9. Обобщенные схемы регрессионного анализа

9.1. Основные теоретические сведения

Гребневая регрессия

В условиях высокой коррелированности экзогенных переменных оценки коэффициентов регрессии, вычисленные с помощью МНК проявляют неустойчивость. Это обусловлено тем, что дисперсия оценок увеличивается при увеличении взаимной корреляции входящих переменных. При этом коэффициент множественной корреляции экзогенной переменной x_j с остальными стремится к единице, что приводит к вырожденности матрицы $X^T X$. В этих условиях лучше использовать смещенные оценки, имеющие меньшую дисперсию, чем несмещенные. Метод, использующий минимум среднего квадратического отклонения в качестве критерия выбора оптимальных оценок коэффициентов регрессии, называется методом гребневой или ридж-регрессии.

Для вычисления смещенных оценок в методе гребневой регрессии вводится параметр k, $0 \le k \le 1$ и ридж-оценки параметров вычисляются по формулам:

$$b = \quad X'^T X' + E \cdot k \quad^{-1} X'^T Y'.$$

Устойчивость оценок параметров характеризуется суммой их дисперсий

$$L \quad b \quad = \sum_{i=1}^{m} s_{b_i}^2 .$$

Найдем все собственные числа λ_i и собственные векторы U_i матрицы $X^T X = U D_\lambda U^T$, где D_λ – диагональная матрица собственных чисел, а U – ортогональная матрица собственных векторов.

Вычисляем $U^T b = g$.

Тогда сумма дисперсий параметров модели может быть представлена в форме:

$$L \quad b \quad = \sum_{i=1}^{m} \frac{\lambda_i s_e^2 \quad k \quad + k^2 g_i}{\lambda_i + k \quad^2} .$$

При мультиколлинеарности (вырождении матрицы $X^T X$) одно или несколько собственных чисел λ_i близки к нулю и при $k = 0$ получается очень большое (если не бесконечное) значение $L\ \ b$. Введение даже очень малого k сразу исправляет положение. Выбирают такое наименьшее значение параметра k, при котором поведение всех коэффициентов регрессии уже стабилизировалось, но еще не начался интенсивный рост остаточной дисперсии.

Заметим, что метод гребневой регрессии эффективнее обычного МНК лишь при наличии значительных корреляций экзогенных переменных.

Модели с dummy-переменными

Dummy-переменные предназначены для учета в модели качественных переменных, измеренных на номинальных шкалах. Например, пол: женский, мужской; состояние экономики: до кризиса и после; сезон: зимний, весенний, летний, осенний (это категории качественной переменной). Качественные признаки могут существенно влиять на структуру линейных связей между переменными и приводить к скачкообразному изменению параметров регрессионной модели. Чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены те или иные цифровые метки, то есть качественные переменные преобразованы в количественные. Такого вида сконструированные переменные в эконометрике принято называть *фиктивными* или *dummy–переменными*.

Качественный фактор принимает только два состояния, которым соответствуют значения 1 и 0. Если же число градаций качественного признака-фактора превышает два, то в модель вводится несколько фиктивных переменных, число которых должно быть на единицу меньше числа качественных градаций. Только при соблюдении этого положения матрица исходных фиктивных переменных будет линейно независима и возможна оценка параметров модели. В сумме все *dummy-переменные* образуют $X_0 \equiv 1$.

Новейшие (Advanced) методы регрессионного анализа

В новейших методах регрессионного анализа результативный признак может быть дискретным и даже качественным, выраженным категориями или номинациями. В этих методах изучается вероятность этих категорий в зависимости от значений объясняющих переменных.

В общем регрессионном анализе на основе внешнего поведения результативного признака, заданного некоторым количеством случайных наблюдений, определяют форму связи и далее методом наименьших квадратов оценивают ее параметры или коэффициенты регрессии. В линейном регрессионном анализе форма связи принимается линейной относительно параметров модели, другими словами, возможны предварительные преобразования переменных. Так, для описания зависимости от одной объясняющей переменной в традиционном регрессионном анализе принимают линейную модель, имеющую вид:

$$\hat{y}(x) = b_0 + b_1 x,$$

где $\hat{y}(x) = y_p(x)$ – ожидаемое (среднее) значение y для каждого x; возможны предварительные функциональные преобразования начальных переменных $(x; y)$.

Коэффициенты регрессии a, b оценивают методом наименьших квадратов, после чего определяют несмещенную оценку остаточной $s_e^2 = const$ и дисперсии вычисленных значений для каждого x:

$$\sigma^2(\hat{y}) = \frac{s_e^2}{N}\left(1 + \frac{(x - \bar{x})^2}{s_x^2}\right),$$

где N – число наблюдений;

s_x^2 – дисперсия объясняющей переменной x.

Границы 95 %-го доверительного интервала на каждое вычисленное значение $\hat{y}(x)$ вычисляется по формулам:

$$y_d = \hat{y} - t_{0,05}(N-2) \cdot \sigma(\hat{y});$$

$$y_u = \hat{y} + t_{0,05}(N-2) \cdot \sigma(\hat{y}),$$

где y_d – нижняя граница (*down)*;

y_u – верхняя граница (*up*);

$t_{0,05}$ *(dfe)* – квантиль распределения Стьюдента;

$dfe = N - 2$ – остаточное число степеней свободы.

В обобщенной линейной модели считается, что результативный признак Y является случайной величиной с известным видом закона распределения, который описывается функцией плотности вероятности $f(y|\mu,\nu)$, где μ и ν – параметры функции распределения (вероятности – для дискретной или плотность вероятности – для непрерывной случайной величины). Считается, что один из этих параметров, а именно линейный предиктор μ, зависит от объясняющих переменных, другой параметр, если закон распределения двухпараметрический, от объясняющих переменных не зависит – $\nu = const$. Линейный предиктор является комбинацией объясняющих переменных и эта комбинация линейна относительно параметров:

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$$

(допускаются также предварительные функциональные преобразования начальных переменных). По экспериментальным данным $(x_i; y_i)$ методом наименьших квадратов оцениваются коэффициенты линейного предиктора β_j и константа ν. Предполагается, что функциональная зависимость между предиктором и ожидаемыми значениями результативной переменной $\mu = g(M(y)) = g(\hat{y})$. Функция g считается известной и называется функцией связи.

Традиционная модель регрессионного анализа является отдельным случаем обобщенной линейной модели при тождественной форме связи $\hat{y} = \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$ и нормальном законе распределения $Y = N(\mu, \sigma)$. Заметим, что даже в этом случае регрессионная модель и обобщенная линейная модель не эквивалентны, поскольку для них разные методы определения ширины доверительной полосы на вычисленные значения.

9.2. Примеры

Пример 9.1. В табл. 9.1 приведены данные для линейного трехфакторного уравнения регрессии:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \varepsilon_t \quad t = 1, \ldots, T \ .$$

Таблица 9.1

Исходные данные

x_{1t}	9,8	12,4	11,2	16,8	9,2
x_{2t}	18,6	26,0	21,0	32,5	23,0
x_{3t}	4,5	19,6	8,6	7,9	15,6
y_t	50	70	55	60	65

Необходимо:

1. Определить гребневые оценки параметров для гребневой константы, равной 0,5 и 0,8.

2. По обоим оцененным уравнениям построить точечный прогноз математического ожидания целевой переменной y_0 для значений экзогенных переменных $x_{10}, x_{20}, x_{30} = 10; 20; 8$.

Решение

1. Гребневые оценки определяются по формулам:

$$b = \left(X'^T X' + E \cdot k\right)^{-1} X'^T Y',$$

где элементами матрицы X' являются центрированные переменные $x'_{it} = x_{it} - \bar{x}_i$, а вектор Y' составляют $y'_t = y_t - \bar{y}$.

Для исходного трехфакторного уравнения матрица X имеет вид:

$$X = \begin{pmatrix} 9,8 & 18,6 & 4,5 \\ 12,4 & 26,0 & 19,6 \\ 11,2 & 21,0 & 8,6 \\ 16,8 & 32,5 & 7,9 \\ 9,2 & 23,0 & 15,6 \end{pmatrix}.$$

Для экзогенных переменных x_1, x_2 и x_3 рассчитаем средние значения: $\bar{x}_1 = 11,88$; $\bar{x}_2 = 24,22$; $\bar{x}_3 = 11,24$.

