

1 Эконометрика. Методы оценки параметров моделей

Эконометрика описывается как дисциплина, целью которой является наполнение эмпирическим содержимым экономических отношений. Главным средством достижения этой цели является применение математических и статистических методов к исследованию экономических данных.

Термин «данные» очень широко используется в самых разных областях человеческой деятельности. В широком смысле он означает фактический материал, который является основой для анализа и принятия решений. С точки зрения статистики, данные — это информация, пригодная для анализа и интерпретации. Часто в качестве синонима к слову «данные» используют «наблюдения», т.е. реализации некоторой случайной величины.

Данные получают путем измерения. Формально, измерение — это приписывание изучаемым объектам набора символов по определенным правилам. Такое широкое определение необходимо, поскольку случайные величины, интересующие исследователя, могут иметь различную природу. Например, для каждого жителя страны можно измерить такие показатели, как: пол, социальное положение, температура, рост. Сравнение двух людей по первому показателю может лишь показать, что их пол одинаков или различен. Для второго показателя можно указать, что эти люди отличны и, более того, кто из них занимает более высокое положение в обществе. Третий показатель, помимо отличий и порядка между людьми, показывает, на сколько они отличаются. Наконец, четвертый показатель позволяет указать, что один человек выше другого во сколько-то раз, т.е. наличествует некая абсолютная точка отсчета. Эти четыре случайные величины представляют примеры четырех различных шкал.

Номинальная шкала используется для классификации объектов, она лишь указывает, какому классу он принадлежит. Разумеется, арифметические действия на такой шкале невозможны.

Порядковая шкала используется для классификации и упорядочивания объектов.

Интервальная шкала позволяет количественно оценить разницу между объектами. Для этого вводится единица измерения и начало отсчета. Структура этой шкалы инвариантна к линейному преобразованию случайной величины.

Шкала сравнений, как и интервальная, количественно оценивает различие, однако начало отсчета в ней абсолютное, обусловленное природой случайной величины (в приведенном примере рост не может быть меньше нуля). Структура этой шкалы не меняется при «растягивании» или «сжатии» шкалы, т.е. масштабировании единицы измерения.

1.1 Структура экономических данных

1.1.1 Перекрестные данные

1.1.2 Временные ряды

1.1.3 Объединенные перекрестные данные

1.1.4 Панельные (продольные) данные

1.2 Регрессионный анализ. Метод наименьших квадратов для оценки линейной регрессии.

Регрессионный анализ — совокупность методов исследования связи одной или нескольких переменных, называемых зависимыми (регрессандами), с несколькими другими,

называемыми независимыми (регрессорами). Связь эта представляется в виде математической модели, т.е. уравнения, которое связывает зависимую переменную с независимыми с учетом множества предположений. Обычно предполагается что эта зависимость (функция регрессии) известна с точностью до набора параметров. Модель, функция регрессии которой линейна относительно параметров модели, также называется линейной регрессионной моделью. Типичными задачами, которые решаются регрессионным анализом, являются:

- получение наилучших точечных и интервальных оценок неизвестных параметров регрессии;
- проверка гипотез относительно этих параметров;
- проверка адекватности предполагаемой модели;
- проверка соблюдения предположений, принятых при составлении модели;
- получение наилучших точечных и интервальных оценок значений зависимой переменной (прогноз).

Регрессионная модель позволяет, во-первых, описать зависимость между интересующими исследователя переменными, вскрыть или прояснить структуру причинных связей. Во-вторых, она может служить основой для прогноза поведения зависимой переменной, что очень важно, когда прямые измерения либо невозможны, либо экономически невыгодны.

