)'(I_n\otimes i_T) = (I_n\otimes i_T')(I_n\otimes i_T) = \\ &= I_nI_n\otimes i_T'i_T = I_n\otimes T = TI_n.\end{aligned}$$

Тогда

$$\mathcal{D}(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}' = \frac{1}{T}(I_nI_n\otimes i_Ti_T') = I_n\otimes\frac{1}{T}J_T,$$

где $J_T = i_Ti_T'$ – матрица из единиц, и

$$Q = I_{nT} - I_n\otimes\frac{1}{T}J_T.$$

Теперь гораздо прозрачнее видна структура матрицы Q, и не только самой матрицы, но и преобразований, совершаемых с её помощью. Теперь понятно, что взятие остатков в коротких регрессиях означает удаление средних по времени. Например,

$$Qy = y_{it} - \overline{y}_i = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}.$$

Значит, оценивание β сводится к применению МНК к преобразованной системе

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)'\beta + v_{it} - \overline{v}_i, \quad i = 1,\ldots,n, \quad t = 1,\ldots,T,$$

где

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \quad \overline{v}_i = \frac{1}{T}\sum_{t=1}^{T} v_{it}.$$

Такое преобразование, удаляющее из характеристик каждого объекта их среднее по времени, называется преобразованием «внутри». С помощью внутри-преобразования мы избавились от индивидуальных эффектов. Таким образом, если нас интересуют только параметры β, мы можем их оценить с помощью МНК на внутри-преобразованных уравнений. Численно получится то же самое, что получилось бы, если бы мы применили МНК к первоначальной системе, но без всяких проблем с необходимостью обращения больших матриц и заведения огромных. Соответственно, третье название для оценки $\widehat{\beta}$ – это *внутри-оценка.*

Рассмотрим свойства внутри-оценки. Её условная дисперсия есть

$$Var(\widehat{\beta}|X) = \sigma_v^2 (X'QX)^{-1}.$$

Дисперсию σ_v^2 можно оценить, взяв сумму квадратов остатков и разделив на количество наблюдений, скорректированное на количество степеней свободы, т.е. на количество параметров, равное $n + k$. Многие теоретики и практики считают эту корректировку очень важным фактором в анализе панельных данных, так как она получается довольно существенной из-за того, что n большое, а T маленькое.

$$\widehat{\sigma}_v^2 = \frac{RSS}{nT - n - k},$$

где

$$RSS = y'Qy - y'QX(X'QX)^{-1}X'Qy.$$

Последнее выражение, конечно, лучше вычислять не «в лоб», а с помощью внутри-регрессии, чтобы не иметь дело с матрицей Q явным образом.

Почему МНК-оценка остается хорошей для внутри-преобразованной системы? Ведь ошибка тоже преобразовывается, да так, что ковариационная матрица в модели вообще получается вырожденной (состоящая из блоков на главной диагонали, каждый из которых заполнен одним и тем же числом). Оказывается, что мы имеем дело с одним из тех частных случаев, когда ОМНК-оценка в точности совпадает с МНК-оценкой.

Далее. Мы оценили только коэффициенты β, но если мы хотим оценить ещё и μ, мы можем воспользоваться второй частью теоремы Фриша–Во–Ловелла, которая говорит, что МНК-оценки оставшихся коэффициентов можно получить с помощью ещё одной короткой регрессии на тех вспомогательных регрессорах, от которых мы избавляемся. В данном случае это фиктивные переменные, так что прорегрессировав на них разницу между левой частью и подогнанным значением части, связанной с x-ми, получим

$$\widehat{\mu} = (\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}'(y - X\widehat{\beta}),$$

что эквивалентно расчёту каждого индивидуального эффекта по средним для этого объекта:

$$\widehat{\mu}_i = \overline{y}_i - \overline{x}_i'\widehat{\beta}.$$

Какими асимптотическими свойствами обладают построенные нами оценки? Оценка β и оценки μ_i сильно отличаются по свойствам:

- Внутри-оценка β состоятельна и асимптотически нормальна при $n \to \infty$ и фиксированном T. Поэтому, если мы построим стандартные ошибки, исходя из вышеуказанной формулы для дисперсии, найдем t-статистики делением самих оценок на их стандартные ошибки, всю инференцию для β можно проводить стандартным образом.

