ГЛАВА 7. НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ И ИХ ЛИНЕАРИЗАЦИЯ

После отбора объясняющих переменных $X = (x_1, x_2, ..., x_k)^T$ и результативного показателя y решается задача сбора статистической информации и построения матрицы наблюдений (y, X), где i-я строка матрицы наблюдений имеет вид $(y_i, x_{i1}, ..., x_{ij}, ..., x_{ik})$ и x_{ij} — значение j-й объясняющей переменной для i-го наблюдения ($i = 1, 2, ..., n; j = 1, 2, ..., k$). В дальнейшем ключевой становится задача выбора параметрического семейства функций $f(X, \beta)$. Во многих практических случаях моделирование экономических зависимостей линейными уравнениями дает вполне удовлетворительные результаты и может использоваться для анализа и прогнозирования. Однако в силу многообразия и сложности экономических процессов ограничиться рассмотрением лишь линейных регрессионных моделей невозможно. Многие экономические зависимости не являются линейными по своей сути, и поэтому их моделирование возможно лишь на основе нелинейных уравнений регрессии. Выбор формы зависимости должен осуществляться на основании содержательного анализа исследуемого явления, а также по результатам анализа взаимосвязи переменных, входящих в модель. Например, зависимость между объемом произведенной продукции и основными факторами производства — трудом и капиталом (производственная функция Кобба–Дугласа), зависимость спроса на товары или услуги от цены и среднедушевого дохода семьи (функция спроса) являются нелинейными. Если в результате предварительного анализа приходят к выводу о нелинейном виде зависимости результативного показателя y от объясняющих переменных $x_1, x_2, ..., x_k$, то вначале делается попытка линеаризации уравнения, т. е. подбора таких преобразований переменных $y, x_1, ..., x_k$, которые позволили бы представить искомую зависимость в виде линейного соотношения между преобразованными переменными.

В таблице 7.1 приведем основные виды нелинейных регрессионных моделей, поддающихся линеаризации.

Таблица 7.1

Основные виды нелинейных регрессионных моделей, используемых для анализа социально-экономических явлений и процессов

Наименование модели	Вид модели
Полиномиальная	$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_k x^k + \varepsilon$
Гиперболическая (обратная)	$y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$

Окончание таблицы 7.1

Наименование модели	Вид модели
Обращенные полиномиальные функции	$y = x/(\beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k) + \varepsilon$, $y = 1/(\beta_0 + \beta_1 x + \varepsilon)$,
Логлинейная	$\ln y = \beta_0 + \beta_1 \ln x + \varepsilon$
Полулогарифмические	$\ln y = \beta_0 + \beta_1 x + \varepsilon$ $y = \beta_0 + \beta_1 \ln X + \varepsilon$
Степенная	$y = \beta_0 x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_k^{\beta_k} \varepsilon$
Экспоненциальная	$y = \beta_0 e^{a+\beta x} \varepsilon$
Функция Гомперца	$\ln y = b - ae^{-x} + \varepsilon$
Логистическая	$y = \dfrac{1}{\beta_0 + \beta_1 e^{-x} + \varepsilon}$
Экспоненциально-степенная	$y = \beta_0 e^{\beta_1 x_1} \cdot x_2^{\beta_2} \varepsilon$
Обращенная экспоненциальная	$y = 1/(\beta_0 + \beta_1 \varepsilon^{-x})$
Показательная	$y = \beta_0 a^{\beta_1 x} \varepsilon$

Современные компьютерные средства представляют большую возможность обработки информации и выбора вида уравнения регрессии экспериментальным методом, например путем сравнения величины остаточной дисперсии ($s^2_{ост}$), рассчитанной для различных моделей.

Если остаточная дисперсия оказывается примерно одинаковой для различных моделей, то предпочтение отдается более простому виду функций, так как они легче интерпретируются и требуют меньшего объема наблюдений. Если один и тот же фактор вводится в модель в разных степенях, то каждая степень рассматривается как самостоятельный фактор. Желательно, чтобы число наблюдений в три–пять раз превышало число параметров модели, подлежащих оцениванию.

Различают два класса нелинейных регрессионных моделей.

1. Нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам (β_j).

Примером могут служить следующие модели.

– полиномиальная

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon;$$

– гиперболическая (обратная модель)

$$y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon.$$

2. Нелинейные как относительно включенных в анализ объясняющих переменных, так и по оцениваемым параметрам.

Примером таких нелинейных моделей являются:

– степенная

$$y = \beta_0 x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_k^{\beta_k} \varepsilon;$$

– показательная

$$y_i = \beta_0 a^{\beta_1 x},$$

где a – положительная константа;

– экспоненциальная

$$y_i = \beta_0 e^{\beta_1 \xi} \varepsilon_i.$$

Для оценки параметров нелинейных моделей используют два подхода.

Первый подход основан на линеаризации модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую нелинейную зависимость представляют в виде линейного соотношения между преобразованными переменными.

Второй подход обычно применяется, когда подобрать соответствующее преобразование не удается. В этом случае применяют методы нелинейной оптимизации на основе исходных переменных.

Для линеаризации модели в рамках первого подхода могут использоваться модели нелинейные по переменным и нелинейные по параметрам.

Далее рассмотрим способы и возможности линеаризации нелинейных моделей.

7.1. Линеаризация нелинейных моделей

Для оценки неизвестных параметров в нелинейных моделях используют следующие методы:

– замена переменных;

– логарифмирование обеих частей уравнения;

– комбинированный.

Замена переменных

Суть первого метода состоит в замене нелинейных объясняющих переменных новыми линейными переменными и сведении нелинейной регрессии к линейной.

Полиномиальные модели

Полиномиальная модель k-го порядка $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon$ может быть приведена к линейному виду путем следующей замены переменных $x^j = x_j'$:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j' + \varepsilon.$$

Таким образом, полином любого порядка сводится к линейной регрессионной модели с ее методами оценивания параметров и проверки гипотез.

Среди нелинейных полиномиальных регрессионных моделей чаще всего используются параболические модели второй ($y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$) и третьей ($y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$) степени. Ограничения в использовании полиномов более высоких степеней связаны с требованием однородности исследуемой совокупности: чем выше порядок полинома, тем больше изгибов имеет кривая и соответственно менее однородна совокупность по результативному признаку.

Параболу второй степени целесообразно применять, если для определенного интервала значений фактора меняется характер связи с результатом: прямая связь меняется на обратную или обратная на прямую. В этом случае определяется значение фактора, при котором достигается максимальное (минимальное) значение результативного признака. Если же исходные данные не обнаруживают изменения направленности связи, то параметры параболы второй степени становятся трудно интерпретируемыми.

Чаще всего исследователь имеет дело лишь с отдельными сегментами параболы, а не с полной параболической формой. Кроме того, параметры параболической связи не всегда могут быть логически истолкованы. Поэтому если график зависимости не демонстрирует четко выраженной параболы (нет смены направленности связи признаков), то она может быть заменена другой нелинейной функцией, например степенной. Равносторонняя парабола может быть использована, например, для характеристики связи удельных расходов сырья, материалов, топлива с объемом выпускаемой продукции.

Гиперболические модели

Кривая регрессии гиперболического типа

$$f(x,\beta) = f(x,\beta_0,\beta_1) = \beta_0 + \beta_1 / x + \varepsilon$$

(рис. 7.1) характеризуется двумя асимптотами (т. е. прямыми, к которым график функции неограниченно приближается, не достигая их) — горизонтальной ($y = \beta_0$) и вертикальной ($x = 0$). С помощью преобразования объясняющей переменной $\tilde{x} = 1/x$ (т. е. при переходе к новой объясняющей переменной $\tilde{x}$) эта зависимость приводится к линейному виду $y = \beta_0 + \beta_1 \tilde{x} + \varepsilon$. Соответственно при вычислении МНК-оценок параметров β_0 и β_1 второй столбец матрицы X должен быть сформирован из чисел $1/x_1, 1/x_2, ..., 1/x_n$.

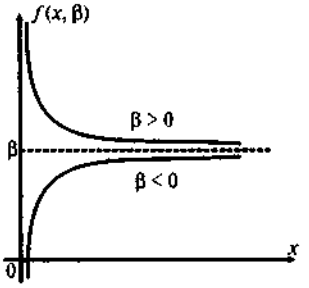

Рис. 7.1. Гиперболическая зависимость вида $f(x,\beta) = \beta_0 + \beta_1 / x$

Данная модель обычно применяется в тех случаях, когда неограниченное увеличение объясняющей переменной x асимптотически приближает зависимую переменную y к некоторому пределу. При $\beta_1 > 0$ имеем обратную зависимость, которая при $x \to \infty$ характеризуется нижней асимптотой (т. е. минимальным предельным значением y), оценкой которого служит параметр β_0. При $\beta_1 < 0$ имеем медленно повышающуюся функцию с верхней асимптотой при $x \to \infty$ (т. е. максимально предельным уровнем y), оценкой которого служит параметр β_0. Данная модель может быть использована, например, для характеристики связи удельных расходов сырья материалов и топлива с объемом выпускаемой продукции.

Классическим примером равносторонней гиперболы является *кривая Филлипса*, характеризующая нелинейное соотношение между нормой безработицы x и процентом прироста заработной платы y. Английский ученый экономист А. В. Филлипс, анализируя данные более чем за 100-летний период, в конце 50-х гг. XX в. установил обратную зависимость процента прироста заработной платы от уровня безработицы.

Если переменные x, y и случайные регрессионные остатки связаны между собой статистической зависимостью вида

$$y = \frac{1}{\beta_0 + \beta_1 x + \varepsilon}, \left(-\frac{\beta_0}{\beta_1} < x < \infty\right)$$

(рис. 7.2), то прийти к *линейной* модели $y' = \beta_0 + \beta_1 x + \varepsilon$ можно, если в качестве результирующего признака рассматривать переменную $y' = 1/y$. Отметим, что при вычислении МНК-оценок β_0 и β_1 надо использовать в качестве вектора наблюденных значений зависимой переменной вектор $y' = (1/y_1, 1/y_2, ..., 1/y_n)^T$.

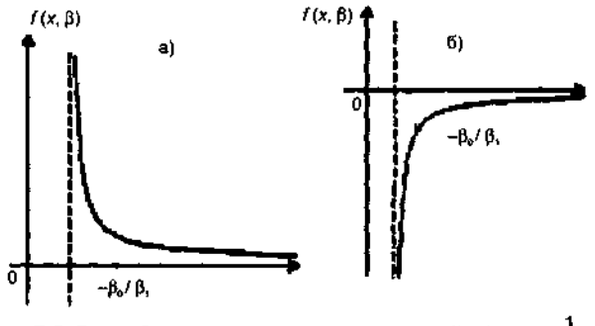

Рис. 7.2. Гиперболическая зависимость вида $f(x,\beta) = \dfrac{1}{\beta_0 x + \beta_1}$

а) случай $\beta_0 < 0$; $\beta_1 > 0$ (для $x > \dfrac{-\beta_0}{\beta_1}$); б) случай $\beta_0 > 0$; $\beta_1 < 0$ (для $x > \dfrac{-\beta_0}{\beta_1}$)

Если этап параметризации модели регрессии приводит нас к зависимости вида

$$y = \frac{x}{\beta_0 x + \beta_1 + \varepsilon x}, (-\frac{\beta_0}{\beta_1} < x < \infty)$$

(рис. 7.3), то линеаризацию исследуемой связи обеспечит переход к новым переменным $y' = 1/y$ и $x' = 1/x$. Легко увидеть, что эти переменные будут связаны между собой зависимостью вида $y' = \beta_0 + \beta_1 x' + \varepsilon$.