Introductory conometrics. Matrix form of linear regression

1.3 Нелинейная регрессия и метод наименьших квадратов

Пусть теперь набор данных $y_1, y_2, \dots, y_n \in \mathbb{R}$, измеренный при известных значениях k факторов (регрессоров) x_1, x_2, \dots, x_k мы пытаемся аппроксимировать нелинейной функцией регрессии $\varphi(x_1, \dots, x_k; \alpha)$ так, чтобы

$$y_t = \varphi_t(x_{t1}, x_{t2}, \dots, x_{tk}; \alpha) + \varepsilon_t.$$

Полагая значения x_{ti} заданными и фиксированными, применим метод наименьших квадратов для нахождения набора параметров $\alpha \in \mathbb{R}^m$, дающего наилучшее приближение. Введя обозначение $\varphi_t(x_{t1}, x_{t2}, \dots, x_{tk}; \alpha) = f_t(\alpha)$, $t = \overline{1, n}$ и составив сумму квадратов невязок, получаем задачу математического программирования вида

$$Q(\alpha) = \sum_{t=1}^n (y_t - f_t(\alpha))^2 \longrightarrow \min_{\alpha}$$

Заметим, что случайные погрешности ε_t в уравнении нелинейной регрессии предполагаются имеющими нулевое математическое ожидание и постоянную дисперсию, и не коррелирующими между собой.

Вообще говоря, для нелинейной регрессии минимум суммы квадратов невязок может и не достигаться. Кроме того, задача математического программирования может быть многоэкстремальной. Однако за многие годы использования и развития МНК эти

проблемы неплохо изучены и существуют теоретические критерии, позволяющие устанавливать корректность поставленной задачи регрессии. К сожалению, проверка выполнения этих критериев часто представляет собой очень сложную задачу и до конца может быть проведена только в редких случаях.

1.3.1 Минимизация суммы квадратов методом Ньютона-Гаусса

Пусть α_0 — некоторое достаточно хорошее приближение к оптимальному значению α . Метод Ньютона минимизации функций многих переменных строит последовательность точек по правилу

$$\alpha_{s+1} = \alpha_s - \left(\frac{\partial^2 Q(\alpha^s)}{\partial \alpha^2} \right)^{-1} \frac{\partial Q(\alpha_s)}{\partial \alpha}, \quad s = 0, 1, \dots$$

Особая форма минимизируемой функции позволяет модифицировать эту формулу. Во-первых, гессиан $Q(\alpha)$ имеет вид

$$\frac{1}{2} \frac{\partial^2 Q}{\partial \alpha^2} = \left(\frac{\partial f_i}{\partial \alpha_j} \right)^T \left(\frac{\partial f_i}{\partial \alpha_j} \right) - \sum_{t=1}^n (y_t - f_t(\alpha)) \frac{\partial^2 f_t(\alpha)}{\partial \alpha^2}.$$

Во-вторых, поскольку α_0 близка к точке минимума, невязки $y_t - f_t(\alpha)$ в некоторой ее окрестности будут невелики. В-третьих, если предположить, что функция регрессии «не очень нелинейна», то $\frac{\partial^2 f_t(\alpha)}{\partial \alpha^2}$ тоже будут малы.

Учитывая это, суммой в гессиане $Q(\alpha)$ можно пренебречь, т.е.

$$\frac{\partial^2 Q}{\partial \alpha^2} \approx 2 \left(\frac{\partial f_i}{\partial \alpha_j} \right)^T \left(\frac{\partial f_i}{\partial \alpha_j} \right)$$

Введя матрицу первых производных $F(\alpha) = \left(\frac{\partial f_i(\alpha)}{\partial \alpha_j} \right)$, можем записать

$$\frac{\partial^2 Q}{\partial \alpha^2} \approx 2F^T(\alpha)F(\alpha), \quad \frac{\partial Q}{\partial \alpha} = -2F^T(\alpha)(y - f(\alpha)).$$

Формулу итераций метода Ньютона можно теперь записать в виде

$$\alpha_{s+1} = \alpha_s + \left(F^T(\alpha_s)F(\alpha_s) \right)^{-1} F^T(\alpha_s)(y - f(\alpha_s)), \quad s = 0, 1, \dots \quad (*)$$

Очевидно, в последней формуле, как и в исходной, предполагается, что $\text{rank}(F(\alpha)) = m$.