Тогда матрица X' принимает вид:

$$X' = \begin{pmatrix} 9{,}8 - 11{,}88 & 18{,}6 - 24{,}22 & 4{,}5 - 11{,}24 \\ 12{,}4 - 11{,}88 & 26{,}0 - 24{,}22 & 19{,}6 - 11{,}24 \\ 11{,}2 - 11{,}88 & 21{,}0 - 24{,}22 & 8{,}6 - 11{,}24 \\ 16{,}8 - 11{,}88 & 32{,}5 - 24{,}22 & 7{,}9 - 11{,}24 \\ 9{,}2 - 11{,}88 & 23{,}0 - 24{,}22 & 15{,}6 - 11{,}24 \end{pmatrix} = \begin{pmatrix} -2{,}08 & -5{,}62 & -6{,}74 \\ 0{,}52 & 1{,}78 & 8{,}36 \\ -0{,}68 & -3{,}22 & -2{,}64 \\ 4{,}92 & 8{,}28 & -3{,}34 \\ -2{,}68 & -1{,}22 & 4{,}36 \end{pmatrix}.$$

После транспонирования получаем матрицу:

$$X'^T = \begin{pmatrix} -2{,}08 & 0{,}52 & -0{,}68 & 4{,}92 & -2{,}68 \\ -5{,}62 & 1{,}78 & -3{,}22 & 8{,}28 & -1{,}22 \\ -6{,}74 & 8{,}36 & -2{,}64 & -3{,}34 & 4{,}36 \end{pmatrix}.$$

Найдем произведение матриц $X'^T X'$:

$$X'^T X' = \begin{pmatrix} -2{,}08 & 0{,}52 & -0{,}68 & 4{,}92 & -2{,}68 \\ -5{,}62 & 1{,}78 & -3{,}22 & 8{,}28 & -1{,}22 \\ -6{,}74 & 8{,}36 & -2{,}64 & -3{,}34 & 4{,}36 \end{pmatrix} \begin{pmatrix} -2{,}08 & -5{,}62 & -6{,}74 \\ 0{,}52 & 1{,}78 & 8{,}36 \\ -0{,}68 & -3{,}22 & -2{,}64 \\ 4{,}92 & 8{,}28 & -3{,}34 \\ -2{,}68 & -1{,}22 & 4{,}36 \end{pmatrix} =$$

$$= \begin{pmatrix} 36{,}448 & 58{,}812 & -7{,}956 \\ 58{,}812 & 115{,}168 & 28{,}286 \\ -7{,}956 & 28{,}286 & 152{,}452 \end{pmatrix}.$$

При $k = 0{,}5$ вычислим значение выражения $X'^T X' + E \cdot k$:

$$X'^T X' + 0{,}5 \cdot E = \begin{pmatrix} 36{,}948 & 58{,}812 & -7{,}956 \\ 58{,}812 & 115{,}668 & 28{,}286 \\ -7{,}956 & 28{,}286 & 152{,}952 \end{pmatrix}.$$

Тогда

$$X'^T X' + 0{,}5 \cdot E \quad ^{-1} = \begin{pmatrix} 0{,}2757 & -0{,}1505 & 0{,}0422 \\ -0{,}1505 & 0{,}0912 & -0{,}0247 \\ 0{,}0422 & -0{,}0247 & 0{,}0133 \end{pmatrix}.$$

Следовательно

$$\left(X'^T X' + E\cdot k\right)^{-1} X'^T = \begin{pmatrix} -0,0119 & 0,2280 & 0,1857 & -0,0305 & -0,3713 \\ -0,0331 & -0,1223 & -0,1261 & 0,0972 & 0,1843 \\ -0,0386 & 0,0891 & 0,0157 & -0,0414 & -0,0249 \end{pmatrix}$$

и окончательно получаем: $b = \begin{pmatrix} -0,3862 \\ 0,6599 \\ 1,0739 \end{pmatrix}$.

Теперь вычислим значение выражения $X'^T X' + E\cdot k$ при $k = 0,8$:

$$X'^T X' + 0,8\cdot E = \begin{pmatrix} 37,248 & 58,812 & -7,956 \\ 58,812 & 115,968 & 28,286 \\ -7,956 & 28,286 & 153,252 \end{pmatrix}.$$

Тогда

$$\left(X'^T X' + 0,5\cdot E\right)^{-1} = \begin{pmatrix} 0,2485 & -0,1353 & 0,0379 \\ -0,1353 & 0,0827 & -0,0223 \\ 0,0379 & -0,0223 & 0,0126 \end{pmatrix}.$$

Следовательно

$$\left(X'^T X' + E\cdot k\right)^{-1} X'^T = \begin{pmatrix} -0,0119 & 0,2050 & 0,1666 & -0,0238 & -0,3359 \\ -0,0330 & -0,1095 & -0,1154 & 0,0933 & 0,1645 \\ -0,0385 & 0,0854 & 0,0127 & -0,0402 & -0,0194 \end{pmatrix}$$

и $b = \begin{pmatrix} -0,3428 \\ 0,6351 \\ 1,0787 \end{pmatrix}$.

2. Построим точечный прогноз математического ожидания целевой переменной y_0 по обоим оцененным уравнениям для значений экзогенных переменных $\left(x_{10}, x_{20}, x_{30}\right) = \left(10; 20; 8\right)$.

$\hat{y}_1 = \bar{y} + b_{11} \cdot x'_{10} + b_{21} \cdot x'_{20} + b_{31} \cdot x'_{30} = \bar{y} + b_{11} \cdot (x_{10} - \bar{x}_1) + b_{21} \cdot (x_{20} - \bar{x}_2) +$
$+ b_{31} \cdot (x_{30} - \bar{x}_3) = 60 + (-0,3862) \cdot (10 - 11,88) + 0,6599 \cdot (20 - 24,22) +$
$+ 1,0739 \cdot (8 - 11,24) = 54,4615;$
$\hat{y}_2 = 60 + (-0,3428) \cdot (10 - 11,88) + 0,6351 \cdot (20 - 24,22) + 1,0787 \cdot$
$\cdot (8 - 11,24) = 54,46951.$

Пример 9.2. Рассмотрим применение фиктивных переменных для функции спроса. Предположим, что по группе лиц мужского и женского пола изучается линейная зависимость потребления кофе y от цены x. В общем виде для совокупности обследуемых уравнение регрессии имеет вид: $y = a + bx + \varepsilon$.

Аналогичные уравнения могут быть найдены отдельно для лиц мужского пола: $y_1 = a_1 + b_1 \cdot x_1 + \varepsilon_1$
и женского пола: $y_2 = a_2 + b_2 \cdot x_2 + \varepsilon_2$.

Различия в потреблении кофе проявятся в различии средних $\overline{y_1}$ и $\overline{y_2}$. Вместе с тем сила влияния x на y может быть одинаковой, то есть $b \approx b_1 \approx b_2$. В этом случае возможно построение общего уравнения регрессии с включением в него фактора "пол" в виде фиктивной переменной. Объединив уравнения y_1 и y_2 и введя фиктивные переменные, приходим к следующему выражению:

$$y = a_1 \cdot z_1 + a_2 \cdot z_2 + b \cdot x + \varepsilon,$$

где z_1 и z_2 – фиктивные переменные, принимающие значения:

$$z_1 = \begin{cases} 1 - \textit{мужской пол} \\ 0 - \textit{женский пол} \end{cases}; \qquad z_2 = \begin{cases} 0 - \textit{мужской пол} \\ 1 - \textit{женский пол} \end{cases}.$$

В общем уравнении регрессии зависимая переменная y рассматривается как функция не только цены x, но и пола (z_1, z_2). Переменная z рассматривается как переменная, принимающая всего два значения: 1 и 0. При этом когда $z_1 = 1$, то $z_2 = 0$ и, наоборот, при $z_1 = 0$ переменная $z_2 = 1$.

Для лиц мужского пола, когда $z_1 = 1$, то $z_2 = 0$, объединенное уравнение регрессии составит: $\hat{y} = a_1 + b \cdot x$, а для лиц женского пола, когда $z_1 = 0$ и $z_2 = 1$ $\hat{y} = a_2 + b \cdot x$. Иными словами, различия в потреблении для лиц мужского и женского пола вызваны различиями свободных членов уравнения регрессии: $a_1 \neq a_2$. Параметр b является общим для всей совокупности лиц как для мужчин, так и для женщин.

Следует иметь в виду, что при введении фиктивных переменных z_1 и z_2 в модель $y = a_1 \cdot z_1 + a_2 \cdot z_2 + b \cdot x + \varepsilon$ применение МНК для оценивания параметров a_1 и a_2 приведет к вырожденной матрице исходных данных, а следовательно, и к невозможности получения их оценок. Объясняется это тем, что при использовании МНК в данном уравнении появляется свободный член, то есть уравнение примет вид:

$$\hat{y} = a_1 \cdot z_1 + a_2 \cdot z_2 + b \cdot x + A.$$

Предполагая при параметре A независимую переменную, равную 1, имеем матрицу исходных данных:

$$\begin{pmatrix} 1 & 1 & 0 & x_1 \\ 1 & 1 & 0 & x_2 \\ 1 & 0 & 1 & x_3 \\ 1 & 1 & 0 & x_4 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & x_n \end{pmatrix}.$$

В рассматриваемой матрице существует линейная зависимость между первым, вторым и третьим столбцами: первый равен сумме второго и третьего столбцов. Поэтому матрица исходных факторов вырождена. Выходом из этой ситуации может явиться переход к уравнению:

$$y = A + A_1 \cdot z_1 + b \cdot x + \varepsilon \text{ или } y = A + A_2 \cdot z_2 + b \cdot x + \varepsilon,$$

то есть каждое уравнение включает только одну фиктивную переменную z_1 или z_2.

Предположим, что определено уравнение $y = A + A_1 \cdot z_1 + b \cdot x + \varepsilon$, где z_1 – принимает значения 1 для мужчин и 0 для женщин.

Теоретические значения уровня потребления кофе для мужчин будут получены из уравнения $\hat{y} = A + A_1 + b \cdot x$.

Для женщин соответствующие значения получим из уравнения $\hat{y} = A + b \cdot x$.

Сопоставив эти значения, видим, что различия в уровне потребления мужчин и женщин состоят в различии свободных членов данных уравнений: A – для женщин и $A + A_1$ – для мужчин.

Пример 9.3. Проанализируем зависимость цены двухкомнатной квартиры от ее полезной площади. При этом в модель могут быть введены фиктивные переменные, отражающие тип дома: "хрущевка", панельный, кирпичный.

При использовании трех категорий домов вводятся две фиктивные переменные z_1 и z_2. Пусть переменная z_1 принимает значение 1 для панельного дома и 0 для всех остальных типов домов, переменная z_2 принимает значение 1 для кирпичных домов и 0 для всех остальных, тогда переменные z_1 и z_2 принимают значения 0 для домов типа "хрущевки".