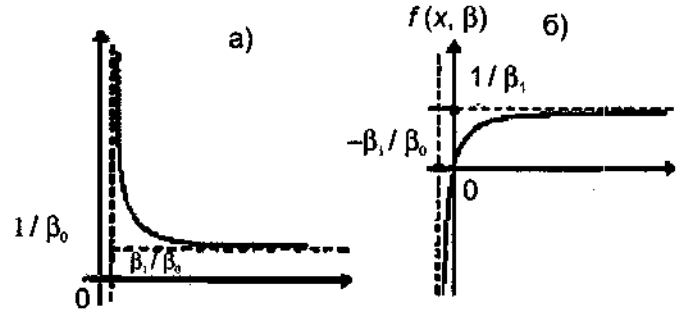

Рис. 7.3. Гиперболическая зависимость вида $f(x,\beta) = \frac{x}{\beta_0 x + \beta_1}$

а) случай $\beta_0 > 0$; $\beta_1 < 0$ (для $x > \frac{-\beta_1}{\beta_0}$); б) случай $\beta_0 > 0$; $\beta_1 > 0$ (для $x > \frac{-\beta_1}{\beta_0}$)

Очевидно, что при вычислении оценок β_0 и β_1 должны использоваться не наблюденные значения x_i и y_i, а обратные к ним величины $x'_i = 1/x_i$ и $y'_i = 1/y_i$.

Заметим, что функции, изображенные на рис. 7.1 (вариант $\beta_1 < 0$) и 7.3 б), используются в определенных ситуациях при построении так называемых кривых Энгеля, которые описывают зависимость спроса на определенный вид товаров или услуг y от уровня доходов x потребителей. При этом спрос определяется либо абсолютными, либо относительными (по отношению к общим потребительским расходам) расходами на данный вид товаров или услуг. Функции, изображенные на рис. 7.1 (вариант $\beta_1 > 0$), 7.2 а) и 7.3 а), могут оказаться полезными при изучении спроса на товар (y) в зависимости от его цены (x).

Функция $y = \beta_0 x / (x + \beta_1)$ с помощью замены переменных: $y' = 1/y$; $x' = 1/x$; $1/\beta_0 = \beta'_0$, $\beta_1/\beta_0 = \beta'_1$ может быть приведена к линейному виду: $y' = \beta'_0 + \beta'_1 x'$.

Полулогарифмические модели

Модели вида

$$\ln y = \beta_0 + \beta_1 x + \varepsilon \text{ и } y = \beta_0 + \beta_1 \ln X + \varepsilon$$

называются полулогарифмическими моделями.

Такие модели обычно используются в тех случаях, когда необходимо определить темп роста или прироста каких-либо экономических показателей. Например, при анализе банковского вклада по первоначальному вкладу и процентной ставке, при исследовании зависимости прироста объема выпуска от относительного (процентного) увеличения затрат ресурсов, бюджетного дефицита от темпа роста ВНП, темпа роста инфляции от объема денежной массы и т. д.

Полулогарифмическая модель $\ln y = \beta_0 + \beta_1 x + \varepsilon$ легко сводится к линейной модели путем следующей замены переменных: $y' = \ln y$.

Коэффициент β_1 в модели имеет смысл темпа прироста переменной y по переменной x, т. е. характеризует отношение относительного изменения y к абсолютному изменению x. Действительно, продифференцировав данное уравнение по x, получим:

$$\frac{1}{y}\frac{dy}{dx} = \beta_1 \to \beta_1 = \frac{Dy/y}{dx} = \frac{\text{относительное изменение } y}{\text{абсолютное изменение } x}.$$

Умножив β на 100, получим процентное изменение y (темп прироста переменной y).

Модель $y = \beta_0 + \beta_1 \ln x + \varepsilon$ используется обычно в тех случаях, когда необходимо исследовать, как процентное изменение независимой переменной влияет на абсолютное изменение зависимой переменной. Так, например, если y = ВНП (валовой национальный продукт), а $x = M$ (денежная масса), получим: ВНП $= \alpha + \beta \ln M + \varepsilon$. Из данной формулы следует, что увеличение предложения денег M на 1% ведет к росту среднего значения ВНП на $0{,}01\beta$.

Данная модель сводится к линейной путем замены переменных $x' = \ln x$. Коэффициент β определяет изменение переменной y вследствие единичного относительного прироста x (например, на 1%), т. е. характеризует отношение абсолютного изменения y к относительному изменению x.

Необходимо также отметить *недостаток метода замены переменных*, который связан с тем, что вектор неизвестных оценок b получается не из условия минимизации суммы квадратов отклонений исходных переменных, а из условия минимизации суммы квадратов отклонений преобразованных переменных. В связи с этим необходимо определенное уточнение полученных оценок.

Логарифмирование обеих частей уравнения

Оценка неизвестных коэффициентов в мультипликативных моделях более проблематична, так как непосредственное применение метода наименьших квадратов для их оценивания невозможно. В ряде случаев путем логарифмирования обеих частей уравнения мультипликативная модель (показательная, экспоненциальная) может быть приведена к линейному виду.

Степенные модели

Степенные модели вида $y = \beta_0 x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_p^{\beta_k} \varepsilon$ широко используются в эконометрических исследованиях. Прологарифмировав левую и правую

часть уравнения и обозначив $y' = \ln y$, $x'_j = \ln x_j$ и $\varepsilon' = \ln\varepsilon$ ($j = 1, 2, ..., k$), получим линейную модель регрессии относительно новых переменных y', x'_j и ε', а именно: $y' = \beta'_0 + \beta_1 x'_1 + ... + \beta_k x'_k + \varepsilon$, где $\beta'_0 = \ln\beta_0$.

При оценке параметров $\beta'_0, \beta'_1, ..., \beta_k$, участвующих в формулах МНК, вектор-столбец Y' и матрица X' определяются по исходным наблюдениям $\{x_{i1}, x_{i2}, ..., x_{ik}; y_i\}$, где $i = 1, 2, ..., n$ следующим образом: $Y' = (\ln y_1, \ln y_2, ..., \ln y_n)^T$, а j-й столбец матрицы X' есть $(\ln x_{1j}, \ln x_{2j}, ..., \ln x_{nj})^T$, $j = 1, 2, ..., k$ (напомним, что первый столбец матрицы X' составлен из одних единиц). Графики зависимостей данного типа для случая $k = 1$ представлены на рис. 7.4.

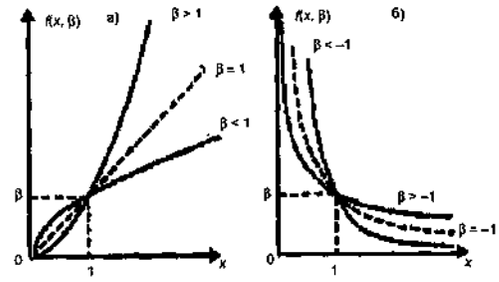

Рис. 7.4. Степенная зависимость вида $f(x,\beta) = \beta_0 x_1^{\beta_1}$
а) случай $\beta_1 > 0$; б) случай $\beta_1 < 0$

К классу степенных функций относятся кривые спроса и предложения, кривые Энгеля, производственные функции, кривые освоения для характеристики связи между трудоемкостью продукции и масштабами производства в период освоения и выпуска нового вида изделий, а также зависимость валового национального дохода от уровня занятости.

Если изучается зависимость спроса на товар или услугу y от его цены x (в данном случае $\beta_1 < 0$) или от дохода X (в данном случае $\beta_1 > 0$), то при такой интерпретации переменных степенная функция называется функцией Энгеля.

Если степенная функция отражает зависимость объема выпуска продукции y от использования ресурса x (в которой $0 < \beta_1 < 1$), то она называется производственной функцией. Например, производственная функция Кобба—Дугласа связывает объем производства Y с затратами капитала K и затратами труда L:

$$Y = AK^{\alpha} L^{\beta}\varepsilon.$$

Автономная зависимость от времени выражена в коэффициенте научно-технического прогресса A. Показатели α и β являются коэффициентами частной эластичности объема производства Y соответственно по затратам капитала K и труда L. Это означает, что при увеличении затрат капитала (труда) на 1% объем производства увеличивается на α% (β%).

Сумма коэффициентов является важным экономическим показателем, который носит название *отдача от масштаба*. При $\alpha+\beta > 1$ имеет место *возрастающая отдача от масштаба* (увеличение объема выпуска больше увеличения затрат ресурсов). При $\alpha + \beta < 1$ — *убывающая отдача от масштаба* (увеличение объема выпуска меньше увеличения затрат ресурсов). При $\alpha + \beta = 1$ говорят о *постоянной отдаче от масштаба* (во сколько раз увеличиваются затраты ресурсов, во столько же раз увеличивается выпуск). Функцию Кобба—Дугласа можно также представить в виде

$$\frac{Y}{L} = \frac{AK^{\alpha}}{L^{\alpha}}\varepsilon.$$

Таким образом, получается зависимость производительности труда (Y/L) от его капиталовооруженности (K/L). Для оценки параметров данной модели ее логарифмируют с целью приведения к линейному виду (для i-го наблюдения):

$$\ln(Y / L)_i = \ln A + \alpha \ln(K / L)_i + \ln\varepsilon_i, \quad i = 1, 2, ..., n.$$

Рассмотренные модели являются нелинейными относительно оцениваемых параметров, так как включают параметры α и β мультипликативно. Однако их можно считать внутренне линейными, так как логарифмирование уравнения приводит его к линейному виду.