- Внутри-оценки μ_i несостоятельны при $n \to \infty$ и фиксированном T, ибо их количество растёт пропорционально с ростом размера выборки.

Рассмотрим вопрос тестирования на индивидуальные эффекты. Работая с панельными данными, мы можем не только контролировать индивидуальные эффекты, но и тестировать их наличие. Получив оценки, мы хотели бы задаться вопросом, есть ли вообще неоднородность между объектами, или, возможно, её не стоит контролировать. В качестве нулевой гипотезы возьмём неотличимость индивидуальных компонент между собой:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n.$$

К сожалению, асимптотический тест здесь невозможен из-за того, что оценки $\widehat{\mu}_i$, равенство которых мы собираемся проверять, несостоятельны, а количество ограничений $n-1$ пропорционально размеру выборки и растёт в асимптотике. Формальные тесты для $\widehat{\mu}_i$ проводить можно, но не асимптотические, а с помощью точного подхода, которого мы всегда избегали из-за заложенной в нём необходимости предполагать распределение данных. Тем не менее, если мы очень хотим этот тест провести, воспользуемся тем фактом, что при нормальности ошибок v_{it},

$$F = \frac{(RSS^R - RSS^U)/(n-1)}{RSS^U/(nT-n-k)} \sim F_{n-1,nT-n-k},$$

где RSS^U просчитывается по полной модели (выписанная ранее формула), RSS^R просчитывается по модели с удалёнными индивидуальными компонентами (как бы по одной большой кросс-секции с наличием константы), а $F_{\cdot,\cdot}$ – это распределение Фишера с соответствующими степенями свободы.

4 Двунаправленная МСО с фиксированными эффектами

В двунаправленной модели, кроме индивидуальных, присутствуют еще временные эффекты. Поскольку они фиксированы, и μ_i, и λ_t обрабатываются как неизвестные параметры. Соответственно, в регрессионную функцию добавляются ещё регрессоры:

$$y_{it} = \alpha + \mu_i + \lambda_t + x_{it}'\beta + v_{it}.$$

Необходимо избавиться от мультиколлинеарности среди фиктивных переменных. Теперь константа дублируется дважды, как сумма всех μ_i и как сумма всех λ_t. В данной модели принято α оставлять, а избавляться от одного из временных и одного из индивидуальных эффектов, для определённости, скажем, последних по счёту:

$$\mu_n = 0, \quad \lambda_T = 0.$$

Матричная форма выглядит следующим образом:

$$\underset{nT\times 1}{y} = \underset{nT\times 1}{i_{nT}}\,\alpha + \mathcal{D}_\mu \mu + \mathcal{D}_\lambda \lambda + X\beta + v,$$

где

$$\underset{(n-1)\times 1}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n-1} \end{pmatrix}, \quad \underset{(T-1)\times 1}{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{T-1} \end{pmatrix},$$

$$\underset{nT\times(n-1)}{\mathcal{D}_\mu} = \begin{pmatrix} i_T & 0 & 0 & \dots & 0 \\ 0 & i_T & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & i_T \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \underset{nT\times(T-1)}{\mathcal{D}_\lambda} = \begin{pmatrix} & I_{T-1} & \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ & I_{T-1} & \\ 0 & \dots & 0 \end{pmatrix}.$$

X имеет размерность $nT \times k$, β – $k \times 1$, v – $nT \times 1$, как и в однонаправленной модели.

Объединим матрицы временных и индивидуальных фиктивных переменных в одну матрицу $\mathcal{D}$, а все соответствующие коэффициенты – в вектор γ:

$$\mathcal{D}\gamma = \underbrace{(i_{nT}\ \mathcal{D}_\mu\ \mathcal{D}_\lambda)}_{nT\times(n+T-1)} \underset{(n+T-1)\times 1}{\begin{pmatrix} \alpha \\ \mu \\ \lambda \end{pmatrix}}.$$

