ЭКОНОМЕТРИКА

Лекция 8.

§ 8. Нелинейная регрессия.

Многие зависимости, описывающие связи между параметрами экономических процессов не являются линейными, поэтому их моделирование с помощью линейных моделей может привести к неточностям или даже ошибкам.

Например, если изучается зависимость спроса от цены товара, то можно ограничиться линейным уравнением регрессии. Если же анализируется эластичность спроса по цене, то целесообразно рассмотреть *логарифмическую модель*.

При анализе издержек от объема выпуска наиболее обоснованной является *полиномиальная (кубическая) модель*.

При рассмотрении производственных функций используются *степенные модели* (например, *производственная функция Кобба-Дугласа* $Y = AK^{\alpha} L^{\beta}$, где Y - объем выпуска, K, L - затраты капитала и труда, соответственно).

Применяются также другие нелинейные модели, в частности, *обратная и экспоненциальная модели*.

Наиболее простые нелинейные модели из числа указанных выше – нелинейные модели, сводящиеся к линейным с помощью некоторых преобразований. В данном курсе будут рассматриваться только такие модели.

1 . Логарифмические (лог-линейные) модели

Пусть некоторая зависимость между экономическими показателями моделируется формулой

$$Y = A \cdot X^{\alpha}, \tag{8.1}$$

где A и α - параметры модели, которые следует определить., а ε.

Например, подобное соотношение может моделировать зависимость спроса Y на некоторый товар от его цены X (в таком случае $\alpha < 0$) или от дохода X (в таком случае $\alpha > 0$, а зависимость (8.1) носит название *функции Энгеля*). Функция (8.1) может отражать также зависимость объема выпуска Y от объемов используемого ресурса X. Тогда (8.1) – *производственная функция*, причем $0 < \alpha < 1$.

Модель (8.1) не является линейной, поскольку зависимость Y от X не является в данном случае пропорциональной. Более того, предположение о существовании «чистой» зависимости (8.1) чаще всего не является точным, поэтому уравнение (8.1) корректируется положительным множителем (мультипликативным случайным отклонением $\mathrm{E} > 0$)

$$Y = A \cdot X^{\alpha} \cdot \mathrm{E}. \tag{8.2}$$

Стандартный подход к исследованию такой модели – ее преобразование к линейной модели с помощью логарифмирования:

$$\ln Y = \alpha_0 + \alpha \cdot \ln X + \varepsilon, \tag{8.3}$$

$\alpha_0 = \ln A, \quad \varepsilon = \ln \mathrm{E}$. Делая в (8.3) замены исследуемых переменных $Y^* = \ln Y, X^* = \ln X$, приходим к линейной модели

$$Y^* = \alpha_0 + \alpha \cdot X^* + \varepsilon, \tag{8.4}$$

которая была рассмотрена ранее. При выполнении классических предпосылок метода наименьших квадратов, можно определить оптимальные несмещенные оценки для параметров α_0, α.

Дифференцируя соотношение (8.1) (или (8.2), или (8.3)), получим, что коэффициент α представляет собой эластичность переменной Y по переменной X :

$$\frac{1}{Y}\frac{dY}{dX} = \alpha \cdot \frac{1}{X} \Rightarrow \alpha = \frac{dY}{dX} \cdot \frac{X}{Y} = E_X(Y). \qquad (8.5)$$

Поскольку коэффициент α является постоянным, то (двойная) логарифмическая модель (8.2) характеризует процессы, обладающие свойством *постоянной эластичности зависимой переменной по фактору* (такая модель называется также моделью постоянной эластичности). В случае парной регрессии обоснованность применения такой модели вытекает из следующих соображений. Если нанести на корреляционное поле «преобразованные точки наблюдения» $(\ln x_t, \ln y_t), \quad t = 1, \ldots, T$, и при этом окажется, что данные точки располагаются близко к некоторой прямой, то применение логарифмической модели обосновано.

Данная модель легко обобщается на случай большего числа факторов

$$\ln Y = \alpha_0 + \alpha_1 \cdot \ln X_1 + \ldots + \alpha_m \cdot \ln X_m + \varepsilon, \qquad (8.6)$$

где коэффициент α_j представляют собой эластичность зависимой переменной Y по независимой (факторной) переменной X_j. В частности, при $m = 2$ такой вид имеет производственная функция Кобба-Дугласа

$$Y = A \cdot K^{\alpha} \cdot L^{\beta},$$

где α, β - эластичности выпуска по затратам капитала и труда, соответственно. Сумма этих коэффициентов называется (коэффициентом) *отдачи от масштаба*. При $\alpha + \beta = 1$ говорят о *постоянной отдаче от масштаба* (во сколько раз увеличиваются совокупные затраты ресурсов, во столько раз увеличивается выпуск). При $\alpha + \beta < 1$ имеет место *убывающая отдача от масштаба* (прирост выпуска меньше прироста совокупных затрат на ресурсы), а при $\alpha + \beta > 1$ - *возрастающая отдача от масштаба*.