Зависимости показательного (экспоненциального) типа

Достаточно широкий класс экономических показателей характеризуется приблизительно постоянным темпом относительного прироста во времени. Этому соответствует следующая форма зависимости показателя y от времени x:

$$y = \beta_0 e^{\beta_1 x + \varepsilon} \quad (\text{где } e = 2{,}7182818).$$

Действительно, если пренебречь влиянием случайной остаточной компоненты ε (т. е. положить $\varepsilon = 0$), то результаты расчетов дают кривые, представленные на рис. 7.5а.

$$\frac{dy}{dx} = \beta_1 \beta_0 e^{\beta_1 x} = \beta_1 y,$$

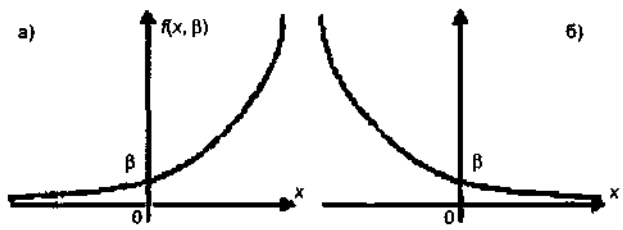

Рис. 7.5. Показательная (экспоненциальная) зависимость вида $f(x,\beta) = \beta_0 e^{\beta_1 x}$
а) случай $\beta_1 > 0$; б) случай $\beta_1 < 0$

Относительный прирост y за единицу времени (т. е. за единицу «количества» x) определяется выражением $\frac{dy}{dx} / y = \beta_1$ (в долях y).

Легко увидеть, что переход к новой переменной $y' = \ln y$ позволяет свести исследуемую зависимость к линейному виду:

$$y' = \beta'_0 + \beta_1 x + \varepsilon, \text{ где } \beta'_0 = \ln \beta_0.$$

Располагая наблюдениями $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ и формируя вектор-столбец Y' из $\ln y_1, \ln y_2, \ldots, \ln y_n$, мы с помощью МНК можем построить оценки b'_0 и b_1 параметров β'_0 и β_1, а затем получить оценку $b_0 = e^{b'_0}$ для параметра β_0 исходного уравнения.

Комбинированный метод

Если в результате параметризации модели мы пришли к необходимости исследовать экспоненциальную статистическую зависимость вида

$$y = \beta_0 e^{\frac{\beta_1}{x} + \varepsilon}$$

(рис. 7.6), то линеаризация искомой зависимости достигается с помощью логарифмирования и замены переменных: $y' = \ln y$, $x' = 1/x$. Очевидно, в терминах переменных (x', y') исследуемая модель будет линейной:

$$y' = \beta'_0 + \beta_1 x' + \varepsilon,$$

где $\beta'_0 = \ln \beta_0$. Соответственно вектор-столбец Y' и матрица X', участвующие в формулах МНК, определяются по исходным наблюдениям $(x_i, y_i), i = 1, 2, ..., n$ следующим образом:

$$Y' = (\ln y_1, \ln y_2, \ldots, \ln y_n)^T; \; X' = \begin{pmatrix} 1 & \ldots & 1 \\ \frac{1}{x_1} & \ldots & \frac{1}{x_n} \end{pmatrix}^T$$

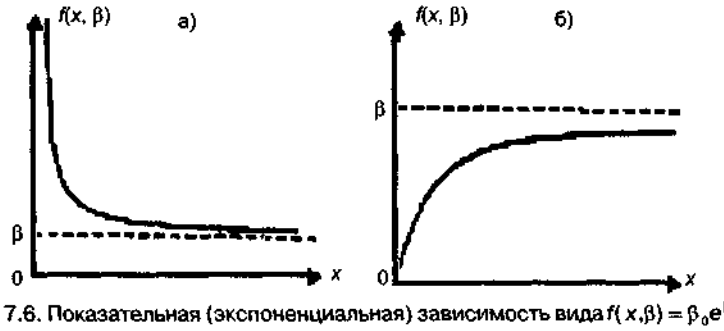

Рис. 7.6. Показательная (экспоненциальная) зависимость вида $f(x, \beta) = \beta_0 e^{\beta_1 / x}$
а) случай $\beta_1 > 0$; б) случай $\beta_1 < 0$

Логистическая кривая

Весьма гибкую форму параметризации искомой регрессионной модели представляет один из частных случаев, так называемой логистической кривой (рис. 7.7):

$$y = \frac{1}{\beta_0 + \beta_1 e^{-x} + \varepsilon} (-\infty < x < \infty).$$

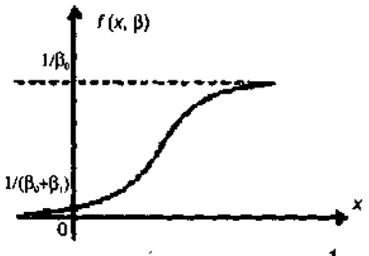

Рис. 7.7. Логистическая кривая вида $f(x,\beta) = \dfrac{1}{\beta_0 + \beta_1 e^{-x}}$. (Случай $\beta_1 > 0$)

Кривая $f(x,\beta)$ имеет две горизонтальные асимптоты $y = 0$ и $y = 1/\beta_0$ и «точку перегиба» $x_0 = \ln(\beta_1/\beta_0, y_0 = 1/2\beta_0)$. Линеаризация этой зависимости производится с помощью перехода к переменным $y' = 1/y$ и $x' = e^{-x}$. Соответственно, вектор-столбец Y' и матрица X', участвующие в формулах МНК, определяются по исходным наблюдениям (x_i, y_i), $i = 1, 2, ..., n$ следующим образом:

$$Y' = (1/y_1, 1/y_2, ..., 1/y_n)^T, \quad X' = \begin{pmatrix} 1 \ldots\ldots 1 \\ e^{-x_1} \ldots e^{-x_n} \end{pmatrix}^T.$$

Логистические кривые используются для описания поведения показателей, имеющих определенные «уровни насыщения», например для описания зависимости спроса на товар y от дохода x, развития производства новых товаров, роста численности населения и т. д. Впервые такая кривая была использована А. Кетле (1796–1874) для расчета численности населения.

При использовании линеаризуемых функций, использующих преобразование зависимой переменной y, необходимо тщательно проверять наличие предпосылок МНК (чтобы они не нарушались при преобразовании). При нелинейных соотношениях рассматриваемых признаков, приводимых к линейному виду, возможно интервальное оценивание параметров нелинейной функции. Так, для показательной кривой сначала строят доверительные интервалы для параметров нового преобразованного уравнения $\ln y = \ln a + x \ln b$, т. е. для $\ln a$ и $\ln b$. Далее с помощью обрат-

ного преобразования определяются доверительные интервалы для параметров в исходном соотношении.

Логлинейная модель

Модель вида $y_t = y_0(1+r)^t\varepsilon_t$ широко используется в банковском и финансовом анализе. Здесь y_0 – начальная величина переменной y (например, первоначальный вклад в банке); r — сложный темп прироста величины y (процентная ставка); y_t — значение величины y в момент времени t (вклад в банке в момент времени t). Данная модель легко сводится к полулогарифмической модели $\ln y = \beta_0 + \beta_1 x + \varepsilon$. Действительно, прологарифмировав данное уравнение, получим:

$$\ln y_t = \ln y_0 + t\ln(1+r) + \ln\varepsilon_t.$$

Обозначив: $\ln y_0 = \beta_0$, $\ln(1+r) = \beta_1$ и $\ln\varepsilon_t = \varepsilon'_t$, получим следующее уравнение

$$\ln y_t = \beta_0 + \beta_1 t + \varepsilon'_t.$$

Дополнительное случайное слагаемое ε'_t используется в силу возможной изменчивости процентной ставки.

Отметим, что в моделях, нелинейных по оцениваемым параметрам, но приводимых к линейному виду, МНК применяется к преобразованным уравнениям. В таких моделях преобразованию подвергается *результативный признак* y в отличие от нелинейных моделей 1-й группы, в которых результативный признак y остается неизменным, а преобразуется только факторный признак.

Если в линейной модели и моделях нелинейных по переменным при оценке параметров исходят из критерия

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min_{\beta},$$

то в моделях, нелинейных по оцениваемым параметрам, требования МНК применяются не к исходным данным результативного признака, а к их преобразованным величинам ($\ln y$ или $1/y$). Это означает, что оценка неизвестных коэффициентов регрессии в уравнении основывается на минимизации суммы квадратов отклонения логарифмов:

$$\sum_{i=1}^{n}(\ln y_i - \ln\hat{y}_i)^2 \to \min_{\beta}.$$

Соответственно если в линейных моделях и моделях, нелинейных по переменным, $\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$, то в моделях, нелинейных по оцениваемым параметрам, $\sum_{i=1}^{n}(\ln y_i - \ln\hat{y}_i) = 0$, а $\sum_{i=1}^{n}(y_i - anti\ln\hat{y}_i) = \sum(y_i - \hat{y}_i) \neq 0$. Вследствие этого оценка параметров с помощью МНК для нелинейных моделей, внутренне линейных, оказывается смещенной.

Расчет коэффициентов эластичности

Одним из важнейших этапов в построении регрессионных моделей является интерпретация полученных результатов. Рассмотрим расчет коэффициентов эластичности и их интерпретацию для различных форм регрессионных моделей (табл. 7.2).

Таблица 7.2

Коэффициенты эластичности для ряда математических функций

Функция y	Первая производная y'_x	Коэффициент эластичности Э
$y=a+bx+\varepsilon$ линейная	b	$\frac{b\cdot x}{a+b\cdot x}$
$y=a+bx+cx^2+\varepsilon$ парабола	$b+2\cdot c\cdot x$	$\frac{(b+2\cdot c\cdot x)\cdot x}{a+b\cdot x+c\cdot x^2}$
$y=a+\frac{b}{x}+\varepsilon$ гипербола	$-\frac{b}{x^2}$	$\frac{-b}{a\cdot x+b}$
$y=a\cdot b^x\cdot\varepsilon$ показательная	$\ln b\cdot a\cdot b^x$	$x\cdot\ln b$
$y=a\cdot x^b\cdot\varepsilon$ степенная	$a\cdot b\cdot x^{b-1}$	b
$y=a+b\cdot\ln x+\varepsilon$ полулогарифмическая	$\frac{b}{x}$	$\frac{b}{a+b\cdot\ln x}$
$y=\frac{a}{1+b\cdot e^{-cx+\varepsilon}}$ логистическая	$\frac{a\cdot b\cdot c\cdot e^{-cx}}{(1+b\cdot e^{-cx})^2}$	$\frac{c\cdot x}{\frac{1}{b}\cdot e^{cx}+1}$
$y=\frac{1}{a+b\cdot x+\varepsilon}$ обратная	$\frac{-b}{(a+b\cdot x)^2}$	$\frac{-b\cdot x}{a+b\cdot x}$

Среди нелинейных функций, которые могут быть приведены к линейному виду, в экономических исследованиях широко используются степенные функции, что связано с четкой интерпретацией параметра b, представляющего собой коэффициент *эластичности*. Величина коэффициента b показывает, на сколько процентов изменится результативный признак y, если факторный признак изменится на 1%. О правомерности подобного толкования параметра b для степенной функции $y=a+x^b$ можно судить, если рассмотреть формулу расчета коэффициентов эластичности

$$Э=f'(x)\frac{x}{y},$$

где $f'(x)$ — первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи.