Предположим, что уравнение регрессии с фиктивными переменными составило:

$$\hat{y} = 320 + 500 \cdot x + 2200 \cdot z_1 + 1600 \cdot z_2 .$$

Частные уравнения регрессии для отдельных типов домов, свидетельствуя о наиболее высоких ценах квартир в панельных домах, будут иметь следующий вид:

- "хрущевки" $\hat{y} = 320 + 500 \cdot x$,
- панельные $\hat{y} = 2520 + 500 \cdot x$,
- кирпичные $\hat{y} = 1920 + 500 \cdot x$.

Параметры при фиктивных переменных z_1 и z_2 представляют собой разность между средним уровнем результативного признака для соответствующей группы и базовой группы. В рассматриваемом примере за базу сравнения цены взяты дома "хрущевки", для которых $z_1 = z_2 = 0$. Параметр при $z_1 = 2\,200$ означает, что при одной и той же полезной площади квартиры цена ее в панельных домах в среднем на \$ 2 200 выше, чем в "хрущевках". Соответственно параметр при z_2 показывает, что в кирпичных домах цена выше в среднем на \$ 1 600 при неизменной величине полезной площади по сравнению с указанным типом домов.

Пример 9.4. Известны затраты на газ и электроэнергию в США в течении 1977 – 1982 гг.

Таблица 9.2

Исходные данные

T	Y	T	Y
1	7,33	13	7,74
2	4,7	14	5,1
3	5,1	15	5,67
4	5,46	16	5,92
5	7,65	17	8,04
6	4,92	18	5,27
7	5,15	19	5,51
8	5,55	20	6,04
9	7,96	21	8,26
10	5,01	22	5,51
11	5,05	23	5,41
12	5,59	24	5,83

Имеем типичную ситуацию, когда желая увеличить объем выборки, вместо среднегодовых используют квартальные данные. Несмотря на то, что выборка при этом увеличилась в четыре раза, значимость модели резко ухудшилась, в коэффициент детерминации уменьшился до нуля. Причина заключается в том, что при переходе к квартальным данным появились сезонные колебания, на интенсивном фоне которых совсем потерялся исследуемый эффект – временной линейный тренд (рис. 9.1).

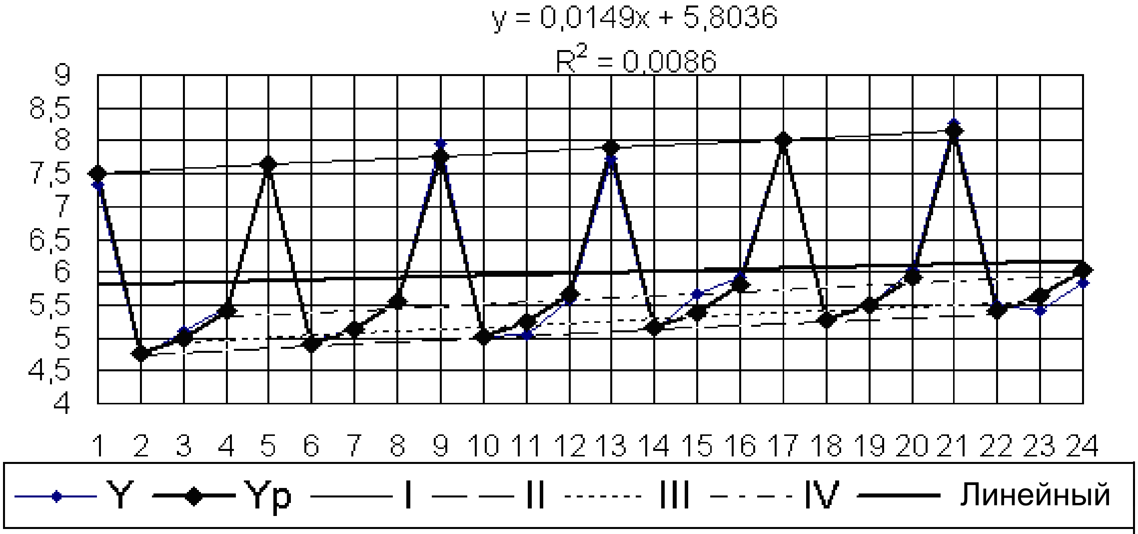


Рис. 9.1. **Сезонные колебания данных по кварталам**

Следовательно, необходимо учесть в модели эту особенность процесса, и поэтому самым экономным способом является введение *dummy-переменных* для описания межквартальных отличий.

Поэтому необходимо ввести в модель четыре дополнительные *dummy-переменные*:

$z_1 = 1$ – для 1-го квартала и $z_1 = 0$ – для всех остальных кварталов;

$z_2 = 1$ – для 2-го квартала и $z_2 = 0$ – для всех остальных кварталов;

$z_3 = 1$ – для 3-го квартала и $z_3 = 0$ – для всех остальных кварталов;

$z_4 = 1$ – для 4-го квартала и $z_4 = 0$ – для всех остальных кварталов.

Если за эталон взять 1-й квартал, то переменную z_1 не следует включать в модель. В результате вычислений будет получено уравнение регрессии для эталонной категории с исправлениями для свободного члена для других категорий.

Можно подключить в модель все сразу *dummy-переменные*, но тогда в модель следует включить свободный член, поскольку $z_1 + z_2 + z_3 + z_4 = 1$. В результате таких вычислений будут получены уравнения регрессии для каждой категории отдельно и правильно подсчитаны все статистические характеристики за исключением статистики Фишера, которая будет занижена в $\frac{m+1}{m} = \frac{5}{4}$ раза.

При таких разных подходах по критерию Стьюдента оцениваются разные эффекты качественного фактора.

По другой методике оценивается значимость эффекта каждой категории, а по первой, рекомендуемой, значимость расхождений каждой категории от эталонной.

Сейчас неявно предусматривается, что для каждой совокупности (для каждой категории) сохраняются неизменными все закономерности в зависимостях от количественных переменных, а влияние качественного признака проявляется только в управлениях свободного члена. Исходные данные вместе с введенными *dummy-переменными* представлены в табл. 9.3, где T -номер квартала, Y -затраты (результативный признак).

Коэффициенты регрессии перед z_2, z_3, z_4 в этой модели показывают, что затраты на газ и электроэнергию в эталонном 1-м квартале существенно отличаются от других кварталов.

Приведем уравнения регрессии для каждого квартала:

$$Y_p = 7,4825 - 2,777 \cdot z_2 - 2,278 \cdot z_3 - 2,193 \cdot z_4 + 0,0316 \cdot T\ ;\ R^2 = 0,9867\ .$$

(tb) (98,5) (33,2) (30,7) (25,9) (7,3)

$$Y_p\ \ I\ \ = 7,4825 + 0,0316 \cdot T\ ,\ (z_1 = 1);$$

$$Y_p\ \ II\ \ = 7,4825 - 2,777 + 0,0316 \cdot T = 4,7059 + 0,0316 \cdot T\ ,\ (z_2 = 1);$$

$$Y_p\ \ III\ \ = 7,4825 - 2,278 + 0,0316 \cdot T = 4,9034 + 0,0316 \cdot T\ ,\ (z_3 = 1);$$

$$Y_p\ \ IV\ \ = 7,4825 - 2,193 + 0,0316 \cdot T = 5,2894 + 0,0316 \cdot T\ ,\ (z_4 = 1).$$

На фоне интенсивных сезонных колебаний значимость исследуемого линейного тренда оказалась сниженной до нуля ($R^2 = 0,0086$), оценка углового коэффициента занижена более чем в два раза.

Таблица 9.3

Данные для построения модели

Z_1	Z_2	Z_3	Z_4	T	Y	Y_p
1	0	0	0	1	7,33	7,5141
0	1	0	0	2	4,7	4,7691
0	0	1	0	3	5,1	4,9991
0	0	0	1	4	5,46	5,4158
1	0	0	0	5	7,65	7,6405
0	1	0	0	6	4,92	4,8955
0	0	1	0	7	5,15	5,1255
0	0	0	1	8	5,55	5,5421
1	0	0	0	9	7,96	7,7668
0	1	0	0	10	5,01	5,0218
0	0	1	0	11	5,05	5,2518
0	0	0	1	12	5,59	5,6685
1	0	0	0	13	7,74	7,8932
0	1	0	0	14	5,1	5,1482
0	0	1	0	15	5,67	5,3782
0	0	0	1	16	5,92	5,7948
1	0	0	0	17	8,04	8,0195
0	1	0	0	18	5,27	5,2745
0	0	1	0	19	5,51	5,5045
0	0	0	1	20	6,04	5,9212
1	0	0	0	21	8,26	8,1459
0	1	0	0	22	5,51	5,4009
0	0	1	0	23	5,41	5,6309
0	0	0	1	24	5,83	6,0476

С помощью трех дополнительных *dummy-переменных* z_2, z_3, z_4 очень хорошо описана вся зависимость с сезонными колебаниями. Следует отметить, что попытки описать эти колебания аналитическим выражением в виде трех гармоник (с такими периодами, как год, полгода, квартал) будут неэкономными и потребуют 6-ти дополнительных параметров (по два параметра на каждую гармонику). Можно в модель включить все сразу *dummy-переменные*, но тогда модель не должна иметь свободный член. Вычисляем параметры такой линейной модели:

$$Y_p = b_1 z_1 + b_2 z_2 + b_3 z_3 + b_4 z_4 + b_5 T$$ (константа отсутствует).

В таком случае будут получены все уравнения для каждого квартала (не надо пересчитывать свободный член). Однако теперь статистики Стьюдента будут оценивать отклонения среднего уровня затрат в квартале от нуля, а не от выбранного эталона. Все характеристики, кроме F, получатся такие же, только статистику Фишера следует увеличить в $5/4$ раза: $F = 280{,}93 \cdot \dfrac{5}{4} = 351{,}16$.