2 . Полулогарифмические модели

Модели вида

$$\ln Y = \alpha_0 + \alpha \cdot X + \varepsilon, \qquad (8.7)$$

$$Y = \alpha_0 + \alpha \cdot \ln X + \varepsilon, \qquad (8.8)$$

носят название полу-логарифмических. Такие модели обычно употребляют, когда необходимо оценивать темп роста или темп прироста каких-либо экономических показателей. Например, при анализе банковского вклада по первоначальному вкладу и процентной ставке, при исследовании прироста объема выпуска от относительного увеличения затрат ресурсов, при изучении бюджетного дефицита от темпа роста ВНП, при оценке темпа роста инфляции от объема денежной массы и т.п.

Рассмотрим, в частности, известную в банковском и финансовом анализе модель начисления сложных процентов

$$Y_t = Y_0 (1 + r)^t, \qquad (8.9)$$

где Y_0 - начальная величина переменной Y (например, первоначальный вклад в банке), r - темп прироста величины Y (например, процентная ставка банка), Y_t - значение величины Y в момент времени t. Модель (8.9) является по сути полу-логарифмической, поскольку

$$\ln Y_t = \alpha_0 + \alpha \cdot t + \varepsilon_t, \qquad (8.10)$$

где $\alpha_0 = \ln Y_0, \alpha = \ln (1 + r)$. Дополнительное случайное слагаемое вводится в уравнение (8.9) с целью характеризации возможной изменчивости процентной ставки.

Коэффициент α в (8.7) имеет смысл темпа прироста переменной Y по переменной X, т.е. характеризует отношение относительного изменения Y к абсолютному изменению показателя X:

$$\frac{1}{Y}\cdot\frac{dY}{dX}=\alpha\Rightarrow\alpha=\frac{dY/Y}{dX}. \tag{8.11}$$

Данный коэффициент может быть выражен в процентах.

В случае модели (8.10) из соотношения $\alpha=\ln(1+r)$ определяется темп прироста r величины Y:

$$r=e^{\alpha}-1. \tag{8.12}$$

Коэффициент α определяет мгновенный темп прироста, а r в соотношении (8.9) определяет обобщенный (сложный) темп прироста. Поэтому в общем случае они отличаются друг от друга.

Рассмотрим *линейно-логарифмическую модель* (8.8). Она сводится к линейной с помощью преобразования $X^*=\ln X$. В данной модели коэффициент α характеризует отношение абсолютного изменения зависимой переменной Y к относительному изменению переменной X:

$$\frac{dY}{dX}=\frac{\alpha}{X}\Rightarrow\alpha=\frac{dY}{dX/X}\Rightarrow\Delta Y\approx\alpha\frac{dX}{X}.$$

Таким образом, изменению $\dfrac{\Delta X}{X}$ на 0,01 соответствует изменение Y на $0{,}01\cdot\alpha$.

Линейно-логарифмическая модель используется в тех случаях, когда необходимо исследовать влияние процентного изменения независимой (факторной) переменной на абсолютное изменение зависимой переменной. Например, если $Y=GNP$ - валовой национальный продукт, а $X=M$ - объем денежной массы, то связь между ними можно моделировать следующим соотношением

$$GNP=\alpha_0+\alpha\ln M+\varepsilon,$$

из которого вытекает, что если увеличить объем денежной массы M на 1%, то валовой национальный продукт в среднем вырастет на $0{,}01\cdot\alpha$.

3 . Обратная модель

Модель вида

$$Y=\alpha_0+\alpha_1\cdot\frac{1}{X}+\varepsilon \tag{8.13}$$

носит название *обратной модели*. Она сводится к линейной с помощью замены $X^*=\dfrac{1}{X}$.

Данная модель применяется в тех случаях, когда соотношение между переменными имеет характер обратно-пропорциональной связи. Модель также можно охарактеризовать следующим образом: неограниченному увеличению переменной X соответствует асимптотическое приближение зависимой переменной к определенному пределу. Последнее свойство можно проиллюстрировать графически. При этом, например, в случае $\alpha_0>0,\alpha_1>0$ уравнение (8.13) может использоваться для моделирования зависимости средних фиксированных издержек Y от объема выпуска продукции X. В случае $\alpha_0>0,\alpha_1<0$ с помощью (8.13) можно моделировать, например, зависимость спроса на блага Y от уровня доходов X. Такая функциональная зависимость носит название функции Торнквиста, при этом значение $X = -\frac{\alpha_1}{\alpha_0}$ определяет минимально необходимый уровень дохода. При $\alpha_0 < 0, \alpha_1 > 0$ одно из графических описаний уравнения (8.13) носит название кривой Филлипса, которая описывает как уровень безработицы X (в процентах) определяет процентное изменение уровня заработной платы Y. При этом кривая Филлипса пересекает ось OX в некоторой точке x_0, которая определяет естественный уровень безработицы.