Задача. Как еще можно получить формулу (*)?

Формула (*) является основой метода Ньютона-Гаусса нахождения параметров нелинейной регрессии. Основной недостаток метода — сходимость гарантируется при довольно жестких условиях. Если эти условия нарушаются (например, матрица $\frac{\partial^2 f_t(\alpha)}{\partial \alpha^2}$ сильно отлична от нулевой и/или невязки $|y_t - f_t(\alpha)| \gg 0$), последовательность приближений α_s расходится. Простейший выход из ситуации заключается в регуливовке длины шага:

$$\alpha_{s+1} = \alpha_s + \lambda_s p_s,$$

где $p_s = \left(F^T(\alpha_s)F(\alpha_s) \right)^{-1} F^T(\alpha_s)(y - f(\alpha_s))$ — направление минимизации, (такое же, как и в исходном методе), а $\lambda_s > 0$ — длина шага в этом направлении. Очевидный

метод регулировки шага — метод редукции — позволяет строить строго релаксационные последовательности, т.е. такие, что

$$Q(\alpha_{s+1}) < Q(\alpha_s) .$$

Метод редукции состоит в следующем: выберем два параметра метода, $0 < \tau < 1$ и $0 < \varkappa < 1$. Далее, на каждой итерации

1. Полагаем $\lambda = 1$.
2. Проверяем неравенство

$$Q(\alpha_s + \lambda p_s) - Q(\alpha_s) < -\tau q_s^T p_s, \quad (**)$$

где $q_s = -\frac{\partial Q(\alpha_s)}{\partial \alpha}$, $q_s \neq 0$.

3. Если неравенство выполняется, полагаем $\lambda_s = \lambda$. В противном случае уменьшаем шаг в \varkappa раз, т.е. $\lambda := \varkappa \lambda$ и возвращаемся к шагу 2.

Можно показать, что, если $q_s \neq 0$, то для любого $0 < \tau < 1$ найдется такое $\bar{\lambda}$, что для всех $0 < \lambda < \bar{\lambda}$ неравенство $(**)$ выполняется.

1.3.2 Минимизация суммы квадратов методом Левенберга-Марквардта

Помимо условий сходимости, еще одним ограничением на применение метода Ньютона является условие полного ранга матрицы производных функции регрессии: $\text{rank}(F(\alpha)) = m$. Рассмотрим, например, модель нелинейной регрессии вида

$$y_t = \alpha_1 e^{\alpha_2 x_t} + \alpha_3 e^{\alpha_4 x_t} + \varepsilon_t .$$

Легко проверить, что на подпространстве (линейном) $\alpha_2 = \alpha_4$ не выполняется условие $|F^T(\alpha)F(\alpha)| > 0$.

Задача. Проверьте последнее утверждение.

Даже если условие формально выполняется, определитель матрицы может быть настолько мал, что из-за погрешности вычислений обратная матрица $(F^T(\alpha)F(\alpha))^{-1}$ не будет положительно определенной. Это может привести, в частности, к заикливанию итераций. На практике маловероятно, что $\text{rank}(F(\alpha)) < m$, однако плохая обусловленность матрицы может представлять реальную проблему. Один из методов, который ее обходит — метод Левенберга-Марквардта.