Применим МНК к нашей системе уравнений, поскольку, благодаря диагональности дисперсионной матрицы ошибок, ситуация идеальна для МНК. Вновь необходимо гарантировать условие полного ранга (по колонкам) матрицы всех регрессоров ($\mathcal{D}$ X), равного $n + T + k - 1$. Отсюда, в частности, следует, как и для однонаправленной модели, что X не должна содержать переменные, не изменяющиеся во времени и таким образом повторяющие индивидуальные эффекты (например, образование, если в выборке только взрослые). Также X не должна содержать изменяющиеся во времени переменные, одинаковые для всех объектов (например, цены). Если все эти условия выполнены, можно применить МНК, являющийся эффективным методом. Опять же мы не хотим иметь дело с огромными матрицами, поэтому вновь пытаемся применить эквивалентное преобразование и свести дело к серии коротких регрессий. Для этой цели представим Q в виде

$$\begin{aligned} Q &= I_{nT} - \mathcal{D}(\mathcal{D}'\mathcal{D})^{-1}\mathcal{D}' = \\ &= I_{nT} - I_n \otimes \frac{1}{T}J_T - \frac{1}{n}J_n \otimes I_T + \frac{1}{nT}J_{nT}, \end{aligned}$$

где J_k – матрица из единиц размера $k \times k$. Короткая регрессия Qy на QX называется регрессией «внутри». Что означает такое преобразование? Выпишем внутри-оценку:

$$\widehat{\beta} = ((QX)'QX)^{-1}(QX)'Qy = (X'QX)^{-1}X'Qy.$$

Из структуры матрицы Q видно, что $\widehat{\beta}$ есть МНК-оценка на внутри-преобразованных уравнениях

$$y_{it} - \overline{y}_i - \overline{y}_t + \overline{y} = (x_{it} - \overline{x}_i - \overline{x}_t + \overline{x})'\beta + v_{it} - \overline{v}_i - \overline{v}_t + \overline{v},$$

где

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \overline{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{it}, \quad \overline{y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it},$$

и аналогично для x и Qv. Условная дисперсия внутри-оценки равна

$$Var(\widehat{\beta}|X) = \sigma_v^2(X'QX)^{-1}.$$

Асимптотические свойства внутри-оценки: $\widehat{\beta}$ состоятельна и асимптотически нормальна. Состоятельную оценку σ_v^2 можно построить как

$$\widehat{\sigma}_v^2 = \frac{RSS}{nT - n - T - k + 1}.$$

Это правильная корректировка на число степеней свободы. Здесь нужна осторожность, ибо программный пакет, оценивающий внутри-регрессию, не знает о том, что оцениваются еще индивидуальные и временные эффекты. Оценки дисперсии необходимо вручную подправлять.

Про асимптотические свойства оценок остальных параметров говорить трудно. Очевидно, что индивидуальные компоненты оцениваются несостоятельно, поскольку их количество растёт пропорционально размеру выборки. С λ-ми другая ситуация, их фиксированное количество $T-1$, и оно не растёт, т.к. T асимптотически фиксировано. С ними, правда, тоже нужно обходиться осторожно, поскольку присутствует некоторая взаимозаменяемость между μ_i, λ_t и свободным членом.

Если мы хотим протестировать модель на отсутствие только индивидуальных эффектов, нулевая гипотеза выглядит как

$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_{n-1} = 0.$$

Здесь $n-1$ ограничение, поэтому возможен только точный тест. Асимпотический тест невозможен, ибо количество ограничений растёт пропорционально размеру выборки. F-тест строится так же, как в однонаправленной модели.

Если мы хотим протестировать на отсутствие и индивидуальных, и временных эффектов, то нулевой гипотезой будет

$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_{n-1} = \lambda_1 = \lambda_2 = \cdots = \lambda_{T-1} = 0.$$

Ограничений еще больше, $n + T - 2$, асимптотический тест снова невозможен, приходится предполагать нормальность и пользоваться точной инференцией.

Если же мы хотим протестировать на отсутствие только временных эффектов, то

$$H_0: \quad \lambda_1 = \lambda_2 = \cdots = \lambda_{T-1} = 0.$$

Здесь количество ограничений $T - 1$ фиксировано. Точный тест (F-тест), конечно, можно построить, но появляется возможность провести асимптотический χ^2_{T-1}-тест.

5 Однонаправленная МСО со случайными эффектами

Ограничимся только однонаправленной структурой, поскольку двунаправленная означала бы случайные временные эффекты, а это, при обычной короткости панелей по времени, не очень удобно. Фиксированность временных эффектов является более адекватным предположением. В этой модели структура ошибки становится сложной, а структура регрессионной функции – простой:

$$u_{it} = \mu_i + v_{it}, \quad \mu_i \sim i.i.d.(0, \sigma_\mu^2), \quad v_{it} \sim i.i.d.(0, \sigma_v^2).$$

Обе компоненты независимы друг от друга и от всех x-ов. Предположение о полной независимости от x-ов, опять же, введено для упрощения. Сама панельная регрессия:

$$y_{it} = x_{it}'\beta + u_{it}.$$

В модели с фиксированными эффектами мы явно указывали константу α, поскольку она могла повторять фиксированные индивидуальные эффекты. Теперь же последние отсутствуют, и нет смысла указывать константу специально; если она есть, то она уже среди x-ов.