Для степенной функции она составит: $f'(x) = a \cdot b \cdot x^{b-1}$. Соответственно коэффициент эластичности равен:

$$Э = a \cdot b \cdot x^{b-1} \cdot \frac{x}{a \cdot x^b} = \frac{a \cdot b \cdot x^b}{a \cdot x^b} = b.$$

Таким образом, для степенной функции коэффициент эластичности представляет собой постоянную величину, равную параметру b. В других функциях коэффициент эластичности зависит от значений фактора x. Так, для линейной регрессии первая производная $\hat{y}_x = a + bx$ функции и эластичность следующие:

$$f'(x) = b \text{ и } Э = b \cdot \frac{x}{a + b \cdot x}.$$

В силу того, что коэффициент эластичности для линейной функции не является величиной постоянной, а зависит от соответствующего значения x, обычно рассчитывается *средний показатель эластичности* по формуле

$$\overline{Э} = b \cdot \frac{\bar{x}}{\bar{y}}.$$

Несмотря на широкое применение в эконометрике коэффициентов эластичности, возможны случаи, когда их расчет не имеет экономического смысла. Это происходит, когда для рассматриваемых признаков бессмысленно определять изменения значений в процентах. Например, на сколько процентов изменится урожайность пшеницы, если качество почвы, измеряемое в баллах, изменится на 1%, и т. д. В таких случаях степенная функция, даже если она оказывается наилучшей по формальным соображениям (исходя из наименьшего значения остаточной вариации), не может быть экономически интерпретирована.

7.2. Гармонический анализ

При моделировании рядов динамики с помощью линейной, полиномиальной или экспоненциальной функции не всегда удается получить удовлетворительный результат. Наличие во временном ряду периодической составляющей $S(t)$ делает непригодными для использования данные модели, так как МНК-оценки будут обладать «плохими» свойствами. Необходимо учесть и периодические колебания уровней ряда. В таких случаях следует использовать гармонический анализ, целью которого является выявление и измерение периодических колебаний.

Функцию, заданную в каждой точке изучаемого интервала времени, можно представить бесконечным рядом синусоидальных и косинусоидальных функций. Нахождение конечной суммы уровней с использовани-

ем функций косинусов и синусов времени называется гармоническим анализом (анализом Фурье).

Уравнение, на основе которого строятся гармоники, имеет вид

$$v_t = y_t - \tilde{y}_t = \sum_{j=1}^{k} \beta_j' \sin\omega_j t + \sum_{j=1}^{k} \beta_j'' \cos\omega_j t + \varepsilon_t$$

где $\tilde{y}_t = \beta_0 + \beta_1 t$, или парабола второго или третьего порядка; j — номер гармоники; $\omega_j = \frac{2\pi}{n} j$ — угловая частота j-й гармоники; k — номер гармоники; ε_t — случайная ошибка; n — период (частота) повторяемости сезонных колебаний.

В статистической практике построение моделей с периодическими колебаниями чаще всего используется, когда явления подвержены периодическим колебаниям с периодом $n = 12$ (т. е. изучаются помесячная динамика показателя за несколько лет).

Значения $\omega(j)$ при $n = 12$ для разных гармоник представлены в следующей таблице.

№ гармоники Фурье	Значение ω (j)
0	0
1	$(2\pi/12) = (\pi/6)$
2	$(4\pi/12) = (\pi/3)$
3	$(6\pi/12) = (\pi/2)$
4	$(8\pi/12) = (3\pi/4)$
5	$(10\pi/12) = (5\pi/6)$
6	$(12\pi/12) = (\pi)$
7	$(14\pi/12) = (7\pi/6)$
8	$(16\pi/12) = (4\pi/3)$
9	$(18\pi/12) = (3\pi/2)$
10	$(20\pi/12) = (5\pi/3)$
11	$(22\pi/12) = (11\pi/6)$
12	$(24\pi/12) = (2\pi)$

Гармонический анализ проводится поэтапно: начинается с построения значимой модели, содержащей первую гармонику (т. е. $\omega(j) = 2\pi/12$). Модель, содержащая каждую последующую гармонику Фурье, включает также и предыдущую гармонику. Так, вторая гармоника имеет в своем составе первую, третья — вторую и т. д. Увеличивая порядок гармоники, ис-

следователь увеличивает частоту колебаний $\omega(j)$ (т. е. график синусоиды становится более «плотным», учащается).

В практике статистического моделирования, как правило, используется не более четырех первых гармоник Фурье. Выбор гармоники, наилучшим образом аппроксимирующей фактические данные, определяется из условия оптимизации среднеквадратической ошибки модели ($S \to \min$) и других характеристик модели.

7.2.1. Примеры

Регрессионная модель объема продаж торгового дома, включающая линейную и гармонические составляющие

На основании данных объема продаж торгового дома за ($n = 12$) месяцев (табл. 7.3) необходимо построить регрессионную модель зависимости объема продаж от времени.

Таблица 7.3

Данные объема продаж торгового дома, млн руб.

Месяц	t	y_t
Январь	1	200
Февраль	2	310
Март	3	320
Апрель	4	260
Май	5	190
Июнь	6	210
Июль	7	310
Август	8	410
Сентябрь	9	430
Октябрь	10	370
Ноябрь	11	300
Декабрь	12	320

Решение

Первоначально аппроксимируем временной ряд линейным уравнением регрессии вида $y_t = \beta_0 + \beta_1 t$.

Используя ППП SPSS, получим оценку уравнения регрессии, найденную с помощью метода наименьших квадратов, следующего вида

$$\hat{y}_t = \underset{(5,3)}{230} + \underset{(1,9)}{11{,}15}t$$

$$R^2 = 0{,}267;\ \hat{s} = 69{,}9;\ F = 3{,}64;\ DW = 1{,}01.$$

В скобках указаны расчетные значения t-статистики для проверки значимости коэффициентов уравнения. Уравнение значимо и содержит все значимые по t-критерию коэффициенты β_j. Критическое значение $t_{кр}$= 1,76, найденное при α = 0,1 и $\nu = n - 2 = 10$ по таблице t-распределения, меньше расчетного $t_j = \frac{b_j}{\hat{S}_{b_j}}$, где j = 0;1.

Статистические характеристики адекватности уравнения: $\hat{s}$ = 69,9; R^2=0,267, $\left|\bar{\delta}\right| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| * 100\% = 19{,}92\%$ и DW = 1,01 свидетельствуют о наличии положительной коррелированности случайных остатков и о недостаточно хороших аппроксимирующих свойствах модели.

Временной ряд объема продаж торгового дома y_t графически представлен на рис. 7.8.

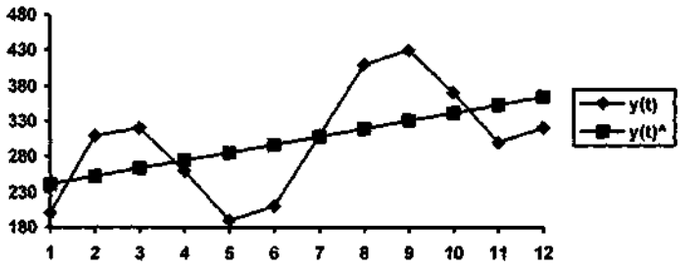

Рис 7.8. Временной ряд объема продаж торгового дома y_t и линейный тренд $\hat{y}_t$

Из рис. 7.8 видно, что ряд характеризуется циклическими колебаниями и, по всей вероятности, недостаточно его аппроксимировать только с применением модели линейного тренда. Можно предположить наличие периодической (сезонной) составляющей временного ряда. Временной диапазон n = 12 вмещает в себя два полных периода циклических колебаний анализируемого показателя. Отсюда можно предположить, что для адекватного описания ряда y_t достаточно второй гармоники (j = 2) с угловой частотой $\omega_2 = \frac{2\pi}{12}\cdot 2$.

Первоначально включим в модель объема продаж две гармоники с угловыми частотами $\omega_1 = \frac{2\pi}{12}1$ и $\omega_2 = \frac{2\pi}{12}2$ и будем строить линейное уравнение регрессии относительно переменных:

$$t, \sin\frac{2\pi}{12}t, \cos\frac{2\pi}{12}t, \sin\frac{4\pi}{12}t, \cos\frac{4\pi}{12}t.$$

t	$\sin\frac{2\pi}{12}t$	$\cos\frac{2\pi}{12}t$	$\sin\frac{4\pi}{12}t$	$\cos\frac{4\pi}{12}t$
1	0,5	0,866 025	0,866 025	0,5
2	0,866 025	0,5	0,866 025	–0,5
3	1	6,13E–17	1,23E–16	–1
4	0,866 025	–0,5	–0,866 03	–0,5
5	0,5	–0,866 03	–0,866 03	0,5
6	1,23E–16	–1	–2,5E–16	1
7	–0,5	–0,866 03	0,866 025	0,5
8	–0,866 03	–0,5	0,866 025	–0,5
9	–1	–1,8E–16	3,68E–16	–1
10	–0,866 03	0,5	–0,866 03	–0,5
11	–0,5	0,866 025	–0,866 03	0,5
12	–2,5E–16	1	–4,9E–16	1

Получим:

$$\hat{y}^{*\prime} = 183{,}2 + \underset{(28,7)}{18{,}36}t + \underset{(0,55)}{1{,}52}\sin\frac{2\pi}{12}t + \underset{(0,52)}{0{,}81}\cos\frac{2\pi}{12}t + \underset{(26,5)}{47{,}67}\sin\frac{4\pi}{12}t - \underset{(-54,2)}{84{,}19}\cos\frac{4\pi}{12}t;$$

$$R^2 = 0{,}997;\ \hat{s} = 3{,}32;\ F = 1106{,}024;\ DW = 3{,}3.$$

Данное уравнение содержит два незначимых коэффициента регрессии, относящихся к 1-й гармонике. Характеристики адекватности уравнения значительно лучше предыдущего.

После исключения незначимых коэффициентов и реализации пошагового алгоритма регрессионного анализа было получено следующее уравнение:

$$\hat{y}_t^* = \underset{(86,7)}{184{,}79} + \underset{(61,4)}{18{,}11}t + \underset{(33,5)}{47{,}245}\sin\frac{4\pi}{12}t - \underset{(-62,3)}{83{,}94}\cos\frac{4\pi}{12}t;$$

$$R^2 = 0{,}998;\ \hat{s} = 3{,}22;\ DW = 3{,}025;\ \bar{\delta} = 0{,}75\%.$$

Все входящие в уравнение коэффициенты значимы при $\alpha = 0{,}05$ ($t_{кр} = 2{,}306$). Полученное уравнение характеризуется остаточным среднеквадратическим отклонением $\hat{s} = 3{,}22$; множественным коэффициентом детерминации $R^2 = 0{,}998$; средней относительной ошибкой аппроксимации $|\bar{\delta}| = 0{,}75\%$ и статистикой Дарбина–Уотсона $DW = 3{,}025$, что свидетельствует об адекватности модели.

Статистические характеристики свидетельствуют об адекватности модели.