В 1944 г. Кеннет Левенберг предложил использовать простой прием, который обеспечивал бы положительную определенность матрицы в итерациях метода Ньютона-Гаусса. Прием этот заключался в добавлении матрицы, пропорциональной единичной следующим образом:

$$\alpha_{s+1} = \alpha_s + (F^T(\alpha_s)F(\alpha_s) + \mu_s I)^{-1} F^T(\alpha_s)(y - f(\alpha_s)), \quad s = 0, 1, \dots$$

Прием имел простое эмпирическое обоснование: матрица $F^T(\alpha_s)F(\alpha_s)$ неотрицательно определена, поэтому $F^T(\alpha_s)F(\alpha_s) + \mu_s I$ определена положительно ($\mu_s > 0$). Следовательно, направление минимизации будет составлять острый угол с направлением антиградиента минимизируемой функции. Эта идея была развита двадцать лет спустя, когда в 1963 г. Дональд Марквардт предложил вместо единичной матрицы использовать

диагональную $D = \text{diag}(F_s^T F_s)$, что равносильно масштабированию направления минимизации. Для улучшения сходимости метода можно использовать переменный шаг, подбирая его методом редукции. Расчетная формула метода Левенберга-Марквардта имеет вид

$$\alpha_{s+1} = \alpha_s + \lambda_s (F^T(\alpha_s)F(\alpha_s) + \mu_s D_s)^{-1} F^T(\alpha_s)(y - f(\alpha_s)), \quad s = 0, 1, \dots$$

Самым сложным вопросом при использовании этого метода является выбор μ_s . Д. Марквардтом было показано, что при больших значениях μ_s алгоритм близок к градиентному, при малых — к методу Ньютона-Гаусса. Поскольку последний хорошо сходится только при наличии достаточно близкого к точке минимума начального приближения, а градиентный метод, наоборот, плохо ведет себя в «овраге», разумной представляется следующая общая рекомендация: значения μ_s следует постепенно уменьшать, устремляя к нулю.

1.4 Метод максимального правдоподобия. Оценка линейной модели

С точки зрения статистики, набор измерений регрессанда (y_1, y_2, \dots, y_n) является случайной выборкой из генеральной совокупности с неизвестными свойствами. И задача восстановления зависимости от регрессоров эквивалентна восстановлению этих свойств: наилучшей моделью зависимости будет та, которой больше всего соответствуют имеющиеся данные.

Свойства случайной величины, как известно, описываются функцией распределения вероятностей и/или функцией плотности вероятностей. Как правило, справедливо предположение, что конкретная зависимость описывается параметрическим семейством плотностей вероятностей, т.е. каждому значению некоторого набора параметров соответствует определенная функция плотности вероятности. Таким образом, восстановление зависимости сводится к оценке параметров плотности вероятности, наилучшим образом соответствующей измерениям.

Предположим, что (y_1, y_2, \dots, y_n) — независимые и одинаково распределенные реализации некоторой случайной величины. Обозначим истинную плотность вероятности этой величины через $f_0(y)$. Пусть эта функция принадлежит параметризованному семейству $\{f(\cdot|\theta) : \theta \in \Omega\}$, так, что $f_0(y) = f(y|\theta_0)$. Значение θ_0 (вообще говоря — векторного) будем называть истинным значением параметра распределения. Задача состоит в том, чтобы найти оценку $\hat{\theta}$, наиболее близкую к θ_0 .

Поскольку измерения y_t предполагаются независимыми, вероятность реализации всех n значений равна произведению вероятностей реализации каждого из них. Поэтому

$$f(y_1, y_2, \dots, y_n|\theta) = \prod_{t=1}^n f(y_t|\theta).$$

Теперь, поскольку набор (y_1, y_2, \dots, y_n) действительно реализовался, можно смотреть на последнее выражение как на функцию θ при фиксированных y_t . Эта функция называется функцией правдоподобия:

$$L(\theta|y_1, y_2, \dots, y_n) = f(y_1, y_2, \dots, y_n|\theta) = \prod_{t=1}^n f(y_t|\theta).$$

На практике часто используется так называемая логарифмическая функция правдоподобия:

$$\ln L(\theta|y_1, y_2, \dots, y_n) = \sum_{t=1}^n \ln f(y_t|\theta)$$

или ее масштабированный вариант:

$$l(\theta|y_1, y_2, \dots, y_n) = \frac{1}{n} \sum_{t=1}^n \ln f(y_t|\theta) .$$

Оценкой метода максимального правдоподобия называется решение задачи

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Omega} L(\theta|y_1, y_2, \dots, y_n),$$

где вместо функции правдоподобия L могут быть использованы любые функции, получаемые из нее монотонным преобразованием, так как абсолютные значения функции правдоподобия не важны.