Запишем уравнения в матричной форме:

$$y = X\beta + u.$$

Регрессия выглядит просто, но теперь дисперсионная матрица ошибок имеет сложную структуру. Пусть

$$\Omega \equiv Var(u).$$

На главной диагонали Ω стоят дисперсии

$$Var(u_{it}) = Var(\mu_i + v_{it}) = \sigma_\mu^2 + \sigma_v^2 \quad \forall\ i\ \ \forall\ t.$$

Вне главной диагонали Ω стоят ковариации:

$$Cov(u_{it}, u_{is}) = Cov(\mu_i + v_{it}, \mu_i + v_{is}) = \sigma_\mu^2 \quad \forall\ i\ \ \forall\ t \neq s,$$

$$Cov(u_{it}, u_{js}) = Cov(\mu_i + v_{it}, \mu_j + v_{js}) = 0 \quad \forall\ i \neq j\ \ \forall\ t, s.$$

Заметим, что в двунаправленной МСО со случайными эффектами появилась бы ненулевая корреляция между u_{it} и u_{jt} из-за общего временного эффекта.

В итоге вся дисперсионная матрица выглядит следующим образом:

$$\underset{nT\times nT}{\Omega} = \begin{pmatrix} \underbrace{\begin{matrix} \sigma_\mu^2 + \sigma_v^2 & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_v^2 \end{matrix}}_{T\times T} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \begin{matrix} \sigma_\mu^2 + \sigma_v^2 & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_v^2 \end{matrix} \end{pmatrix}.$$

Каждый блок матрицы размера $T \times T$, всего n ненулевых блоков.

Временно отвлечёмся от структуры дисперсионной матрицы. Посмотрим, какие есть варианты неэффективного оценивания.

- Можно оценить систему с помощью МНК, поскольку u и x независимы:

$$\widehat{\beta}_{OLS} = (X'X)^{-1}X'y,$$

$$Var(\widehat{\beta}_{OLS}|X) = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

- Можно оценить систему с помощью внутри-преобразования:

$$\widehat{\beta}_W = (X'QX)^{-1}X'Qy,$$

$$Var(\widehat{\beta}_W|X) = (X'QX)^{-1}X'Q\Omega QX(X'QX)^{-1}.$$

- Можно оценить систему с помощью так называемого между-преобразования

$$\overline{y}_i = \overline{x}_i'\beta + \mu_i + \overline{v}_i, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T.$$

  Пусть $P = I_{nT} - Q$. Тогда между-оценка равна

$$\widehat{\beta}_B = (X'PX)^{-1}X'Py,$$

$$Var(\widehat{\beta}_B|X) = (X'PX)^{-1}X'P\Omega PX(X'PX)^{-1}.$$

  Между-преобразование является противоположным внутри-преобразованию, т.к. в сумме они дают «пустое» преобразование, означающее применение МНК.

Теперь рассмотрим эффективное ОМНК-оценивание. ОМНК-преобразование, которое нужно при этом осуществить, диктуется матрицей $\Omega^{-\frac{1}{2}}$. Перепишем Ω в более удобном виде:

$$\Omega = \sigma_v^2 I_{nT} + \sigma_\mu^2 I_n \otimes J_T = \sigma_v^2 Q + (\sigma_v^2 + T\sigma_\mu^2)P.$$

Какая польза от такой записи? Дело в том, что матрицы P и Q являются проекторами, они симметричны и идемпотентны по построению; кроме того $QP = PQ = 0$. Фактически мы имеем разложение по базису из двух ортонормированных матриц. Такая запись матрицы Ω называется *спектральным разложением* Ω. Вся прелесть такого разложения в том, что тогда искомая нами матрица $\Omega^{-\frac{1}{2}}$ элементарно находится:

$$\begin{aligned}
\Omega^{-\frac{1}{2}} &= \frac{1}{\sigma_v} Q + \frac{1}{\sqrt{\sigma_v^2 + T\sigma_\mu^2}} P = \\
&= \frac{1}{\sigma_v}(Q + \sqrt{\theta} P) = \\
&= \frac{1}{\sigma_v}(I_{nT} - (1-\sqrt{\theta}) I_n \otimes \frac{1}{T} J_T),
\end{aligned}$$