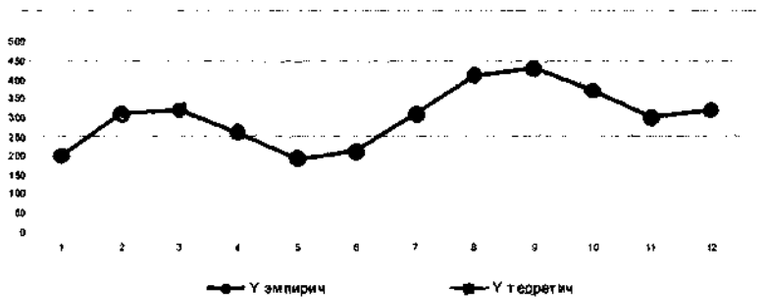

Рис. 7.9. Аппроксимация рядом Фурье исходного временного ряда

В данном примере наглядно продемонстрирована возможность моделирования процессов, содержащих периодические составляющие, путем совместного использования линейного уравнения регрессии и гармоник Фурье.

Степенные регрессионные модели объема потребления мяса цыплят

Данные по Великобритании за (n = 20) лет о потреблении цыплят (y), среднедушевом доходе (x_1), стоимости 1 фунта цыплят (x_2), стоимости 1 фунта свинины (x_3) и стоимости 1 фунта говядины (x_4), представлены в табл. 7.4.

Требуется построить и сравнить уравнения регрессии вида

а) $\tilde{y} = \beta_0 \cdot (x_2)^{\beta_2}$ — функция спроса;

б) $\tilde{y} = \beta_0 \cdot (x_1)^{\beta_1}$ — функция потребления;

в) $\tilde{y} = \beta_0 \cdot (x_1)^{\beta_1} (x_2)^{\beta_2}$ — функция спроса и потребления;

г) $\tilde{y} = \beta_0 \cdot (x_2)^{\beta_2} (x_3)^{\beta_3} (x_4)^{\beta_4}$ — функция спроса с учетом цены на товарозаменители.

Таблица 7.4

Исходные данные

t	y	x_1	x_2	x_3	x_4
1	30,8	459,7	39,5	55,3	79,2
2	31,2	492,9	37,3	54,7	77,4
3	33,3	528,6	38,1	63,7	80,2
4	35,6	560,3	39,3	69,8	80,4
5	36,4	624,6	37,8	65,9	83,9

Окончание таблицы 7.4

t	y	x_1	x_2	x_3	x_4
6	36,7	666,4	38,4	64,5	85,5
7	38,4	717,8	40,1	70,0	93,7
8	40,4	768,2	38,6	73,2	106,1
9	40,3	843,3	39,8	67,8	104,8
10	41,8	911,6	39,7	79,1	114,0
11	40,4	931,1	52,1	95,4	124,1
12	40,7	1 021,5	48,9	94,2	127,6
13	40,1	1 165,9	58,3	123,5	142,9
14	42,7	1 349,6	57,9	129,9	143,6
15	44,1	1 449,4	56,5	117,6	139,2
16	46,7	1 575,5	63,7	130,9	165,5
17	50,6	1 759,1	61,6	129,8	203,3
18	50,1	1 994,2	58,9	128,0	219,6
19	51,7	2 258,1	66,4	141,0	221,6
20	52,9	2 478,7	70,4	168,2	232,6

Решение

При построении степенных уравнений регрессии использовался пакет SPSS, позволяющий путем применения логарифмического преобразования к объясняемой переменной y и всем объясняющим переменным x_j линеаризовать модель для нахождения оценок параметров уравнения регрессии с помощью метода наименьших квадратов. Таким образом, были получены следующие модели:

Функция спроса

$$\tilde{y} = \beta_0 \cdot (x_2)^{\beta_2}.$$

Сначала построим линейную регрессию для прологарифмированных исходных переменных:

$$\ln \tilde{y} = \ln \beta_0 + \beta_2 \ln x_2.$$

Обозначив $y' = \ln \hat{y}$, $\beta'_0 = \ln \beta_0$, $x'_2 = \ln x_2$, получим линейную модель: $y' = \beta'_0 + \beta_2 \cdot x'_2$. Найдем для нее оценки параметров β'_0, β_2. Затем, используя ППП SPSS, получим оценку уравнения регрессии, найденную с помощью метода наименьших квадратов, следующего вида

$$\tilde{y} = 4{,}185 \cdot (x_2)^{\underset{(6{,}53)}{0{,}588}} \qquad (7.1)$$

$$\hat{s} = 2{,}1; \; DW = 0{,}67; \; \bar{\delta} = 0{,}1\%.$$

В скобках указано значение t-критерия для проверки значимости коэффициента регрессии. Уравнение экономически не интерпретируемо, хотя все его коэффициенты значимы. Вряд ли можно согласиться со знаком «+» коэффициента эластичности $э_2 = 0{,}588$, из чего следует, что с ростом цены на цыплят на 1% спрос на них увеличится в среднем на 0,588%. В модели не учитывается инфляционный процесс, рост среднедушевых доходов населения, происходящий в стране за рассматриваемые 20 лет, хотя за это время среднедушевой доход вырос в 5,4 раза, а стоимость цыплят — в 1,8 раза (табл. 7.3).

Функция потребления:

$$\tilde{y} = 5{,}347 \cdot \underset{(17{,}3)}{(x_1)^{0{,}295}} \qquad (7.2)$$

$$\hat{s} = 2{,}154;\ DW = 0{,}658;\ \bar{\delta} = 3{,}07\%.$$

Модель интерпретируема. Однако вывод, что при увеличении среднедушевого дохода на 1% потребление цыплят в среднем растет на 0,295%, не учитывает динамику цены за рассматриваемый период. Модель характеризуется оценкой остаточной дисперсии $\hat{s}^2 = 2{,}154$, средней относительной ошибкой аппроксимации $\bar{\delta} = 3{,}07\%$ и статистикой Дарбина—Уотсона $DW = 0{,}658$, которая свидетельствует о положительной автокоррелированности случайных регрессионных остатков.

Более интересны в содержательном плане следующие две модели.

Функция спроса-потребления

$$\tilde{y} = 7{,}516 \cdot \underset{(11{,}9)}{(x_1)^{0{,}428}} \cdot \underset{(-3{,}92)}{(x_2)^{-0{,}325}} \qquad (7.3)$$

$$\hat{s}^2 = 1{,}246;\ DW = 1{,}77;\ \bar{\delta} = 2{,}21\%;\ F = 54{,}7.$$

Из модели следует ($э_1 = 0{,}428$), что с ростом среднедушевого дохода на 1% при неизменной стоимости цыплят их потребление в среднем увеличится на 0,428%. В модели (7.2), представленной функцией потребления, аналогичный вывод делался на фоне роста стоимости цыплят, чем обусловлена разница в коэффициентах эластичности $э_1$.

Увеличение же стоимости цыплят на 1% при неизменном среднедушевом доходе приводит к уменьшению потребления в среднем на 0,325%. Этот вывод интересно сравнить с выводом, сделанным по модели функции спроса (7.1). Если в модели (7.1) $э_2 = 0{,}588$ есть парный коэффициент эластичности, то в (7.3) $э_2 = -0{,}325$ — это частный коэффициент эластичности. При этом парный и частный коэффициенты эластичности имеют разные знаки.

Модель функции спроса-потребления обладает хорошими аппроксимирующими свойствами, о чем свидетельствуют статистические характеристики адекватности, приведенные под уравнением. Значение критерия

Дарбина—Уотсона указывает на отсутствие автокоррелированности остатков (рис. 7.10).

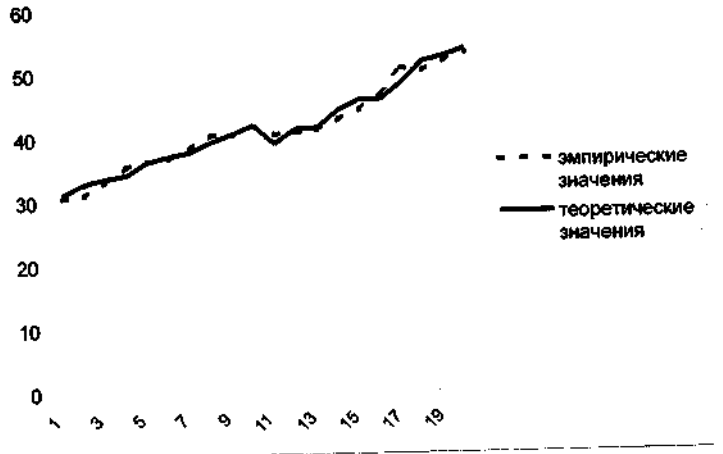

Рис. 7.10. Эмпирические и теоретические значения временного ряда

Функция спроса с учетом цены на товарозаменители

$$\tilde{y} = \beta_0 \cdot (x_2)^{\beta_2} \cdot (x_3)^{\beta_3} \cdot (x_4)^{\beta_4}.$$

Прологарифмировав исходную функцию, получим:

$$\ln \tilde{y} = \ln \beta_0 + \beta_2 \cdot \ln x_2 + \beta_3 \cdot \ln x_3 + \beta_4 \cdot \ln x_4.$$

Обозначив $y' = \ln \hat{y}$, $\beta'_0 = \ln \beta_0$, $x'_2 = \ln x_2$, $x'_3 = \ln x_3$, $x'_4 = \ln x_4$, получим линейную модель: $y' = \beta'_0 + \beta_2 \cdot x'_2 + \beta_3 \cdot x'_3 + \beta_4 \cdot x'_4$. Найдем для нее оценки параметров β'_0, β_2, β_3 и β_4.

Степенная регрессионная модель потребления цыплят с учетом цен на товарозаменители (свинину и говядину) имеет вид

$$\tilde{y} = 11{,}080 \cdot \underset{(-3{,}71)}{(x_2)^{-0{,}630}} \cdot \underset{(2{,}65)}{(x_3)^{0{,}345}} \cdot \underset{(5{,}99)}{(x_4)^{0{,}455}} \qquad (7.4)$$

$$\hat{s}^2 = 1{,}994; \; DW = 1{,}7; \; \bar{\delta} = 3{,}03\%$$

Модель характеризует зависимость объема потребления от стоимости цыплят (x_2) и цены на такие замещающие продукты, как свинина (x_3) и говядина (x_4). Из модели следует, что при неизменной стоимости двух сопутствующих продуктов увеличение на 1% стоимости цыплят приводит к снижению их потребления в среднем на 0,63%, а увеличение стоимости свинины или говядины на 1% при неизменности цен на остальные входящие в модель продукты приводит к росту потребления цыплят в среднем соответственно на 0,345% и 0,455%.

Модель (7.4) менее адекватна данным наблюдений, чем модель (7.3), о чем свидетельствует сравнение их статистических характеристик адекватности.