<Пример из "Эконометрики" Дугерти>

Покажем, как метод максимального правдоподобия позволяет найти оценки параметров линейной регрессии. Согласно идее метода, предположим, что каждое наблюдение y_t есть реализация случайной величины, подчиняющейся распределению $y_t \in N(\beta_0 + \beta_1 x_t, \sigma^2), t = \overline{1, n}$. Параметрами распределения являются величины $\beta_0, \beta_1, \sigma^2$. Нетрудно видеть, что

$$\begin{aligned} f(y_1, y_2, \dots, y_n | \beta_0, \beta_1, \sigma^2) &= \prod_{t=1}^n f(y_t | \beta_0, \beta_1, \sigma^2) = \\ &= \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_t - \beta_0 - \beta_1 x_t)^2}{2\sigma^2}} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 x_t)^2} \end{aligned}$$

Значит, логарифмическая функция правдоподобия имеет вид

$$\ln L(\beta_0, \beta_1, \sigma^2 | y_1, y_2, \dots, y_n) = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 x_t)^2$$

Решение задачи максимизации функции правдоподобия приводит к системе уравнений:

$$\begin{cases} \frac{\partial \ln L}{\partial \beta_0} = \frac{1}{\sigma^2} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 x_t) = 0 \\ \frac{\partial \ln L}{\partial \beta_1} = \frac{1}{\sigma^2} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 x_t) x_t = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 x_t)^2 = 0 \end{cases}$$

Ее решение легко можно найти:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_t (x_t - \bar{x})(y_t - \bar{y})}{\sum_t (x_t - \bar{x})^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_t e_t^2 = \frac{1}{n} \sum_t (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 \end{aligned}$$

Как видно, оценки параметров линейной модели методом максимального правдоподобия совпадают с оценками метода МНК.

Задача. Показать, что оценки параметров β_0 и β_1 являются в данном случае несмещенными. Найти их дисперсию.

Оценки метода максимального правдоподобия обладают такими свойствами, как состоятельность и асимптотическая эффективность, причем при довольно мягких ограничениях на исходные данные и модель.

Задача. Было проведено маркетинговое исследование эффективности рекламной кампании некоторого товара. Через равные промежутки времени проводилось тестирование фокус-группы исследуемых, которое выявляло, какой объем информации о продукте оставался в актуальной памяти. Для этого испытуемым задавалось m вопросов и измерялась величина $\frac{x_t}{m}$, где x_t — число верных ответов при тестировании в момент времени t . Затем она усреднялась по всем испытуемым. Как можно видеть, величина x_t подчиняется биномиальному распределению, вопрос только в том, по какому закону меняется вероятность успешного исхода в каждом из независимых испытаний.

Очевидно, что вероятность верного ответа со временем уменьшается, т.к. из памяти стираются ненужные детали. Были выдвинуты две гипотезы:

$$\begin{array}{ll} \text{Степенная модель } H_0 & : \quad p(w, t) = w_1 t^{-w_2}, \quad w_1, w_2 > 0 \\ \text{Экспоненциальная модель } H_1 & : \quad p(w, t) = w_1 e^{-w_2 t}, \quad w_1, w_2 > 0 \end{array}$$

Необходимо применить метод максимального правдоподобия, чтобы найти оценки параметров w_1, w_2 в каждом случае и выяснить, какая из гипотез лучше объясняет имеющийся набор данных $y = (0.94, 0.77, 0.40, 0.26, 0.24, 0.16)^T$, причем измерения проводились в моменты времени $t = 1, 3, 6, 9, 12, 18$. Объем испытаний при каждом измерении равен $m = 100$. Сравнить графически полученные модели с реальными данными.