где

$$\theta = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2}.$$

ОМНК-преобразование, подразумеваемое матрицей $\Omega^{-\frac{1}{2}}$, следующее:

$$y_{it} - (1-\sqrt{\theta})\overline{y}_i = (x_{it} - (1-\sqrt{\theta})\overline{x}_i)'\beta + v_{it} - (1-\sqrt{\theta})\overline{v}_i.$$

ОМНК-оценка $\widehat{\beta}_{GLS}$ есть МНК-оценка на этой ОМНК-преобразованной системе. Эта оценка является наилучшей из несмещённых, она состоятельна и асимптотически нормальна. Её условная дисперсия равна

$$Var(\widehat{\beta}_{GLS}|X) = (X'\Omega^{-1}X)^{-1} = ((\Omega^{-\frac{1}{2}}X)'\Omega^{-\frac{1}{2}}X)^{-1}.$$

Чтобы осуществить доступное ОМНК-оценивание, необходима состоятельная оценка θ, т.е. состоятельные оценки компонент ковариационной матрицы σ_μ^2 и σ_v^2. Это можно сделать, например, так: с помощью внутри-регрессии найти

$$\widehat{\sigma}_v^2 = \frac{RSS_W}{nT - n - k + 1},$$

а с помощью между-регрессии –

$$\widehat{\sigma}_\mu^2 + \frac{\widehat{\sigma_v}^2}{T} = \frac{1}{T}\frac{RSS_B}{n-k}.$$

Тогда

$$\widehat{\theta} = \frac{\widehat{\sigma_v}^2}{\widehat{\sigma_v}^2 + T\widehat{\sigma}_\mu^2} = \frac{RSS_W}{RSS_B}\,\frac{n-k}{nT - n - k + 1} \xrightarrow{p} \theta.$$

Доступная ОМНК-оценка $\widehat{\beta}_{FGLS}$ является МНК-оценкой системы

$$y_{it} - (1 - \sqrt{\widehat{\theta}})\overline{y}_i = (x_{it} - (1 - \sqrt{\widehat{\theta}})\overline{x}_i)'\beta + v_{it} - (1 - \sqrt{\widehat{\theta}})\overline{v}_i.$$

Доступная ОМНК-оценка остается состоятельной и асимптотически нормальной, но теряет свойство несмещённости.

При тестировании на индивидуальные эффекты в моделях с фиксированными эффектами под нулевой гипотезой подразумевалось равенство всех соответствующих параметров. Здесь же их отсутствие означает отсутствие у них дисперсии:

$$H_0: \quad \sigma_\mu^2 = 0, \qquad H_A: \quad \sigma_\mu^2 > 0.$$

При нормальности ошибок u из теории точной инференции известно, что

$$\frac{RSS_W}{\sigma_v^2} \sim \chi^2_{nT-n-k+1}, \quad \frac{RSS_B}{\sigma_v^2 + T\sigma_\mu^2} \sim \chi^2_{n-k},$$

причём эти статистики между собой независимы из-за ортогональности внутри- и между-преобразований. При нулевой гипотезе

$$F = \frac{RSS_B/(n-k)}{RSS_W/(nT-n-k+1)} = \frac{1}{\widehat{\theta}} \sim F_{n-k,nT-n-k+1}.$$

Этот тест может быть плох не только из-за того, что он точный, но еще и потому, что тестовая статистика при нулевой гипотезе может быть очень близкой к её значению при альтернативной, если σ_μ^2 и T маленькие, т.е. мощность теста может быть небольшой. Чаще применяется асимптотичесикй тест Брюша-Пагана, строящийся на основе оценок максимального правдоподобия при предположении нормальности.

6 Фиксированные или случайные эффекты?

Обсудим концептуальный вопрос: как правильно моделировать индивидуальные эффекты – как случайные или как фиксированные? До сего момента мы касались этого вопроса только косвенно. Существует два подхода. Кто-то придерживаются одного из них, кто-то – другого, но оба подхода желательно иметь в виду, принимая решение о спецификации модели.