Таким образом, из четырех построенных степенных моделей работоспособны все, кроме первой. Так как исходными при построении модели являются временные ряды годовых данных в реальных ценах, то это не позволяет учесть влияние инфляционных процессов и изменения реальных доходов. В этой связи предпочтение следует отдать двум последним моделям, которые экономически более содержательны и обладают достаточно хорошими статистическими характеристиками.

Заслуживает внимания также модель, которую назовем модифицированной функцией спроса. В этой модели в качестве аргумента выступает переменная $\frac{x_2}{x_1}$ – стоимость 1 фунта цыплят, приходящаяся на единицу среднедушевого дохода. Этот удельный показатель более точно характеризует цену и используется при межстрановых и межрегиональных сопоставлениях цен.

Модифицированная функция спроса имеет вид

$$\tilde{y} = 9{,}32 \cdot \left(\frac{x_2}{x_1}\right)^{\underset{(-21{,}4)}{-0{,}488}} \tag{7.5}$$

$$\hat{s}^2 = 1{,}6;\ DW = 2{,}2;\ \bar{\delta} = 2{,}3\%.$$

Статистические характеристики модели свидетельствуют об ее адекватности. Из модели следует, что при увеличении объясняющей переменной на 1% потребление цыплят снизится на 0,488%. Этот вывод согласуется с экономической сущностью явления.

Модель (7.5) можно представить также в виде

$$\tilde{y} = 9{,}32 \cdot (x_1)^{0{,}488} \cdot (x_2)^{-0{,}488}. \tag{7.6}$$

В таком виде уравнение сопоставимо с моделью (7.3) даже по знакам при коэффициентах регрессии. Однако показатели адекватности ($\hat{s}^2, \bar{\delta}$) у модели (7.6) хуже, чем у (7.3). Последнее можно объяснить тем, что при построении модели (7.6) на ее параметры было наложено ограничение (дополнительные условие $b_1 = -b_2$), что и привело к увеличению остаточной дисперсии $\hat{s}^2$.

7.3. Методы нелинейной оптимизации

В некоторых случаях приведенные в табл. 7.1 нелинейные модели нельзя свести к линейной форме. В такой ситуации оценки наименьших квадратов параметров β приходится получать с помощью итерационных

вычислительных процедур, осуществляющих последовательное приближение к минимуму суммы квадратов

$$Q = \sum_{i=1}^{n} (y_i - f(x, \beta))^2.$$

Выполнить это приближение можно стандартными методами нелинейной оптимизации — методами Ньютона—Рафсона, Нелдера—Мидда и другими, включаемыми в большинство пакетов статистических программ [3, 4, 11, 22, 26].

Успешность данных процедур зависит от вида модели и особенностей применения итеративного подхода.

Так, в рассмотренной выше степенной модели $y = \beta_0 x_1^{\beta_1} \varepsilon$ предполагалось, что случайная составляющая ε мультипликативно связана с объясняющей переменной x. Если же модель представлена в виде $y = \beta_0 x_1^{\beta_1} + \varepsilon$, то она становится внутренне нелинейной, так как ее нельзя преобразовать к линейному виду.

7.4. Выбор формы модели. Подбор линеаризующего преобразования (подход Бокса—Кокса)

Многообразие и сложность экономических процессов предопределяет многообразие моделей, используемых для эконометрического анализа. С другой стороны, это существенно усложняет процесс нахождения максимально адекватной формулы зависимости. Для случая парной регрессии подбор модели обычно осуществляется по виду расположения наблюдаемых точек на корреляционном поле. Однако нередки ситуации, когда расположение точек приблизительно соответствует нескольким функциям и необходимо из них выбрать наилучшую. Например, криволинейные зависимости могут аппроксимироваться полиномиальной, показательной, степенной, логарифмической функциями. Еще более неоднозначна ситуация для множественной регрессии, так как наглядное представление статистических данных в этом случае невозможно.

Можно, конечно, действовать методом проб и ошибок: последовательно построить по имеющимся у нас исходным статистическим данным каждую из альтернативного набора линеаризуемых моделей, а затем выбрать из них наилучшую в смысле какого-то «критерия качества» (например, по максимальному значению подправленной на несмещенность оценки коэффициента детерминации $\hat{R}^2$).

Чтобы учитывать при селекции простоту модели, делают поправку на количество регрессоров k. Это дает коэффициент детерминации, скорректированный на количество степеней свободы:

$$R_0^2 = 1 - \frac{(1 - R^2)n}{n - k}.$$

Оценки метода наименьших квадратов в случае линейной зависимости являются одновременно и оценками метода максимального правдоподобия. Поэтому предлагается сравнивать модели на основе максимума логарифмической функции максимального правдоподобия ($\ln \hat{L}$). Если учесть при этом количество наблюдений (n) и ввести «штраф» за большое количество регрессоров (k), то получится информационный критерий Акаике (*Akaike information criterion*):

$$AIC = -\frac{2}{n}(\ln \hat{L} - k).$$

Чем меньше AIC, тем лучшей считается модель.

Английские статистики Г. Бокс и Д. Кокс предложили более формализованную процедуру подбора линеаризующего преобразования. Их метод основан на предположении, что искомое преобразование принадлежит определенному *однопараметрическому семейству* преобразований вида

$$\tilde{y}_i(\lambda) = \frac{y_i^{\lambda} - 1}{\lambda}, \quad \tilde{x}_i(\lambda) = \frac{(x_{ij})^{\lambda} - 1}{\lambda}, \quad i = 1, 2, ..., n. \tag{7.7}$$

Точнее их гипотезу можно сформулировать следующим образом: *существует такое вещественное (положительное или отрицательное) число* λ^*, *что один из двух нижеследующих вариантов представления искомой регрессионной зависимости между наблюдаемыми переменными* $X = (x_1, x_2, ..., x_k)^T$ *будет удовлетворять всем требованиям нормальной классической линейной модели множественной регрессии*:

обычная линейная модель

$$\tilde{y}_i(\lambda^*) = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{7.8'}$$

или линейная модель с использованием преобразованных переменных

$$\tilde{y}_i(\lambda^*) = \beta_0 + \beta_1 x_{i1}(\lambda^*) + ... + \beta_k x_{ik}(\lambda^*) + \varepsilon_i. \tag{7.8}$$

Замечание 1. Преобразования вида (7.7) применяются обычно к переменным, принимающим только *положительные* значения. Поэтому если это не так, то вначале подбирают «сдвиговые» константы $c_1, c_2, ..., c_k$, которые обеспечивают положительность значений $y_{i1} + c_0$ и $y_{ij} + c_j$ ($j = 1, 2, ..., k$), а затем к сдвинутым значениям переменных применяют данное преобразование, т. е.

$$y'_i(\lambda) = \frac{(y_i + c_0)^{\lambda} - 1}{\lambda}, \quad \tilde{x}_i(\lambda) = \frac{(x_{ij} + c_j)^{\lambda} - 1}{\lambda}, \quad i = 1, 2, ..., n. \tag{7.7'}$$

Замечание 2. Семейство степенных преобразований вида (7.7) или (7.7') весьма широко и гибко. При $\lambda = 1$ модели (7.8) и (7.8') являются линейными относительно y_i и $x_{i1}, x_{i2}, ..., x_{ik}$. При $\lambda = 0$ мы имеем степенную зависимость между y и x, поскольку $y'_i(0) = \lim(y_i^{\lambda} - 1)/\lambda = \ln y_i$ и $x'_{ij}(0) = \lim((x_j)^{\lambda} - 1)/\lambda = \ln x_{ij}$.

При других значениях λ уравнения (7.8) и (7.8′) будут связывать между собой какие-то степени исходных переменных.

Оценка неизвестного значения параметра λ, если исходить из справедливости сформулированной выше гипотезы, сводится к оценке параметра λ в формулах (7.7) или (7.7′) по имеющимся в распоряжении исходным статистическим данным. Эта проблема решается с помощью метода максимального правдоподобия.

Правильный выбор вида экономической модели является отправной точкой для ее качественного анализа. Безусловно, на практике неизвестно, какая модель является верной, и зачастую подбирают такую модель, которая наиболее точно соответствует реальным данным. При этом необходимо учитывать, что идеальной модели не существует. Поэтому, чтобы выбрать качественную модель, необходимо ответить на ряд вопросов, возникающих при ее анализе:

1. Каковы признаки «хорошей» модели?
2. Какие ошибки спецификации встречаются и каковы последствия данных ошибок?
3. Как обнаружить ошибку спецификации?
4. Каким образом можно исправить ошибку спецификации и перейти к лучшей модели?

В ряде случаев достаточно очевидно, какая модель лучше. В других случаях для принятия обоснованного решения приходится проводить достаточно кропотливый сравнительный анализ. Для этого необходимо выбрать критерии, которые позволят сделать обоснованный вывод. Обычно для построения «хорошей» работоспособной модели и сравнения ее с другими возможными моделями необходимо учитывать следующие свойства (критерии).

Простота. Модель должна быть максимально простой. Данное свойство определяется тем фактом, что модель не отражает действительность идеально, а является ее упрощением. Поэтому из двух моделей, приблизительно одинаково отражающих реальность, предпочтение отдается модели, содержащей меньшее число объясняющих переменных.

Единственность. Для любого набора статистических данных определяемые коэффициенты должны вычисляться однозначно.

Максимальное соответствие. Уравнение тем лучше, чем большую часть разброса зависимой переменной оно может объяснить. Поэтому стремятся построить уравнение с максимально возможным скорректированным коэффициентом детерминации — R^2.

ным. Другими словами, модель обязательно должна опираться на теоретический фундамент, так как в противном случае результат использования регрессионного уравнения может быть весьма плачевным.

Прогнозные качества. Модель может быть признана качественной, если полученные на ее основе расчетные значения результативного показателя $\hat{y}_i$ согласуются с наблюдаемыми значениями y_i.

Другим критерием прогнозных качеств оцененной модели регрессии может служить следующее отношение

$$V = \frac{\hat{s}}{\bar{y}} 100\%,$$

где $\hat{s} = \sqrt{\frac{\sum_{i=1}^{i=n} e_i^2}{n-k-1}}$ — стандартная ошибка регрессии; $e_i = y_i - \hat{y}_i$, $\bar{y}$ — среднее значение зависимой переменной уравнения регрессии.

Если величина V мала (а она определяет относительную ошибку прогноза в процентах) и отсутствует автокорреляция остатков (проверяется по критерию DW Дарбина—Уотсона), то прогнозные качества модели высоки. Отсутствие автокорреляции проверяется только при работе с временной выборкой.

Если уравнение регрессии используется для прогнозирования, то величина V обычно рассчитывается не для того периода, на котором оценивалось уравнение, а для некоторого следующего за ним временного интервала, для которого известны значения зависимой и объясняющих переменных. Тем самым на практике проверяются прогнозные качества модели. В случае положительного решения, если можно спрогнозировать значения объясняющих переменных на некоторый последующий период, построенная модель обоснованно может быть использована для прогноза значений объясняемой переменной y.