4 . Степенная модель

Модель вида

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \ldots + \alpha_m X^m + \varepsilon \tag{8.14}$$

носит название степенной модели. Например, кубическая функция

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X^3 + \varepsilon \tag{8.15}$$

в микроэкономике моделирует зависимость общих издержек TC (total charge) от объема выпуска Q.

Аналогично, квадратичная функция

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \varepsilon \tag{8.16}$$

Может отражать зависимость средних издержек AC (average charge), либо предельных (маржинальных) издержек MC (marginal charge) объемом выпуска Q; или зависимость прибыли π от расходов на рекламу C (advertising cost) и т.д.

Модель (8.14) является линейной относительно коэффициентов регрессии $\alpha_0, \alpha_1, \ldots, \alpha_m$. Она сводится к линейной множественной модели с помощью преобразований $X_1^* = X, X_2^* = X^2, \ldots, X_m^* = X^m$.

5 . Показательная модель

Показательная функция

$$Y = \beta_0 e^{\alpha X} \tag{8.17}$$

также широко используется в эконометрическом анализе. В частности, она используется при моделировании ситуаций описывающих изменение переменной Y с постоянным темпом прироста по времени t (т.е. зависимая переменная $X = t$). Данная модель сводится к линейной с помощью логарифмирования

$$\ln Y = \ln \beta_0 + \alpha \cdot t.$$

Ряд экономических показателей моделируется функциями, являющимися композициями рассмотренных выше функций. Например, широко известна производственная функция Кобба-Дугласа с учетом научно технического прогресса

$$Y = A \cdot K^{\alpha} \cdot L^{\beta} \cdot e^{\gamma t}. \tag{8.18}$$

6 . Выбор формы модели

Многообразие и сложность экономических процессов предопределяет использование множества моделей, используемых для эконометрического анализа. С другой стороны, это существенно усложняет процесс нахождения максимально адекватной функциональной зависимости. Для случая парной регрессии такой выбор можно обосновать, анализируя расположение точек на корреляционном поле. В случае множественной регрессии такой анализ невозможен. При подборе подходящей модели учитывается опыт предыдущего моделирования, а также задачи, которые надлежит решить с помощью эконометрического анализа. Однако следует отметить, что идеальной модели не существует.

Чтобы выбрать качественную модель необходимо ответить на ряд вопросов, возникающих при моделировании:

1. Каковы признаки качественной модели?
2. Какие ошибки спецификации встречаются, и каковы последствия этих ошибок?
3. Как обнаружить ошибку спецификации?
4. Каким образом можно исправить ошибку спецификации и перейти к лучшей (более качественной) модели?

Признаки качественной модели:

- *скупость (простота)*;
- *единственность*;
- *максимальное соответствие*;
- *согласованность с теорией*;
- *хорошие прогнозные качества*.

Виды ошибок спецификации:

а) Отбрасывание значимой переменной.

Рассмотрим суть этой ошибки на следующем теоретическом примере. Пусть модель, отражающая некоторую экономическую зависимость, имеет вид:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \varepsilon. \tag{8.19}$$

Предположим, что этой модели соответствует эмпирическое уравнение регрессии:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \varepsilon. \tag{8.20}$$

Исследователь (по каким-то причинам) пришел к выводу, что объясняемая переменная Y зависит (в основном) только от одной переменной X_1. С учетом этого он рассматривает вместо модели (8.19) другую модель– только лишь с одной факторной переменной:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \delta, \tag{8.21}$$

и, соответственно, вместо эмпирического уравнения (8.20) – другое уравнение

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \delta. \tag{8.22}$$

При этом отбрасывание переменной X_2 может представлять собой ошибку. Одно из последствий этого – смещенность оценки b_1 для параметра α_1. Действительно, используя предпосылки метода наименьших квадратов, получаем

$$b_1 = \frac{S_{xy}}{S_{x^2}} = \frac{\operatorname{cov}(X_1, Y)}{D(X_1)} = \frac{\operatorname{cov}(X_1, \alpha_0) + \operatorname{cov}(X_1, \alpha_1 X_1) + \operatorname{cov}(X_1, \alpha_2 X_2) + \operatorname{cov}(X_1, \varepsilon)}{D(X_1)} =$$

$$= \frac{1}{D(X_1)} \left[0 + \alpha_1 D(X_1) + \alpha_2 \operatorname{cov}(X_1, X_2) + \operatorname{cov}(X_1, \varepsilon)\right] = \alpha_1 + \alpha_2 \frac{\operatorname{cov}(X_1, X_2)}{D(X_1)}.$$

Таким образом,

$$M(b_1) = \alpha_1 + \alpha_2 \frac{\operatorname{cov}(X_1, X_2)}{D(X_1)}.$$

Единственно возможным вариантом выполнения условия несмещенности b_1 является некоррелированность переменных X_1, X_2. Но в этом случае переход от модели (8.19) к модели (8.21) является оправданным.