Пример 9.5. Имеем исследование влияния затрат на рекламу на объемы реализации мобильных телефонов. В табл. 9.4 приведены данные об относительной доли реализованных телефонов в филиалах фирмы, имеющих разные затраты на рекламу. Количество реализованных телефонов m в партиях товара из n единиц для каждого уровня затрат на рекламу x распределено по биномиальному закону $P(Y = m) = P_n(m) = C_n^m p^m (1 - p)^{n-m}$ с математическим ожиданием (средним) $y_p = M(m) = np$ и дисперсией $D(m) = np(1 - p)$.

Таблица 9.4

Затраты на рекламу в филиалах

Филиал	Затраты на рекламу (x)	Объем партии товара (n)	Количество проданных телефонов (m)	Относительная доля реализованных телефонов (p)
1	10,2	50	44	0,8800
2	7,7	49	42	0,8571
3	5,1	46	24	0,5217
4	3,8	48	16	0,3333
5	2,6	50	6	0,1200

Вероятность $P_n\ m$ зависит от двух параметров: n – количества товара в партии и p – удельной относительной доли реализованных телефонов. Относительная часть реализованных телефонов $0 \le p \le 1$ будет большой (близкой к 1) при больших затратах на рекламу и маленькой (близкой к нулю) при небольших.

Необходимо аппроксимировать зависимость $p\ x$ достаточно простым выражением. По данным табл. 9.4 построен график зависимости p от x (рис. 9.2). Экспериментальных точек мало, хотя каждая точка является результатом $n \approx 50$ наблюдений. Очевидно, что эта сигмоидальная зависимость должна быть похожей на интегральную функцию распределения (не обязательно нормального закона) или на логистическую зависимость. Забегая вперед, следует отметить, что обе аппроксимации привели к практически одинаковым результатам (на рис. 9.2 сплошная и пунктирная линии).

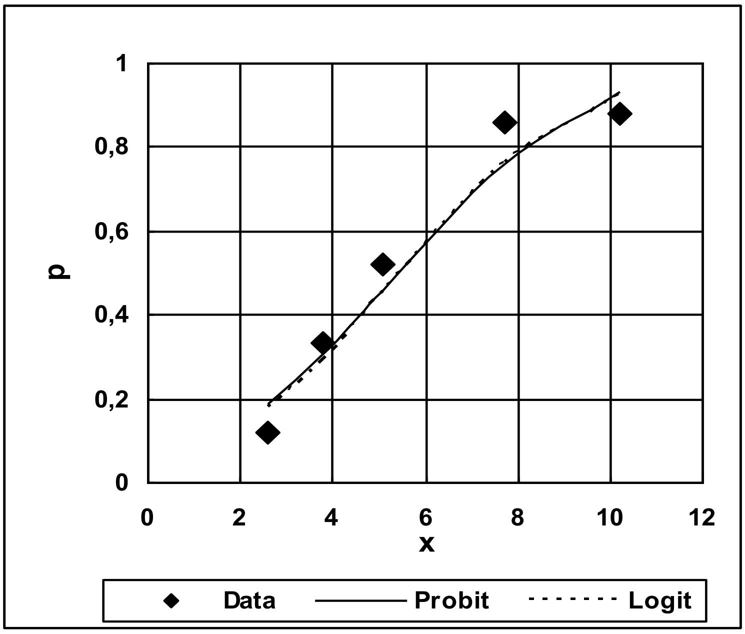


Рис. 9.2. **Аппроксимация относительной доли реализованного товара в зависимости от затрат на рекламу**

Если применяется аппроксимация функцией нормального распределения, используется название "пробит-анализ"; если применяется аппроксимация логистической функцией, используется название "логит-анализ". Сравним оба вида биномиальной модели.

Пробит-анализ. Аппроксимируем зависимость параметра $p(x)$ биномиального распределения интегральной функцией стандартного нормального закона:

$$p(x)=\Phi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{t} e^{-\frac{z^2}{2}}dz,$$

где $t=\Phi^{-1}(p)$ – функция, обратная $\sqrt{}$ к функции распределения нормального закона;

$t=\dfrac{x-a}{\sigma}=b_0+b_1x$ – линейный предиктор (откуда и следует название "линейная" модель).

Для каждого известного значения p из табл. 9.4 находим квантили стандартного нормального распределения $t=\Phi^{-1}(p)$ (табл. 9.5) и обозначаем их в виде точек $x-t$ на рис. 9.3а. Далее методом наименьших квадратов оцениваем коэффициенты регрессии линейного предиктора $t=0,3113x-1,6922$ и вычисляем расчетные (сглаженные) значения (табл. 9.4): $p_n=\Phi(0,3113x-1,6822)$.

Таблица 9.5

Расчетные значения $p(x)$ в пробит- p_n и логит- анализе p_s

№п/п	x	p	t	p_n	s	p_s
1	10,2	0,8800	1,174987	0,930978	1,99243	0,924514
2	7,7	0,8571	1,067571	0,759548	1,81529	0,767224
3	5,1	0,5217	0,054519	0,458368	0,080043	0,457004
4	3,8	0,3333	-0,43073	0,305291	-0,70819	0,298392
5	2,6	0,1200	-1,17499	0,188669	-1,99243	0,184671

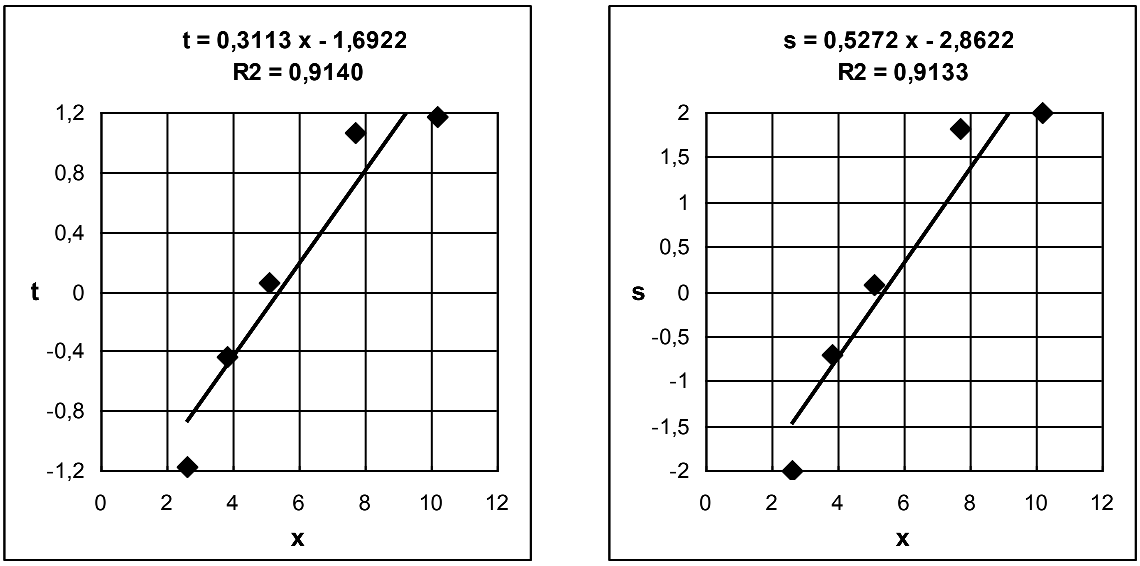
Для сравнения сразу приведем вычисления с помощью логит-анализа.

Логит-анализ. Аппроксимируем зависимость параметра $p(x)$ биномиального распределения логистической функцией:

$$p(x) = \frac{1}{1+e^{-s}} = \frac{1}{1+e^{-(a+bx)}} \quad \leftrightarrow \quad s = \ln\frac{p}{1-p},$$

где $s = a + bx$ – линейный предиктор.

Для каждого известного значения p из табл. 9.4 находим значение $s = \ln\dfrac{p}{1-p}$ (табл. 9.5) и отмечаем их в виде точек $(x-s)$ на графике рис. 9.3б. Далее методом наименьших квадратов оцениваем коэффициенты регрессии линейного предиктора $s = 0{,}5272x - 2{,}8622$ и вычисляем расчетные (сглаженные) значения $p_s = \dfrac{1}{1+e^{-(0{,}5272x-2{,}8622)}}$ (табл. 9.5).

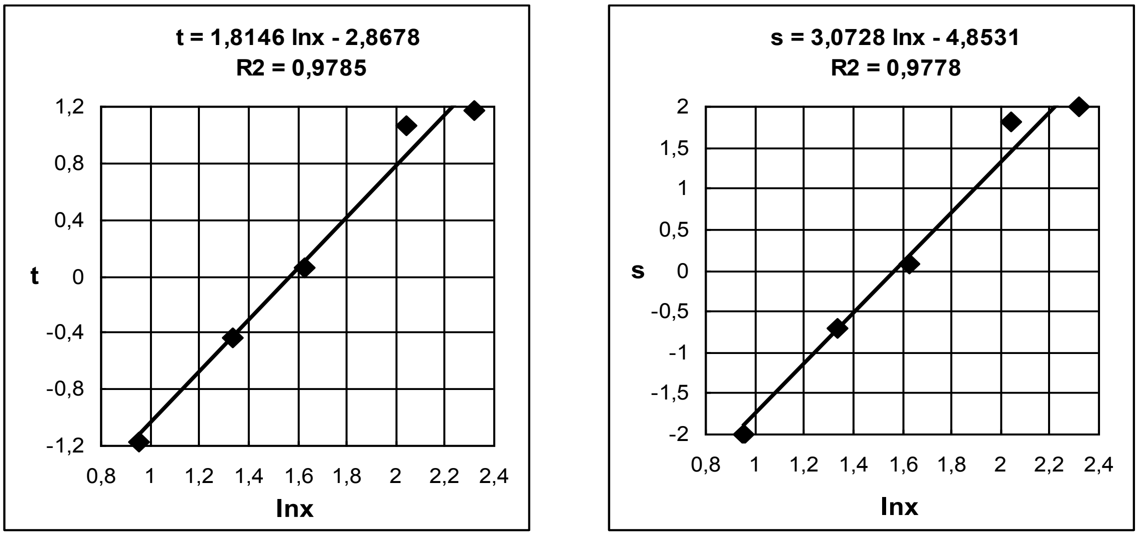


а) **пробит-анализ** б) **логит-анализ**

Рис. 9.3. **Аппроксимация зависимости *p(x)* в функциональных переменных**

На рис. 9.3 построены графики $p_n(x)$ – сплошной линией, $p_s(x)$ – пунктирной линией. Видим, что вычисленные значения p_n и p_s по двум видам анализа (пробит- и логит-) близки между собой, то есть оба вида анализа эквивалентны. Это не случайно, поскольку логистической зависимостью можно хорошо аппроксимировать функцию распределения нормального закона. Как определенное преимущество логит-анализа отметим более простой вид математических формул, записанных с помощью элементарных функций.