«Философский» подход разграничивает две ситуации – проводится ли *условная* или *безусловная* инференция.

- Инференция условная, если мы хотим строить выводы для той конкретной выборки, которую имеем. В некотором смысле, имеющаяся выборка для нас и есть популяция. Тогда фиксированные эффекты считаются более адекватной спецификацией.

- Инференция безусловная, если имеющиеся данные – это лишь выборка из большой популяции, а мы хотим строить выводы для всей популяции. Тогда более адекватны случайные эффекты.

Пример 1. Пусть имеются данные по фирмам, и выборка содержит данные для IBM, Microsoft, General Electrics и т.п. фирмы, которые единственны в своем роде и которые вряд ли можно назвать случайной выборкой из большой популяции. В этом случае стоит использовать модель с фиксированными эффектами, и наши выводы скорей будут применимы к этим конкретным фирмам, нежели к фирмам в принципе. Если же в выборке много однотипных однородных фирм, то модель со случайными эффектами более адекватна, и наши статистические выводы будут применимы ко всей популяции подобных фирм.

Пример 2. Пусть речь идет о каких-то крупных региональных единицах. Применительно к США это 50 штатов, применительно к России это 89 регионов. Если они все имеются в выборке, то все выводы применимы только к этому конкретному составу (другого-то и нет). При этом мы используем фиксированные эффекты. Если же мы рассматриваем более мелкие подразделения, которыми в США являются, например, графства, а в России – районы, их тысячи, и выборка, скорей всего, содержит их малую толику. Здесь случайные эффекты более приемлимы. А наиболее удачной спецификацией при рассмотрении графств или районов были бы комбинированные эффекты – случайные для самих графств или районов плюс фиксированные для штатов или регионов, в которых те находятся.

Альтернативный, «статистический», подход не делает содержательной разницы между фиксированными и случайными эффектами. Нормальная ситуация – это случайные эффекты, а фиксированные эффекты используются вынужденно, когда случайные эффекты не «чистые», а скоррелированы с регрессорами, и роскошь применить ОМНК недоступна, ибо ОМНК даёт несостоятельные результаты. Так что в принципе мы не столько выбираем между типами эффектов, сколько между методами оценивания.

Итак, в случае «чистых» случайных эффектов мы можем использовать ОМНК-оценку, которая в данной ситуации является состоятельной и асимптотически эффективной. Для того, чтобы она была таковой, ошибка μ_i не должны коррелировать с регрессорами x_{it}. В случае же фиксированных эффектов преобладает внутри-оценка, и она состоятельна безотносительно факта скоррелированности или нескоррелированности ошибок с регрессорами.

Нескоррелированность ошибок с регрессорами означает, что панельная регрессия правильно специфицирована, то есть регрессионная функция $x_{it}'\beta$ действительно линейна и содержит все релевантные регрессоры, и остаточная индивидуальная неоднородность, заложенная в индивидуальных эффектах, действительно случайна, не связана с характеристиками объекта, находящимися в x_{it}.

Таким образом, выбор между случайными и фиксированными эффектами при «статистическом» подходе сводится к проверке того, есть ли корреляция между ошибками и обычными регрессорами. Поскольку речь фактически идёт о тестировании спецификации модели, мы можем воспользоваться инструментарием Хаусмана. Наиболее жёсткой гипотезой является нескоррелированность ошибок с регрессорами, поэтому в качестве нулевой гипотезы берём

$$H_0: \ E[u_i|X_i] = 0,$$

в коем случае случайные эффекты – наиболее адекватная спецификация, а ОМНК – наиболее адекватный метод оценивания. В качестве альтернативной гипотезы берём

$$H_A: \ E[u_i|X_i] \neq 0,$$

то есть фиксированные эффекты – наиболее адекватная спецификация, и используется внутри-оценивание.