б) Добавление незначимой переменной.

В некоторых случаях в уравнение регрессии включают слишком много факторных переменных. К чему это приводит? Хотя замена одной модели на другую, содержащую большее количество факторов, не приводит к нарушению условия несмещенности соответствующих оценок. Тем не менее, точность самих оценок уменьшается, поскольку стандартные ошибки (с.к.о.) увеличиваются (и это увеличение может быть существенным).

в) Выбор неправильной функциональной зависимости.

Проиллюстрируем эту ситуацию примером. Пусть корректная модель имеет вид (8.19). Любое эмпирическое уравнение регрессии с теми же переменными, но имеющее другой функциональный вид, может привести к искажению истинной зависимости. А именно, соответствующие оценки могут оказаться смещенными, статистические свойства самих оценок и других показателей качества уравнения регрессии могут ухудшиться. Это, прежде всего, связано с нарушением предпосылок метода наименьших квадратов. Прогнозные свойства модели в этом случае являются низкими.

При построении уравнений регрессии ошибки спецификации встречаются достаточно часто. Сложность процедуры исправления ошибок определяется типом ошибки и нашими знаниями об исследуемом объекте.

- Если в уравнении имеется одна незначимая переменная, то она обнаружит себя малым значением t-статистики.

- Если в уравнении несколько статистически незначимых объясняющих переменных, то следует построить новое уравнение регрессии, не содержащее этих переменных. Затем с помощью F-статистики нужно сравнить коэффициенты детерминации первоначального и нового уравнения регрессии:

$$F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{T - m - 1}{k}.$$

- Если в уравнении несколько незначимых переменных, то возможно это проявление *мультиколлинеарности*. Этот вопрос будет подробно рассмотрен ниже.

Осуществление указанных проверок имеет смысл только при условии, что функциональная зависимость, описывающая взаимосвязь переменных, выбрана правильно. С другой стороны, следует учитывать, что выбор модели не всегда определен однозначно. При определении качества модели обычно анализируются следующие параметры:

а) скорректированный коэффициент детерминации;

б) t-статистики при проверке качества коэффициентов регрессии;

в) статистика Дарбина-Уотсона при анализе коррелированности отклонений;

г) согласованность знаков коэффициентов с теорией;

д) прогнозные качества модели.

Для более детального анализа адекватности модели может быть предложено исследование остаточного члена модели.

Простейший способ такого анализа может быть реализован с помощью графического представления случайных отклонений.

Существует ряд других тестов обнаружения ошибок спецификации, среди которых можно выделить:

1. Тест Рамсея RESET (regression specification error test).
2. Критерий максимального правдоподобия (the likelihood ratio test).
3. Тест множителя Лагранжа (the Lagrange multiplier test).
4. Тест Хаусмана (the Hausman test).
5. Преобразование Бокса-Кокса (Box-Cox transformation).

Подробный анализ использования этих тестов выходит за рамки данного курса. Остановимся на основных идеях метода Рамсея RESET. Суть его состоит в следующем:

На первом этапе оценивают линейную регрессию между переменными

$$y_t = a_0 + a_1 x_{1t} + \ldots + a_m x_{mt} + \varepsilon_t$$

или

$$\hat{y}_t = a_0 + a_1 x_{1t} + \ldots + a_m x_{mt}.$$

Далее графически анализируют зависимость $\varepsilon_t(\hat{y}_t)$. Если данная зависимость может быть выражена некоторой формулой $\varepsilon = f(\hat{y})$, то ее вводят в исходное уравнение регрессии

$$y_t = a_0 + a_1 x_{1t} + \ldots + a_m x_{mt} + f(\hat{y}_t) + \varepsilon_t.$$

Сравнивают коэффициенты детерминации для начального и вновь построенного уравнения на основе следующей F-статистики:

$$F = \frac{R_2^2 - R_1^2}{1 - R_2^2} \cdot \frac{T - k}{r},$$

где T - число наблюдений, k - число параметров в новой модели, r - число новых регрессоров (дополнительных членов в уравнении). Если F-статистика окажется статистически значимой (принимает достаточно большое значение – большее, чем критическое значение F-статистики при выбранном уровне значимости), то это означает, что исходное уравнение было неправильно специфицировано. К сожалению, тест Рамсея не указывает напрямую спецификацию модели лучшую, чем исследуемая. Дополнением (или даже альтернативой) тесту Рамсея может служить тест множителя Лагранжа.