Если правильно принято предположение о логистической форме связи, точки на графике $x-s$ должны группироваться около некоторой прямой, однако здесь (как и для пробит-анализа), есть явные отклонения от прямой. В обобщенных линейных моделях предполагаются предварительные функциональные преобразования переменных. Так, в приведенном примере целесообразно перейти к логарифмам объясняющей переменной. При этом существенно улучшится качество аппроксимации как при пробит-, так и при логит-анализе (рис. 9.4).



а) **пробит-анализ** б) **логит-анализ**

Рис. 9.4. **Аппроксимация зависимости $p\ \ln x$ в функциональных переменных**

Из рис. 9.4 видно, что коэффициент детерминации после перехода от логарифмов увеличился от 0,91 до 0,98. Теперь можно назвать обобщенную линейную модель обобщенной логарифмической моделью. Следовательно, функция распределения результативной переменной (y – количество реализованных телефонов) определена полностью.

Она зависит от двух параметров $n = const$ и $p = f\ \ x$:

$$p = \frac{1}{1+e^{-s}} = \frac{1}{1+e^{-\ 3{,}0728\ln x - 4{,}8531}} \approx \frac{x^3}{x^3+100}.$$

Для известного $n = 50$ и для каждого $x = 0, 1, 2, 3, \ldots, 12$ вычислены средние значения (математические ожидания) результативной переменной $Y = np\ \ x$ и средние квадратические отклонения $\sigma = \sqrt{np\ \ x\ \ \ 1 - p\ \ x}$ (табл. 9.6).

При $n = 50$ биномиальное распределение уже асимптотически нормальное (распределение Лапласа), поэтому границы 95 %-х доверительных интервалов определяем по правилу: $Y_d = Y - t_{0,05} \cdot \sigma$ – нижняя граница (*down*), $Y_u = Y + t_{0,05} \cdot \sigma$ – верхняя (*up*) граница, $t_{0,05} = 2,01$ находим из таблицы критических точек распределения Стьюдента (приложение В) при $dfe = n - 2 = 48$.

На рис. 9.5 изображен график зависимости $y\ \ x$ вместе с 95 %-й доверительной полосой. Точками обозначены наблюдаемые значения m (скорректированные для одинакового объема партий $n = 50$).

Таблица 9.5

Вычисленные значения y с 95 %-ми доверительными полосами

x	s	p	σ	$Y = np$	-95%	+95%
1	-4,853	0,008	0,620	0,387	-0,859	1,633
2	-2,723	0,062	1,700	3,081	-0,337	6,499
3	-1,477	0,186	2,750	9,292	3,763	14,820
4	-0,593	0,356	3,385	17,794	10,989	24,599
5	0,092	0,523	3,532	26,154	19,055	33,253
6	0,653	0,658	3,355	32,880	26,136	39,624
7	1,126	0,755	3,041	37,758	31,646	43,869
8	1,537	0,823	2,699	41,149	35,724	46,574
9	1,899	0,870	2,380	43,486	38,702	48,270
10	2,222	0,902	2,100	45,112	40,890	49,333
11	2,515	0,925	1,860	46,260	42,521	49,999
12	2,783	0,942	1,657	47,086	43,757	50,416

На этом же примере сравним результаты проведенных вычислений (по обобщенной линейной модели) с результатами вычислений по нелинейному регрессионному анализу.

Пусть относительная доля $p(x)$ является результативным признаком. Используя предыдущее исследование, можно получить ту же самую форму связи для описания выбранного результативного признака:

$$y = p = \frac{1}{1+e^{-s(x)}} \leftrightarrow s(x) = \ln\frac{p}{1-p}, \ \hat{s} = 3{,}0728 \cdot \ln x - 4{,}8531.$$

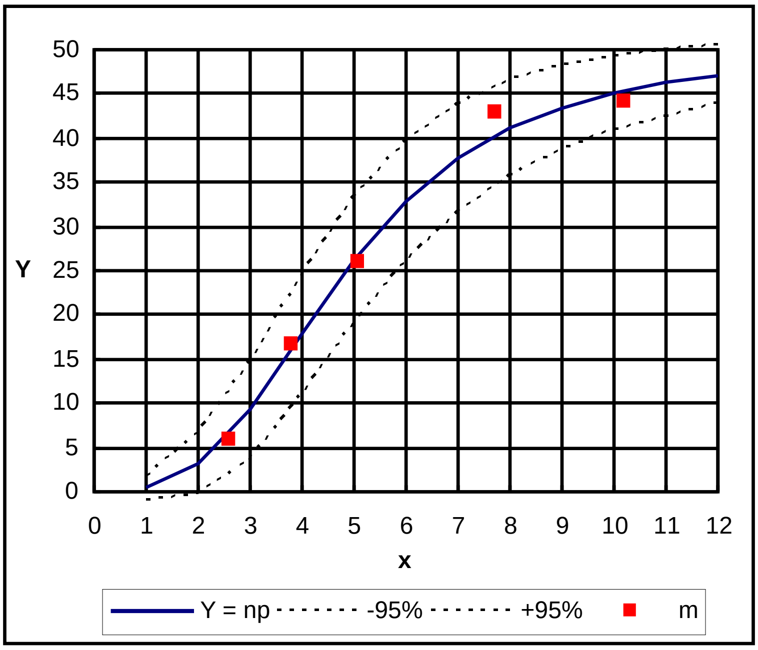


Рис. 9.5. **Результаты анализа по обобщенной модели**

В функциональных координатах $(X = \ln x, s)$ изучаемая зависимость является линейной. Следовательно, до этого момента методики обобщенной линейной модели и регрессионного анализа совпадали (естественно, только для данного простого примера). Однако, далее начинаются существенные расхождения.

В обобщенной линейной модели $p(x)$ считается параметром биномиального распределения результативного признака (второй параметр $n = 50$). Зная функцию распределения, для каждого значения аргумента x вычисляем ожидаемое значение $\hat{y} = M\left(\frac{m}{n}\right) = \frac{n\hat{p}}{n} = \hat{p}$ и дисперсию

$$\sigma^2(\hat{y}) = D\left(\frac{m}{n}\right) = \frac{n\hat{p}(1-\hat{p})}{n^2} = \frac{\hat{p}(1-\hat{p})}{n}.$$

График уравнения регрессии для $y = p\ \ x$ вместе с границами 95 %-го интервала (рис. 9.6а) только масштабом оси ординат отличается от графика на рис. 9.5.

В традиционном регрессионном анализе принята совсем другая методика определения границ доверительной полосы. Параметр $n = const$ (число объектов в группе для вычисления относительной доли $p = \frac{m}{n}$) вообще не входит в модель. Считается, что имеется всего $N = 5$ наблюдений (5 точек) в функциональных координатах X, s . Для этих 5-ти точек находим несмещенную оценку дисперсии остатков линейной регрессионной модели $MSE = const$, где $MSE = \frac{\sum\ s - \hat{s}\ ^2}{N-2} = 0{,}331079$.

Далее для каждого значения аргумента вычисляется дисперсия расчетных значений по формуле:

$$\sigma^2\ \ \hat{s}\ \ = \frac{MSE}{N}\left(1 + \frac{X - \bar{X}\ ^2}{s_X^2}\right) = \frac{0{,}331079}{5}\left(1 + \frac{\ln x - 5{,}88\ ^2}{7{,}5336}\right).$$

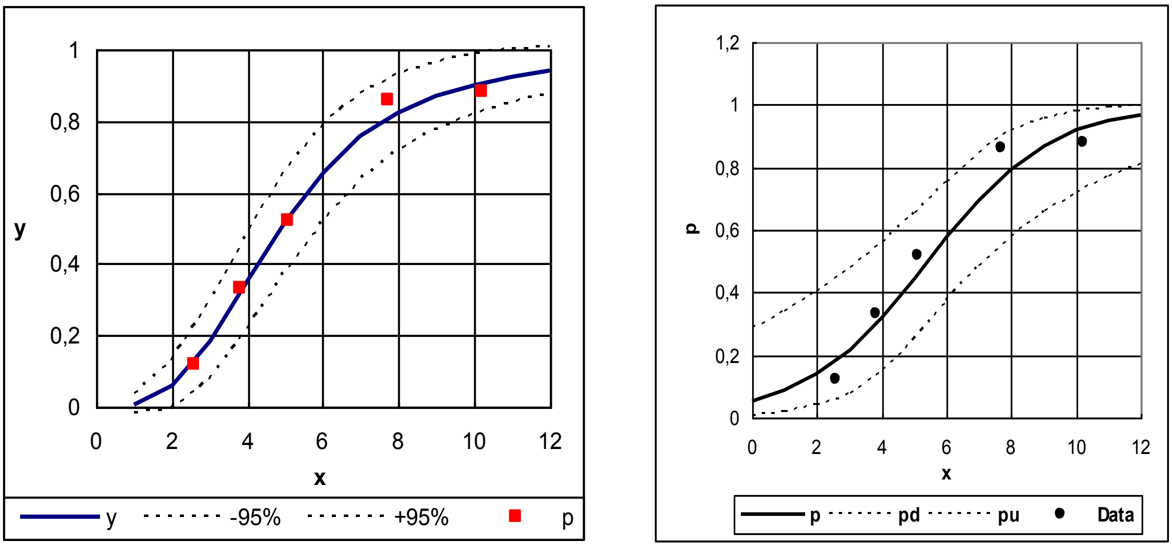
В табл. 9.7 для $x = 0, 1, 2, 3, \ldots, 12$ вычислены расчетные значения $\hat{s}$, их дисперсии $\sigma^2\ \ \hat{s}$ и границы 95 %-й доверительной полосы s_d и s_u: $s_d = \hat{s} - t_{0,05}\ \ N-2\ \ \cdot \sigma\ \ \hat{s}$, $s_u = \hat{s} + t_{0,05}\ \ N-2\ \ \cdot \sigma\ \ \hat{s}$;
где $t_{0,05}\ \ 3\ = 3{,}1824$ найдено по таблице критических точек распределения Стьюдента (приложение В).