Поскольку не существует какого-либо единого правила построения регрессионных моделей, анализ перечисленных свойств позволяет строить более качественные эконометрические модели.

7.5. Виды ошибок спецификации

Одним из базовых предположений построения качественной модели является правильная (хорошая) спецификация уравнения регрессии. Правильная спецификация уравнения регрессии означает, что оно в целом верно отражает соотношение между экономическими показателями, участвующими в модели. Это является необходимой предпосылкой дальнейшего качественного оценивания.

Неправильный выбор функциональной формы или набора объясняющих переменных называется *ошибками спецификации.*

Основные типы ошибок спецификации

1. Отбрасывание значимой переменной

Суть данной ошибки и ее последствия наглядно иллюстрируются следующим примером. Пусть теоретическая модель, отражающая рассматриваемую экономическую зависимость, имеет вид

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \tag{7.9}$$

Данной модели соответствует следующее эмпирическое уравнение регрессии

$$y = b_0 + b_1 x_1 + b_2 x_2. \tag{7.10}$$

Исследователь по каким-то причинам (недостаток информации, поверхностное знание о предмете исследования и т. п.) считает, что на переменную y реально воздействует лишь переменная x_1. Он ограничивается рассмотрением модели

$$y = \gamma_0 + \gamma_1 x_1 + \nu. \tag{7.11}$$

При этом он не рассматривает в качестве объясняющей переменную x_2, совершая ошибку отбрасывания существенной переменной.

Пусть тогда эмпирическое уравнение регрессии, соответствующее теоретической модели (7.9), имеет вид

$$y = \hat{\gamma}_0 + \hat{\gamma}_1 x_1. \tag{7.12}$$

Последствия данной ошибки достаточно серьезны. Оценки, полученные с помощью МНК по уравнению (7.12), являются смещенными ($M(\hat{\gamma}_0) \neq \beta_0$, $M(\hat{\gamma}_1) \neq \beta_1$) и несостоятельными даже при бесконечно большом числе испытаний. Следовательно, возможные интервальные оценки и результаты проверки соответствующих гипотез будут ненадежными.

Отметим, что ошибка данного рода существенно отражается и на коэффициенте детерминации R^2. В нашей ситуации при использовании уравнения (7.12) значение коэффициента детерминации будет завышать роль переменной x_1 в объяснении переменной y.

2. Добавление незначимой переменной

В некоторых случаях в уравнения регрессии включают слишком много объясняющих переменных, причем не всегда обоснованно. Например, пусть теоретическая модель имеет следующий вид

$$y = \beta_0 + \beta_1 x_1 + \varepsilon. \tag{7.13}$$

Пусть исследователь подменяет ее более сложной моделью

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \varepsilon. \tag{7.14}$$

добавляя при этом не оказывающую реального воздействия на y объясняющую переменную x_2. В этом случае совершается ошибка добавления несущественной переменной.

Последствия данной ошибки будут не столь серьезными, как в предыдущем случае. Оценки $\hat{\gamma}_0$, $\hat{\gamma}_1$ коэффициентов, найденные для модели

(7.14), остаются, как правило, несмещенными ($M(\hat{\gamma}_0) = \beta_0$, $M(\hat{\gamma}_1) = \beta_1$) и состоятельными. Однако их точность уменьшится, увеличивая при этом стандартные ошибки, т. е. оценки становятся неэффективными, что отражается на их устойчивости. Данный вывод логически вытекает из формул расчета дисперсий оценок коэффициентов регрессии для этих уравнений

$$s_{b_1}^2 = \frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \text{ где } b_1 \text{ — оценка } \beta_1 \text{ в (7.13)},$$

$$s_{\hat{\gamma}_1}^2 = \frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2(1 - r_{12}^2)}, \text{ где } \hat{\gamma}_1 \text{ — оценка } \gamma_1 \text{ в (7.14)}.$$

Здесь r_{12} — коэффициент корреляции между объясняющими переменными x_1 и x_2. Следовательно, $s_{b_1}^2 \le s_{\hat{\gamma}_1}^2$, причем знак равенства возможен лишь при $r_{12} = 0$.

Увеличение дисперсии оценок может привести к ошибочным результатам проверки гипотез относительно значений коэффициентов регрессии, расширению интервальных оценок.

3. Выбор неправильной функциональной формы

Рассмотрим следующий пример. Пусть правильная регрессионная модель имеет вид

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

Любое эмпирическое уравнение регрессии с теми же переменными, но имеющее другой функциональный вид, приводит к искажению истинной зависимости. Например, в следующих уравнениях

$$\ln y = a_0 + a_1 x_1 + a_2 x_2 + e,$$
$$y = c_0 + c_1 \ln x_1 + c_2 \ln x_2 + u$$

совершена ошибка выбора неправильной функциональной формы уравнения регрессии. Последствия данной ошибки будут весьма серьезными. Обычно такая ошибка приводит либо к получению смещенных оценок, либо к ухудшению статистических свойств оценок коэффициентов регрессии и других показателей качества уравнения. В первую очередь это вызвано нарушением условий Гаусса—Маркова для отклонений. Прогнозные качества модели в этом случае очень низки.

Обнаружение и корректировка ошибок спецификации

При построении уравнений регрессии, особенно на начальных этапах, ошибки спецификации весьма нередки. Они допускаются обычно из-за поверхностных знаний об исследуемых экономических процессах либо из-за недостаточно глубоко проработанной теории, или из-за погрешностей сбора и обработки статистических данных при построении

эмпирического уравнения регрессии. Важно уметь обнаружить и исправить эти ошибки. Сложность процедуры определяется типом ошибки и нашими знаниями об исследуемом объекте.

Если в уравнении регрессии имеется одна несущественная переменная, то она обнаружит себя по низкой t-статистике. В дальнейшем эту переменную исключают из рассмотрения.

Если в уравнении несколько статистически незначимых объясняющих переменных, то следует построить другое уравнение регрессии без этих незначимых переменных. Затем с помощью F-статистики сравниваются коэффициенты детерминации для первоначального и дополнительного уравнений регрессий

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

Здесь n — число наблюдений, k — число объясняющих переменных в первоначальном уравнении, m — число отбрасываемых из первоначального уравнения объясняющих переменных.

При наличии нескольких несущественных переменных, возможно, имеет место мультиколлинеарность. Рекомендуемые выходы из этой ситуации подробно рассмотрены ранее. Однако осуществление указанных проверок имеет смысл лишь при правильном подборе вида (функциональной формы) уравнения регрессии, что можно осуществить, если согласовывать его с теорией. Например, при построении кривой Филлипса, указывающей, что зависимость между заработной платой y и безработицей x является обратной, возможны следующие модели:

$$y = \alpha + \beta x + \varepsilon, \beta < 0;$$

$$\ln y = \alpha + \beta \ln x + \varepsilon, \beta < 0;$$

$$y = \alpha + \beta \frac{1}{x + \gamma} + \varepsilon, \beta < 0;$$

$$y = \alpha + \alpha^{\beta x} + \varepsilon, \beta < 0.$$

Отметим, что выбор модели не всегда осуществляется однозначно, и в дальнейшем требуется сравнивать модель как с теоретическими, так и с эмпирическими данными, совершенствовать ее. При определении качества модели обычно анализируются следующие параметры:

а) скорректированный коэффициент детерминации $\overline{R}^2$;

б) t-статистики;

в) статистика Дарбина—Уотсона (DW);

г) согласованность знаков коэффициентов с экономической теорией;

д) прогнозные качества (ошибки) модели.

Если все эти показатели удовлетворительны, то данная модель может быть предложена для описания исследуемого реального процесса. Если же какая-либо из описанных выше характеристик не является удовлетворительной, то есть основания сомневаться в качестве данной модели (не-

правильно выбрана функциональная форма уравнения; не учтена важная объясняющая переменная; имеется объясняющая переменная, не оказывающая значимого влияния на зависимую переменную).

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

7.6. Преобразование случайного отклонения

Существенную роль для получения качественных оценок играет выполнимость определенных предпосылок МНК для случайных отклонений. Наиболее важные из них требуют, чтобы отклонения ε_i являлись нормально распределенными случайными величинами с нулевым математическим ожиданием и постоянной дисперсией σ^2, а также не коррелировали друг с другом ($\varepsilon_i \in N(0, \sigma^2)$, $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$, $i, j = 1, 2, \ldots, n$). При невыполнимости указанных предпосылок оценки, полученные по МНК, не будут наилучшими в классе линейных оценок.

В моделях с аддитивным случайным членом выполнимость предпосылок МНК имеет место, и проблем с оцениванием не возникает.

Для описания возможных проблем со случайным отклонением воспользуемся степенной моделью. При этом случайный член ε может входить в уравнение по-разному. Рассмотрим три возможных случая:

$$y = \beta_0 x^{\beta_1} e^{\varepsilon}, \tag{7.15}$$

$$y = \beta_0 x^{\beta_1} \varepsilon, \tag{7.16}$$

$$y = \beta_0 x^{\beta_1} + \varepsilon. \tag{7.17}$$

Данные модели являются нелинейными относительно параметра β_1. Прологарифмировав каждое из этих соотношений, соответственно получим:

$$\ln y = \beta'_0 + \beta_1 \ln x + \varepsilon, \tag{7.18}$$

$$\ln y = \beta'_0 + \beta_1 \ln x + \ln\varepsilon, \tag{7.19}$$

$$\ln y = \ln(\beta_0 x^{\beta_1} + \varepsilon). \tag{7.20}$$

Здесь $\beta'_0 = \ln \beta_0$.

Использование (7.18) для оценки параметров в (7.15) не вызывает осложнений, связанных со случайным отклонением.

Преобразование (7.16) в (7.19) приводит к преобразованию случайных отклонений ε_i в $\ln\varepsilon_i$. Использование МНК в (7.19) для нахождения оценок параметров требуется, чтобы отклонения $\nu_i = \ln\varepsilon_i$ удовлетворяли предпосылкам МНК: $\nu_i \sim N(0, \sigma^2)$. Но это возможно только в случае логарифмически нормального распределения случайной величины ε_i с математическим ожиданием $M(\varepsilon_i) = e^{\sigma^2/2}$ и дисперсией $D(\varepsilon_i) = e^{\sigma^2}(e^{\sigma^2} - 1)$.

Логарифмирование соотношения (7.17) не приводит к линеаризации соотношения относительно параметров. В этом случае для нахождения оценок необходимо использовать определенные итерационные процедуры оценки нелинейных регрессий.

Таким образом, при использовании преобразований с целью нахождения оценок необходимо особое внимание уделять рассмотрению свойств случайных отклонений, чтобы полученные в результате оценки имели высокую статистическую значимость.

Графическое исследование остаточного члена модели

Графическое представление поведения остаточного члена e_i $(i = 1, 2, ..., n)$ позволяет проанализировать наличие автокорреляции и гетероскедастичности (непостоянства дисперсий отклонений). Кроме того, с помощью графического представления отклонений e_i, может быть также обнаружена неправильная спецификация уравнения. Для этого строится график зависимости величин отклонений e_i от номера наблюдения i.