Как уже упоминалось, оценка, являющаяся состоятельной и асимптотически эффективной при H_0, но несостоятельной при H_A – это ОМНК-оценка. Возьмём её в качестве $\widehat{\beta}_0$ для теста Хаусмана. Альтернативная, робастная, состоятельная как при H_0, так и при H_A, но неэффективная ни при одной из них – это внутри-оценка, она будет взята в качестве $\widehat{\beta}_1$. Тогда статистика Хаусмана равна

$$\mathcal{H} = (\widehat{\beta}_W - \widehat{\beta}_{GLS})'((X'QX)^{-1}X'Q\Omega QX(X'QX)^{-1}) - (X'\Omega^{-1}X)^{-1})^{-1}(\widehat{\beta}_W - \widehat{\beta}_{GLS}) \xrightarrow{d} \chi^2_k,$$

при условии обратимости обращаемой матрицы, которое, как правило, выполняется. Здесь k – это размерность β. Если $\mathcal{H} > q^{\chi^2_k}_{1-\alpha}$, то мы не можем принять спецификацию со случайными эффектами и вынуждены работать с фиксированными; если же $\mathcal{H} \leqslant q^{\chi^2_k}_{1-\alpha}$, спецификацию со случайными эффектами можно принять.

В случае непринятия случайных эффектов иногда удаётся улучшить модель, найдя дополнительные объясняющие переменные для «нечистых» случайных эффектов и добавив их в модель в качестве регрессоров.

7 Динамическая панельная регрессия

Здесь мы отклоняемся от стандартной модели панельных данных в одном из самых важных направлений. Теперь среди объясняющих переменных появятся лагированные y-ки. Рассмотрим простейшую ситуацию, когда у нас временных эффектов нет, а вся динамика управляется дополнительным регрессором – лагированном y.

Итак, рассмотрим однонаправленную структуру с динамическим слагаемым:

$$y_{i,t} = \delta y_{i,t-1} + x'_{i,t}\beta + \mu_i + v_{i,t},$$

где обе компоненты ошибки μ_i и v_i независимы от $x_{i,t}$, и всего уравнений $n(T-1)$, поскольку $t = 2, \ldots, T$ из-за лагированного y. МНК и ОМНК здесь несостоятельны, так как μ_i коррелирует с $y_{i,t-1}$.

Рассмотрим между-преобразованную систему:

$$\overline{y}_i = \delta\overline{y}_{i,-1} + \overline{x}'_i\beta + \mu_i + \overline{v}_i,$$

где слева – среднее $y_{i,t}$ от $t = 2$ до T, а справа – тоже среднее $y_{i,t}$, но от $t = 1$ до $T-1$. Между-оценка также несостоятельна, по тем же причинам, что и МНК с ОМНК.

Рассмотрим внутри-преобразование, удаляющее индивидуальные эффекты:

$$y_{i,t} - \overline{y}_i = \delta(y_{i,t-1} - \overline{y}_{i,-1}) + (x_{i,t} - \overline{x}_i)'\beta + v_{i,t} - \overline{v}_i.$$

Увы, внутри-ошибка содержит в себе почти все идиосинкратические компоненты, которые коррелируют с $y_{i,t-1}$. Поэтому эта оценка тоже несостоятельна (если панель короткая).

Удалим индивидуальные эффекты другим способом – взятием первых разностей:

$$y_{i,t} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (x_{i,t} - x_{i,t-1})'\beta + v_{i,t} - v_{i,t-1},$$

где $i = 1, \ldots, n$, $t = 3, \ldots, T$, всего уравнений $n(T-2)$. МНК и ОМНК здесь неприменимы из-за корреляции между $v_{i,t-1}$ и $y_{i,t-1}$. Зато в данной ситуации можно применять метод инструментальных переменных, причём инструменты легко находятся.

В матричной форме имеем:

$$\underset{n(T-2)\times 1}{\Delta y} = \delta \underset{n(T-2)\times 1}{\Delta y_{-1}} + \underset{n(T-2)\times k}{\Delta X}\underset{k\times 1}{\beta} + \underset{n(T-2)\times 1}{\Delta v}.$$

Создадим матрицу инструментов:

$$\underset{n(T-2)\times ?}{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix}.$$

В качестве инструментов можно выбирать лагированные разности Δy или сами лагированные y-ки. Опытным путем было получено, что сами лагированные y-ки брать лучше, чем приращения.