Далее выполнено обратное преобразование $p = \frac{1}{1 + e^{-s}}$ и на основании этого построен график (рис. 9.6б). Сравнивая рис. 9.6а и рис. 9.6б отметим, что хотя линии регрессии одинаковые (для данного простого примера), но есть существенные отличия в ширине доверительной полосы (полосе неопределенности). Метод нелинейного регрессионного анализа привел к менее определенным результатам сравнительно с методикой обобщенной линейной модели.

Таблица 9.7

Вычисление нелинейной регерссионной зависимости

X	*s*	σ^2	s_d	s_u
0	-2,862	0,370104	-4,79808	-0,92592
1	-2,3348	0,27553	-4,0053	-0,6643
2	-1,8076	0,198535	-3,22561	-0,38959
3	-1,2804	0,139119	-2,46741	-0,09339
4	-0,7532	0,097281	-1,7458	0,239402
5	-0,226	0,073022	-1,08598	0,633981
6	0,3012	0,066342	-0,5185	1,120903
7	0,8284	0,077241	-0,05608	1,712875
8	1,3556	0,105719	0,320845	2,390355
9	1,8828	0,151775	0,642971	3,122629
10	2,41	0,215411	0,932952	3,887048
11	2,9372	0,296625	1,203936	4,670464
12	3,4644	0,395417	1,463207	5,465593



а б

Рис. 9.6. **Сравнение обобщенной линейной модели** (а) **с нелинейным регрессионным анализом** (б)

9.3. Задачи

Задача 9.1. В табл. 9.8 приведены данные для линейного трехфакторного уравнения регрессии:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \varepsilon_t \qquad t = 1, \ldots, T \quad .$$

Таблица 9.8

Исходные данные

x_{1t}	10,3	16	14	17,1	15
x_{2t}	20,8	28	25	34	21,7
x_{3t}	16	20,3	18	30	21
y_t	50	70	55	60	65

Необходимо:

1. Определить гребневые оценки параметров для гребневой константы, равной 0,3 и 0,5.

2. По обоим оцененным уравнениям построить точечный прогноз математического ожидания целевой переменной y_0 для значений экзогенных переменных $x_{10}, x_{20}, x_{30} \;=\; 15; 18; 10$.

Задача 9.2. Исследовать зависимость между результатами письменных экзаменов по математике. Получены следующие данные о количестве решенных задач студентами на экзамене (x) и при проведении контроля остаточных знаний (y), а также распределение этих студентов по фактору "пол" (табл. 9.9).

Таблица 9.9

Исходные данные

№ студента	Количество решенных задач		Пол студента
	на экзамене, x	при проведении контроля остаточных знаний, y	
1	2	3	4
1	10	8	муж.
2	7	6	жен.
3	8	5	муж.

Окончание табл. 9.9

1	2	3	4
4	8	5	жен.
5	9	7	муж.
6	7	6	жен.
7	6	4	муж.
8	6	5	муж.
9	5	3	жен.
10	8	7	муж.
11	9	6	муж.
12	6	3	жен.

Построить линейную регрессионную модель y по x с использованием фиктивной переменной по фактору "пол". Можно ли считать, что эта модель одинакова для юношей и девушек?

Задача 9. 3. Исследуется эффективность лекарств y в зависимости от возраста пациента x (табл. 9.10).

Таблица 9.10

Исходные данные

Лекарство	y	x	Лекарство	y	x
a	54	69	b	30	40
b	30	48	b	23	41
a	58	73	a	21	55
b	66	64	b	43	45
b	67	60	a	38	58
a	64	62	b	43	58
a	67	70	a	43	64
a	33	52	b	45	55
a	33	63	b	48	57
b	42	48	a	48	63
b	33	46	a	53	60
a	28	55	b	58	62

Построить линейную регрессионную модель y по x с использованием фиктивной переменной по фактору "вид лекарства". Сравнить эффективность лекарств вида a и b.

Задача 9. 4. Исследуется надежность станков трех производителей a, b и c. При этом учитывается возраст станка (в месяцах), x и время (в часах) безаварийной работы до последней поломки, y. Выборка из 40 станков дала следующие результаты (табл. 9.11).

Таблица 9. 11

Исходные данные

Фирма	Время безаварийной работы, y	Возраст станка, x	Фирма	Время безаварийной работы, y	Возраст станка, x
a	280	23	a	200	52
b	230	30	b	265	20
c	112	65	c	148	70
a	176	69	c	150	62
c	90	75	b	176	40
a	176	63	a	123	66
b	216	25	a	245	20
c	110	75	c	176	39
b	45	75	b	260	25
a	236	48	b	220	22
a	205	59	c	194	33
a	240	25	a	156	48
b	65	69	a	100	75
a	115	71	b	240	21
c	200	26	a	170	56
b	126	45	c	116	58
a	225	40	b	120	40
c	210	30	a	240	37
b	45	69	b	88	56
a	260	30	a	120	67

Построить линейную регрессионную модель y по x без учета производителя и с использованием фиктивных переменных по фактору "производитель". Сравнить полученные результаты. Сделать вывод.

Задача 9.5. *Logit-модель* была применена к выборке, в которой $y=1$, если количество занятых в фирме выросло ($y=0$ – в противном случае), x_1– доход фирмы, в млн. \$; $x_2=1$, если фирма относится к области высоких технологий ($x_2=0$ – в противном случае).

Получена следующая модель: $\hat{z}=0,40+0,20\cdot x_1+0,10\cdot x_2$.

Требуется определить оценку вероятности роста занятости для фирмы высокотехнологичной фирмы A с доходом в \$ 5 млн и для фирмы B, не относящейся к сфере высоких технологий и имеющей доход \$ 7 млн.

Задача 9.6. Имеется выборка, состоящая из 528 наблюдений, в которой $y=1$, если заработная плата работника ниже \$ 5 в час ($y=0$ – в противном случае). Предполагается, что уровень заработной платы зависит от следующих факторов: x_1 – образование, лет; x_2 – пол (1–женский, 0 – мужской); x_3 – опыт работы, лет. В табл. 9.12 приведены коэффициенты, полученные при оценке линейной регрессии y от x_1, x_2 и x_3 с помощью МНК, и при оценке *Logit-модели* с помощью нелинейного МНК.

Таблица 9.12

Коэффициенты уравнения регрессии

	Коэффициенты		Выборочные средние
	линейной регрессии	*Logit-модели*	
1	0,94	5,87	1
x_1	–0,05	–0,56	13,09
x_2	0,15	1,26	0,46
x_3	–0,01	–0,06	17,66

Требуется:

1. Определить на основе *Logit-модели*, оценку вероятности для мужчины и для женщины, имеющих 12 лет образования и 15 лет опыта работы, оказаться низкооплачиваемыми работниками.

2. Определить на основе *Logit-модели*, изменение оценки вероятности быть низко оплачиваемым работником для мужчины с характеристиками из п. 1, если он проучится на один год больше.

3. Ответить на вопросы пунктов 1–2 с использованием линейной регрессионной модели.

Задача 9.7. Имеется выборка, состоящая из 528 наблюдений, в которой $y=1$, если работник состоит в профсоюзе ($y=0$ – в противном случае). Предполагается, что членство в профсоюзе зависит от следующих факторов: x_1 – образование, лет; x_2 – пол (1 – женский, 0 – мужской); x_3 – опыт работы, лет; x_4 – опыт работы в квадрате. Выборочные средние равны: $\bar{y}=0{,}18$; $\bar{x}_1=13{,}09$; $\bar{x}_2=0{,}46$; $\bar{x}_3=17{,}66$; $\bar{x}_4=459{,}45$.

На основе выборочных данных была получена следующая *Probit-модель*: $\hat{y}=F\ \ -0{,}900-0{,}015x_1-0{,}599x_2+0{,}029x_3-0{,}0003x_4$

Требуется определить, на сколько снижается вероятность быть членом профсоюза в расчете на год дополнительного образования.

10. Системы одновременных уравнений

10.1. Основные теоретические сведения

Довольно часто сложные экономические процессы моделируются несколькими связанными между собой уравнениями, содержащими как повторяющиеся, так и собственные переменные. Наибольшее распространение в эконометрических исследованиях получила система взаимозависимых или одновременных уравнений. В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Такая система взаимозависимых уравнений называется также структурной формой модели. Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной переменной. В такой системе каждое уравнение не может рассматриваться самостоятельно, а для нахождения его параметров используются специальные приемы оценивания, так как МНК, в этом случае, неприменим.