Если зависимость имеет регулярный (неслучайный) характер, то это означает, что исследуемое уравнение регрессии неверно специфицировано.

Существует и ряд других тестов обнаружения ошибок спецификации, среди которых можно выделить [3, 4, 11, 21, 22]:

1. Тест Рамсея RESET (*Regression specification error test*).
2. Тест (критерий) максимального правдоподобия (*The Likelihood Ratio test*).
3. Тест Валда (*The Wald test*).
4. Тест множителя Лагранжа (*The Lagrange multiplier test*).
5. Тест Хаусмана (*The Hausman test*).
6. Бокса—Кокса преобразование (*Box—Cox transformation*).

Суть указанных тестов состоит либо в осуществлении преобразований случайных отклонений, либо в масштабировании зависимой переменной, с тем чтобы можно было сравнить начальное и преобразованное уравнения регрессии на основе известного критерия.

7.7. Схема анализа нелинейных эконометрических зависимостей

Стандартная схема анализа нелинейных зависимостей должна включать в себя следующие этапы.

Подбор начальной модели. Он осуществляется на основе экономической теории, предыдущих знаний об объекте исследования, опыта исследователя и его интуиции.

Оценка параметров модели на основе имеющихся статистических данных.

Осуществление тестов проверки качества модели (обычно используются t-статистики для коэффициентов регрессии, F-статистика для коэф-

фициента детерминации, статистика Дарбина—Уотсона для анализа отклонений и ряд других тестов).

При наличии хотя бы одного неудовлетворительного ответа по какому-либо тесту модель совершенствуется с целью устранения выявленного недостатка.

При положительных ответах по всем проведенным тестам модель считается адекватной. Она используется для анализа и прогнозирования объясняемой переменной.

Однако необходимо отметить, что даже качественная модель является подгонкой спецификации модели под имеющийся набор данных, поэтому вполне реальна картина, когда исследователи, обладающие одинаковыми наборами данных, строят разные модели для объяснения одной и той же переменной. Проблематичным является и использование модели для прогнозирования значений объясняемой переменной. Иногда хорошие с точки зрения диагностических тестов модели обладают весьма низкими прогнозными качествами.

Одно из главных направлений эконометрического анализа — постоянное совершенствование моделей. Здесь следует отметить отсутствие единственного подхода, определяющего заранее возможные пути совершенствования. Исследователь должен помнить, что совершенной модели не существует. В силу постоянно изменяющихся условий протекания экономических процессов не может быть и постоянно качественных моделей. Новые условия требуют пересмотра даже весьма устойчивых моделей.

Однако отметим, что при всех недостатках моделей принятие на их основе решений приводит в целом к гораздо более точным результатам, чем при принятии решений лишь на основе интуиции и экономической теории.

7.8. Задачи и тесты

1. Данные о прибыли предприятия y (млн долл.) и расходах на рекламу x по за 9 лет представлены в таблице.

y	5	7	12	16	23	21	19	18	16
x	0,8	1,1	1,8	2,5	4,1	5,5	7,3	8,1	8,9

Требуется:

1) построить корреляционное поле и выдвинуть гипотезу о форме зависимости между рассматриваемыми показателями;

2) оценить по МНК коэффициенты линейного уравнения регрессии $\tilde{y} = \beta_0 + \beta_1 x$ и сделать вывод о качестве построенного уравнения;

3) оценить по МНК коэффициенты параболического уравнения регрессии $\tilde{y} = \beta_0 + \beta_1 x + \beta_2 x^2$ и сделать вывод о его качестве;

4) обосновать выбор лучшей модели.

2. По данным за 15 лет построены два уравнения регрессии:

$$\tilde{y} = 3{,}45 - 0{,}55x;\ R^2 = 0{,}68$$

$$t = (20{,}5)\ (-4{,}3),$$

$$\ln \tilde{y} = 0{,}85 - 0{,}25x,\ R^2 = 0{,}78$$

$$t = (44{,}9)\ (-5{,}3),$$

где y — ежедневное среднедушевое потребление кофе (в чашках по 100 г); x — среднегодовая цена кофе (в руб./кг).

Требуется:

1) проинтерпретировать коэффициенты каждой из моделей;

2) обосновать выбор лучшей модели;

3) ответить на вопрос, можно ли о качестве модели судить по коэффициенту детерминации.

3. По статистическим данным некоторой страны за 20 лет построена модель макроэкономической производственной функции:

$$\ln y = -3{,}52 + 1{,}53\ln K + 0{,}47\ln L,\ R^2 = 0{,}875$$

$$t = (2{,}76)\ (5{,}32),$$

где y — реальный ВНП (млн долл.), K — объем затрат капитала, (млн долл.), L — объем затрат труда (человеко-дни).

Требуется:

1) оценить качество построенной модели;

2) проинтерпретировать коэффициенты регрессии и оцените их статистическую значимость.

4. В таблице представлены данные за 9 лет, характеризующие зависимость прибыли предприятия y от расходов на рекламу x:

y	5	7	13	15	20	25	22	20	17
x	0,8	1,0	1,8	2,4	4,1	5,6	7,3	8,2	8,8

Требуется:

1) построить линейную регрессионную модель и оценить ее коэффициенты;

2) построить параболическую регрессионную модель и оценить ее коэффициенты;

3) построить обратную регрессионную модель и оценить ее коэффициенты;

4) обосновать выбор лучшей модели.

5. Модель вида $y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon_i$ носит название:

– *полиномиальная;

– гиперболическая (обратная);

– обращенная полиномиальная функция;

– логлинейная;
– обобщенная логарифмическая;
– полулогарифмическая;
– степенная;
– экспоненциальная.

6. Модель вида $y_i = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon_i$ носит название:

– полиномиальная;
– *гиперболическая (обратная);
– обращенная полиномиальная функция;
– логлинейная;
– обобщенная логарифмическая;
– полулогарифмическая;
– степенная.

7. Модель вида $\ln y = \beta_0 + \beta_1 \ln x + \varepsilon_i$ носит название:
– полиномиальная;
– гиперболическая (обратная);
– обращенная полиномиальная функция;
– *логлинейная;
– обобщенная логарифмическая;
– полулогарифмическая;
– степенная.

8. Модель вида $y = \beta_0 a^{\beta_1 x} \varepsilon$ носит название:
– полиномиальная;
– гиперболическая (обратная);
– экспоненциальная;
– степенная;
– функция Гомперца;
– логистическая;
– *показательная.

9. Модель вида $y = \dfrac{1}{\beta_0 + \beta_1 e^{-x} + \varepsilon}$ носит название:

– полиномиальная;
– гиперболическая (обратная);
– экспоненциальная;
– степенная;
– *функция Гомперца;
– логистическая;
– показательная.

10. Модель вида $y = 1/(\beta_0 + \beta_1 e^{-x})$ носит название:

– полиномиальная;
– гиперболическая (обратная);
– *обращенная экспоненциальная;
– степенная;
– функция Гомперца;
– логистическая;
– показательная.

11. Степенная модель $y_i = \beta_0 x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_k^{\beta_r} \varepsilon_i$ относится к моделям:

– нелинейным относительно включенных в анализ объясняющих переменных, но линейным по оцениваемым параметрам (β_j);
– *нелинейным относительно включенных в анализ объясняющих переменных и нелинейным по оцениваемым параметрам;
– линейным.

12. Кубическая функция $y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 X^3 + \varepsilon_i$ может быть приведена к линейному виду с помощью следующих преобразований:

– *замена переменных;
– логарифмирование обеих частей уравнения;
– замена переменных и логарифмирование обеих частей уравнения;
– исключение лишних переменных.

13. Модель вида $y = \dfrac{1}{\beta_0 + \beta_1 x + \varepsilon}, (-\dfrac{\beta_0}{\beta_1} < x < \infty)$ может быть приведена к линейному виду с помощью следующих преобразований:

– *замена переменных;
– логарифмирование обеих частей уравнения;
– замена переменных и логарифмирование обеих частей уравнения;
– исключение лишних переменных.

14. Функция Кобба—Дугласа может быть приведена к линейному виду с помощью следующих преобразований:

– замена переменных;
– *логарифмирование обеих частей уравнения;
– замена переменных и логарифмирование обеих частей уравнения;
– исключение лишних переменных.

15. Из модели

$$\hat{y} = 9{,}32 \cdot \left(\frac{x_2}{x_1}\right)^{-0{,}488}$$

следует, что при увеличении объясняющей переменной на 1% $\hat{y}$ в среднем:

– увеличится на 0,488%;
– *снизится на 0,488%;

– увеличится на 0,488 единицы своего измерения;
– снизится на 0,488 единицы своего измерения.

16. Построенное уравнение регрессии

$$\hat{y} = -1{,}2x_1^{6{,}57}x_2^{-0{,}22}x_3^{7{,}8}$$

показывает, что:
– рост переменной x_1 на единицу своего измерения приводит к росту среднего значения y на 6,57 единицы своего измерения;
– рост переменной x_1 на единицу своего измерения приводит к уменьшению среднего значения y на 6,57 единицы своего измерения;
– *рост переменной x_1 на % приводит к росту среднего значения y на 6,57%;
– рост переменной x_1 на % приводит к уменьшению среднего значения y на 6,57%.

17. Согласно положениям регрессионного анализа нелинейная функция может быть приведена к линейной форме (т. е. линеаризирована) только в том случае, если ошибка обоих уравнений регрессии остается:
– *аддитивной;
– мультипликативной;
– смешанной.

18. При анализе эластичности спроса по цене целесообразно использовать следующую модель:
– линейную;
– полиномиальную;
– логарифмическую;
– *степенную;
– экспоненциальную.

19. Для оценки неизвестных параметров в нелинейных моделях используют следующие методы:
– *замена переменных;
– *логарифмирование обеих частей уравнения;
– *замена переменных и логарифмирование обеих частей уравнения;
– исключение лишних переменных.

7.9. Вопросы для самопроверки

1. Какие классы нелинейных моделей вы знаете?

2. Приведите примеры регрессионных моделей, нелинейных по оцениваемым параметрам.

3. Приведите примеры регрессионных моделей, нелинейных относительно включенных в анализ объясняющих переменных, но линейных по оцениваемым параметрам.

4. Какой нелинейной функцией может быть заменена парабола второй степени, если не наблюдается смена направленности связи признаков?

5. Чем отличается применение МНК к моделям, нелинейным относительно включаемых переменных от применения к моделям, нелинейным по оцениваемым параметрам?

6. Как определяются коэффициенты эластичности для степенных моделей?

7. Назовите основные виды ошибок спецификации.

8. Назовите методы линеаризации нелинейных моделей.

9. В каких случаях используется гармонический анализ? Приведите примеры.

10. В каких случаях следует прибегать к итерационным вычислительным процедурам? Приведите примеры.