Мы хотим использовать так много инструментов, как это возможно, ради большей эффективности. Возьмём первое уравнение (т.е. соответствующее $t = 3$) для i-го объекта и найдём для него инструменты. При $t = 3$ ошибка состоит из $v_{i,3}$ и $v_{i,2}$. Мы хотим среди y-ков найти те, которые нескоррелированы с ними, который реализовались раньше обоих ошибок. Таковой единственный – $y_{i,1}$, все же остальные y-ки будут коррелировать с ошибкой либо непосредственно, либо через предысторию. Чтобы выбрать инструменты для $x_{i,t}$, можно сделать дополнительное предположение, что $x_{i,t}$ строго экзогенны, что означает, что они определяются вне системы и не коррелируют с этими ошибками, ни с будущими, ни с прошлыми. Тогда в качестве инструментов можно взять все имеющиеся $x_{i,t}$. Выпишем все найденные инструменты в одну строчку, остальное заполняем нулями.

Перейдём к следующей строчке, соответствующей $t = 4$. Теперь ошибка равна $v_{i,4} - v_{i,3}$, и есть два y-ка, которые реализовались до момента времени $t = 3$. Поэтому здесь мы берём инструменты $y_{i,1}$ и $y_{i,2}$. Закономерность понятна. Когда мы дойдём до самой нижней строчки, соответствующей $t = T$, ошибка будет равна $v_{i,T} - v_{i,T-1}$. Это значит, что у нас много y-ков, которые реализовались до момента $t = T - 1$, а именно от $y_{i,1}$ до $y_{i,T-2}$, а потом все те же x-ы, что и ранее.

Выпишем всю матрицу инструментов для i-го объекта:

$$W_i = \begin{pmatrix} y_{i,1}\, x'_{i,1}\, x'_{i,2}\ \dots\ x'_{i,T} & 0 \dots 0 & 0 \dots 0 \\ 0 \dots 0 & y_{i,1}\, y_{i,2}\, x'_{i,1}\, x'_{i,2}\ \dots\ x'_{i,T} & 0 \dots 0 \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ 0 \dots 0 & 0 \dots 0 & y_{i,1}\ \dots\ y'_{i,T-2}\, x'_{i,1}\ \dots\ x'_{i,T} \end{pmatrix}$$

Получилась широкая матрица высоты $n(T - 2)$. Если строгая экзогенность x-ов – слишком сильное и неправдоподобное предположение, можно предположить их *предопределенность*, когда текущие и прошлые значения x-ов не коррелируют с ошибками, тогда как будущие вполне могут коррелировать. Тогда от части инструментов придётся избавиться. В первой строчке из x-ов останутся только $x'_{i,1}$ и $x'_{i,2}$, во второй – $x'_{i,1}$, $x'_{i,2}$ и $x'_{i,3}$, и так далее. В последней строчке будут стоять почти все x-ы, кроме последнего, то есть $x'_{i,1}\ \dots\ x'_{i,T-1}$.

Далее, мы хотим скорректировать на серийную корреляцию в ошибке $v_{i,t} - v_{i,t-1}$

во благо более эффективного оценивания. Для этого построим следующую матрицу:

$$\underset{(T-2)\times(T-2)}{G} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ 0 & 0 & -1 & & \vdots \\ & \dots & & \ddots & -1 \\ & \dots & & -1 & 2 \end{pmatrix}.$$

По главной диагонали этой матрицы стоят дисперсии ошибок, а вне главной диагонали – ковариации, с точностью до общей константы.

Теперь можно выписать эффективную ОММ-оценку:

$$\begin{pmatrix} \widehat{\delta} \\ \widehat{\beta} \end{pmatrix} = ((\Delta y_{-1}\Delta X)'W(W'(I_n \otimes G)W)^{-1}W'(\Delta y_{-1}\Delta X))^{-1}$$

$$\times(\Delta y_{-1}\Delta X)'W(W'(I_n \otimes G)W)^{-1}W'\Delta y.$$

Для подсчёта стандартных ошибок необходимо оценить асимптотическую дисперсию:

$$A\widehat{Var}\begin{pmatrix} \widehat{\delta} \\ \widehat{\beta} \end{pmatrix} = \widehat{\sigma}_v^2((\Delta y_{-1}\Delta X)'W(W'(I_n \otimes G)W)^{-1}W'(\Delta y_{-1}\Delta X))^{-1}.$$

Если бы мы не предположили независимость между ошибками и обычными регрессорами, допустив условную гетероскедастичность, процедуру, конечно же, мы сделали бы двухшаговой. На втором шаге с помощью остатков, полученных в результате описанного выше алгоритма, мы переоценили бы взвешивание, скорректировав на условную гетероскедастичность, и запустили бы алгоритм вновь.