Наибольшее распространение получили такие методы оценивания коэффициентов структурной модели:

1) косвенный метод наименьших квадратов;
2) двухшаговый метод наименьших квадратов;
3) трехшаговый метод наименьших квадратов;
4) метод максимального правдоподобия с полной информацией;
5) метод максимального правдоподобия с ограниченной информацией.

Косвенный метод наименьших квадратов (КМНК) применяется для идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы:

1) структурная модель преобразовывается в приведенную форму модели;

2) для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij};

3) коэффициенты приведенной формы модели трансформируются в параметры структурной модели.

Двухшаговый метод наименьших квадратов (ДМНК) используется для оценки коэффициентов сверхидентифицируемой модели, основная идея которого – на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения. Далле, подставив их вместо фактических значений можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения.

Исходная система уравнений называется:

а) идентифицируемой, если по оценкам коэффициентов приведенных уравнений можно однозначно определить коэффициенты структурных уравнений (обычно это удается сделать, когда количество уравнений для определения коэффициентов структурных уравнений в точности равно количеству этих коэффициентов);

б) неидентифицируемой (недоопределенной), если по коэффициентам приведенных уравнений можно получить несколько вариантов значений коэффициентов структурных уравнений (обычно это происходит когда количество уравнений для определения коэффициентов структурных уравнений меньше числа определяемых коэффициентов);

в) сверхидентифицируемой (переопределенной), если по оценкам коэффициентов приведенных уравнений невозможно определить коэффициенты структурных уравнений. В данном случае система, связывающая коэффициенты структурных уравнений с коэффициентами приведенных уравнений, является несовместной (обычно в таких случаях число уравнений для оценки коэффициентов структурных уравнений больше числа определяемых коэффициентов).

Сверхидентифицируемая структурная модель может быть двух типов:

1) все уравнения системы сверхидентифицируемы;
2) система содержит наряду со сверхидентифицируемыми и идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемы, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если же в системе есть идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Дальнейшим развитием ДМНК является трехшаговый метод наименьших квадратов (ТМНК). Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным является ДМНК.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к достаточно сложным вычислительным процедурам. Поэтому в качестве модификации служит метод максимального правдоподобия с ограниченной информацией (метод наименьшего дисперсионного отношения).

При переходе от приведенной формы модели к структурной исследователь сталкивается с проблемой идентификации. Идентификация – это единственность соответствия между приведенной и структурной формами модели. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы.

Для определения идентифицируемости структурных уравнений применяются необходимые и достаточные условия. Пусть система одновременных уравнений включает в себя N уравнений относительно N эндогенных переменных. Система содержит M экзогенных либо предопределнных переменных. Пусть n и m – количество соответственно эндогенных и экзогенных переменных в проверяемом на идентифицируемость уравнении. Переменные, не входящие в данное уравнение, но входящие в другие уравнения системы, называют исключенными переменными (их количество равно $N-n$ для эндогенных и $M-m$ для экзогенных переменных соответственно).

***Первое необходимое условие идентифицируемости*.**

Уравнение идентифицируемо, если оно исключает по крайней мере $N-1$ переменную (эндогенную или экзогенную), присутствующую в модели:

$$N-n + M-m \geq N-1.$$

Второе необходимое условие идентифицируемости.

Уравнение идентифицируемо, если количество исключенных из уравнения экзогенных переменных не меньше количества эндогенных переменных в этом уравнении, уменьшенного на единицу:

$$M - m \geq n - 1.$$

Знаки равенств в обоих необходимых условиях соответствуют точной идентификации уравнений.

Необходимое и достаточное условие идентифицируемости. В модели, содержащей N уравнений относительно N эндогенных переменных, условие идентифицируемости выполняется тогда и только тогда, когда ранг матрицы, составленной из исключенных их данного уравнений переменных, но входящих в другие уравнения системы, равен $N-1$.

10.2. Примеры

Пример 10.1. Примером системы одновременных уравнений может служить ***модель динамики цены и заработной платы вида:***

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

где y_1 – темп изменения месячной заработной платы;

y_2 – темп изменения цен;

x_1 – процент безработных;

x_2 – темп изменения постоянного капитала;

x_3 – темп изменения цен на импорт сырья.

В рассмотренных классах систем эконометрических уравнений структура матрицы коэффициентов при зависимых переменных различна.

Представим систему эконометрических уравнений в матричном виде:

$$BY + \Gamma X = E,$$

где B – матрица коэффициентов при зависимых переменных;

Y – вектор зависимых переменных;

Γ – матрица параметров при объясняющих переменных;
X – вектор объясняющих переменных;
E – вектор ошибок.

Если матрица B диагональная, то рассматриваемая модель является системой независимых уравнений. Так, при трех зависимых и трех объясняющих переменных модель имеет вид:

$$\begin{cases} y_1 = a_{01} + a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + E_1, \\ y_2 = a_{02} + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + E_2, \\ y_3 = a_{03} + a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + E_3. \end{cases}$$

Матрица параметров при зависимых переменных является диагональной:

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Если матрица B треугольная (или может быть приведена к такому виду), то модель представляет собой систему рекурсивных уравнений. Так, если модель имеет вид:

$$\begin{cases} y_1 = a_{01} + a_{11}x_1 + a_{12}x_2 + E_1, \\ y_2 = a_{02} + b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + E_2, \\ y_3 = a_{03} + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + E_3. \end{cases}$$

То есть зависимая переменная y_1 первого уравнения участвует как объясняющая переменная во втором уравнении системы, а зависимая переменная y_2 второго уравнения рассматривается как объясняющая переменная в третьем уравнении. Тогда матрица коэффициентов при зависимых переменных модели составит:

$$B = \begin{pmatrix} 1 & 0 & 0 \\ -b_{21} & 1 & 0 \\ 0 & -b_{32} & 1 \end{pmatrix},$$

то есть представляет собой треугольную матрицу.

Если матрица B не является ни диагональной, ни треугольной, то модель представляет собой систему одновременных уравнений. Так, для модели вида:

$$\begin{cases} y_1 = a_{01} + b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + E_1; \\ y_2 = a_{02} + b_{21}y_1 + b_{23}y_3 + a_{23}x_3 + E_2; \\ y_3 = a_{03} + b_{31}y_1 + a_{32}x_2 + a_{33}x_3 + E_3 \end{cases}$$

получим матрицу коэффициентов при зависимых переменных:

$$B = \begin{pmatrix} 1 & -b_{12} & 0 \\ -b_{21} & 1 & -b_{23} \\ -b_{31} & 0 & 1 \end{pmatrix},$$

которая не является ни диагональной, ни треугольной. Соответственно это отражается на выборе метода оценки параметров эконометрических систем.

Пример 10.2. Построить эконометрическую модель, которая характеризует зависимость между объемом валового национального продукта от производственных ресурсов, а именно: основных производственных фондов, рабочей силы, материальных ресурсов.

Решение

Для построения такой модели целесообразно строить эконометрическую модель на основе системы одновременных структурных уравнений:

$$X_t = f\ (F_t, L_t, u_t)\ ;$$

$$M_t = f\ (X_t, v_t)\ ;$$

$$Y_t = X_t - M_t,$$

где X_t – выпуск продукции; Y_t – валовой национальный продукт; F_t – основные производственные фонды; L_t – рабочая сила; M_t – материальные ресурсы; t – период времени.

Запишем два первых уравнения в виде:

$$X_t = a_0 F_t^{a_1} L_t^{a_2} u_t;$$

$$M_t = b_0 + b_1 X_t + v_t;$$

$$Y_t = X_t - M_t,$$

где a_0, a_1, a_2, b_0, b_1 – параметры модели, u_t, v_t – остатки.

Таким образом, построенная эконометрическая модель составляется из трех одновременных уравнений: два первых являются регрессионными, а третье – тождественность. Все эти уравнения описывают экономические процессы, которые происходят одновременно и, значит, они должны иметь общее решение.

Пример 10.2. Нужно определить зависимость между заработной платой и продуктивностью работы на предприятии.

Решение

Построим эконометрическую модель, которая описывается системой одновременных уравнений:

$$Y_1 = f\left(Y_2, X_1, X_2, X_3, u_1\right);$$

$$Y_2 = f\left(Y_1, X_1, X_2, X_4, X_5, u_2\right),$$

где Y_1 – заработная плата;

Y_2 – производительность труда;

X_1 – уровень квалификации работающих;

X_2 – стаж работающих;

X_3 – форма оплаты труда;

X_4 – фондоотдача;

X_5 – текучесть рабочей силы;

u_1, u_2 – остатки, в первом и втором уравнениях модели соответственно.

Пример 10.3. Модель "Спрос – предложение" содержит функции спроса, предложения и условие равновесия:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_t, \quad \alpha_1 < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t, \quad \beta_1 < 0, \\ q_t^D = q_t^S. \end{cases}$$

В предоставленной системе первое уравнение – функция спроса, второе уравнение – функция предложения, третье уравнение – условие равновесия, p_t, ϖ_t – цена товара и зарплата в момент времени t, ε_t и ν_t – случайные составляющие.

Имеются следующие результаты наблюдений (табл. 10.1).

Таблица 10.1

Результаты наблюдений

p	10	15	5	8	4
q	6	6	18	12	8
ϖ	2	6	2	7	4

Найдем оценки параметров этой системы с помощью КМНК.

Обозначим $q_t^D = q_t^S$ через q_t, тогда

$$\begin{cases} q_t = \alpha_0 + \alpha_1 p_t + \varepsilon_t; \\ q_t = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t. \end{cases}$$

Приравняем правые части этих уравнений:

$\alpha_0 + \alpha_1 p_t + \varepsilon_{t1} = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t$, отсюда

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2 \varpi_t}{\alpha_1 - \beta_1} + \frac{\nu_t - \varepsilon_t}{\alpha_1 - \beta_1}.$$

Введем обозначения: $\pi_{10} = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$; $\pi_{11} = \dfrac{\beta_2}{\alpha_1 - \beta_1}$; $\psi_t = \dfrac{\nu_t - \varepsilon_t}{\alpha_1 - \beta_1}$.

Тогда $p_t = \pi_{10} + \pi_{11}\varpi_t + \psi_t$. Подставляем это выражение для p_t в первое уравнение, раскроем скобки и введем следующее обозначение:

$\pi_{20} = \alpha_0 + \alpha_1 \pi_{10}$; $\pi_{21} = \alpha_1 \pi_{11}$; $\phi_t = \alpha_1 \psi_t + \varepsilon_t$.

Тогда $q_t = \pi_{20} + \pi_{21}\varpi_t + \phi_t$.

Получаем систему из приведенных уравнений:

$$\begin{cases} p_t = \pi_{10} + \pi_{11}\varpi_t + \psi_t; \\ q_t = \pi_{20} + \pi_{21}\varpi_t + \phi_t. \end{cases}$$

Оценим параметры каждого из этих уравнений при помощи МНК, имеем $\pi_{11} = 0,75$; $\pi_{10} = 5,25$; $\pi_{21} = -0,46$; $\pi_{20} = 11,53$.

Тогда $\alpha_1 = \dfrac{\pi_{21}}{\pi_{11}} = -\dfrac{0,46}{0,75} = -0,61;$

$$\alpha_0 = \pi_{20} - \alpha_1\pi_{10} = 11,53 - \quad -0,61 \quad 5,25 = 14,73.$$

Нельзя оценить коэффициенты β_i на основании полученных результатов. Возникает так называемая проблема идентификации.

Пример 10.4. Кейнсианская модель формирования доходов. Рассматривается закрытая экономика без государственных расходов.

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t; \\ y_t = c_t + i_t, \end{cases}$$

где первое уравнение – функция потребления, второе уравнение – макроэкономическое тождество, y_t и i_t – значения совокупного выпуска и инвестиций в момент времени t, ε_t – случайная составляющая.

Пример 10.5. Одной из возможных нестохастических форм модели равновесия на рынке товаров является такая модель:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t; & (1) \\ \tau_t = \alpha_0 + \alpha_1 y_t; & (2) \\ i_t = \gamma_0 + \gamma_1 r_t; & (3) \\ y_{d\,t} = y_t + \tau_t; & (4) \\ g_t = \overline{g}; & (5) \\ y_t = c_t + i_t + g_t, & (6) \end{cases}$$

где уравнения (1) − функция потребления;

(2) − функция налогов;

(3) − функция инвестиций;

(4) − функция располагаемого дохода;

(5) − государственные расходы;

(6) − макроэкономическое тождество;

y_t − значения в момент времени t национального дохода;

c_t − значения в момент времени t потребления;

i_t – значения в момент времени t желаемого объема частных инвестиций;

g_t – государственных расходов (в данном случае $g_t = \bar{g} = const$);

τ_t – объема налогов;

$y_{d\ t}$ – располагаемого дохода;

r_t – процентной ставки.

$$\begin{cases} y_t = \dfrac{\beta_0}{1-\beta_1} + \dfrac{1}{1-\beta_1} i_t + \dfrac{\varepsilon_t}{1-\beta}; \\ c_t = \dfrac{\beta_0}{1-\beta_1} + \dfrac{\beta_1}{1-\beta_1} i_t + \dfrac{\varepsilon_t}{1-\beta}, \end{cases}$$

где коэффициент $\dfrac{1}{1-\beta_1}$ представляет собой денежный мультипликатор, определяющий, на какую величину увеличивается совокупный доход при увеличении объема инвестиций на единицу.

Пример 10.6. Исследовать модифицированную модель Кейнса:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t; \\ i_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \nu_t; \\ y_t = c_t + i_t + g_t. \end{cases}$$

Здесь g_t – объем государственных расходов.

Эндогенные переменные: $c_t, i_t, y_t \quad N = 3$.

Экзогенные переменные: $g_t, y_{t-1} \quad M = 2$.

Первое уравнение содержит эндогенные переменные $c_t, i_t \quad n = 2$ и не содержит экзогенные переменные $m = 0$.

$N - n \ + \ M - m \ \geq N - 1$, то есть (3-2)+(2-0)≥3-1 и 3≥2.

Поэтому первое уравнение переопределено.

Второе уравнение содержит эндогенные переменные $i_t,\, y_t \quad n=2$ и экзогенную переменную $m=1$.

$N-n \;+\; M-m \;\geq N-1$, то есть (3-2)+(2-1)$\geq$3-1 и 2=2.

Поэтому второе уравнение идентифицируемо.

Третье уравнение содержит эндогенные переменные $c_t,\, i_t,\, y_t$, то есть $n=3$ и экзогенную переменную $g_t \quad m=1$.

$N-n \;+\; M-m \;\geq N-1$, то есть (3-3)+(2-1)$\geq$3-1 и 1$\geq$2 (ложно). Поэтому третье уравнение сверхидентифицировано.

Первое уравнение исходной системы $c_t = \beta_0 + \beta_1 y_t + \varepsilon_t$ переопределено, то есть по оценкам коэффициентов приведенных уравнений невозможно определить оценки коэффициентов β_0 и β_1.

Для переменной y_t строим приведенное уравнение:

$$y_t = \pi_{10} + \pi_{11} y_{t-1} + \pi_{12} g_t + \varpi_t,$$

где ϖ_t – случайное отклонение.

Находим с помощью МНК оценки коэффициентов $\pi_{10},\ \pi_{11},\ \pi_{12}$ и из уравнения получаем оценки $\hat{y}_t$ по экзогенным переменным g_t и y_{t-1}.

Из уравнения $c_t = \beta_0 + \beta_1 \hat{y}_t$ находим оценки коэффициентов β_0 и β_1 с помощью МНК.

Пример 10.7. Рассмотрим эконометрическую модель Клейна, составленную для описания экономики США в период с 1921 по 1941 гг.

Структурная модель содержит три регрессионных уравнения и три тождества:

$$\begin{cases} C_t = b_0 + b_1 \cdot P_t + b_2 \cdot P_{t-1} + b_3 \cdot \; W_t + W_t' \; + e_t; \\ I_t = b_0' + b_1' \cdot P_t + b_2' \cdot P_{t-1} + b_3' \cdot K_{t-1} + e_t'; \\ W_t = b_0'' + b_1'' \cdot X_t + b_2'' \cdot X_{t-1} + b_3'' \cdot \; t - t_{cp} \; + e_t''; \\ X_t = C_t + I_t + G_t; \\ P_t = X_t - W_t - T_t; \\ K_t = K_{t-1} + I_t, \end{cases}$$

где C – фонд потребления (расходы на потребительские товары);

I – капиталовложения;

W – личные доходы в частном секторе;

W' – личные доходы в государственном секторе(зарплата);

X – объем производства;

T – налоги;

K – запасы капитала;

P – прибыль;

G – государственные расходы на товары и услуги;

t – календарное время (год).

Все переменные измеряются в миллионах долларов США в ценах 1931 г.

Требуется оценить параметры модели:

$$b_0,\ b_1,\ b_2,\ b_3;\ b_0',\ b_1',\ b_2',\ b_3';\ b_0'',\ b_1'',\ b_2'',\ b_3''.$$

Исходные данные приведены в табл. 10.2. Для переменных P_t, K_t, X_t в скобках приведены данные за 1920 г., так как в модель входят эти же переменные с запаздыванием P_{t-1}, K_{t-1}, X_{t-1}.

Значение $t_{cp} = 1931$.

Сначала были оценены параметры каждого уравнения отдельно стандартным МНК:

$$C_t = 16,24 + 0,193 \cdot P_t + 0,090 \cdot P_{t-1} + 0,796 \cdot \ W_t + W_t' \ + e_t;$$

$$I_t = 10,13 + 0,480 \cdot P_t + 0,333 \cdot P_{t-1} - 0,112 \cdot K_{t-1} + e_t';$$

$$W_t = 1,50 + 0,439 \cdot X_t + 0,146 \cdot X_{t-1} + 0,130 \cdot \ t - 1931 \ + e_t''.$$

Анализ полученных моделей показал существенные проблемы. Коэффициенты регрессии при переменной P_t и P_{t-1} (прибыль) есть нормы отчислений. Получаем, что коэффициенты перед P_t значительно превышают коэффициенты перед P_{t-1}. А должно быть наоборот, так как существует некоторая временная задержка между получением прибыли и ее инвестициями. Получили смещенные значения параметров – их реальные значения в несколько раз отличаются от вычисленных МНК-оценок. Эконометрическая модель с такими смещенными параметрами – неадекватная, она неверно описывает действительное состояние экономики страны.

Таблица 10.2

Сводные экономические показатели США за 1921 – 1941 гг.

t	C	P	W	W'	I	K	X	G	T
1	2	3	4	5	6	7	8	9	10
(1920)		(12.7)				(182.8)	(44.9)		
1921	41.9	12.4	25.5	2.7	-0.2	182.6	45.6	3.9	7.7
1922	45.0	16.9	29.3	2.9	1.3	184.5	50.1	3.2	3.9
1923	49.2	18.4	34.1	2.9	5.2	189.7	57.2	2.8	4.7
1924	50.6	19.4	33.9	3.1	3.0	192.7	57.1	3.5	3.8
1925	52.6	20.1	35.4	3.2	5.1	197.8	63.0	3.3	5.5
1926	55.1	19.6	37.4	3.3	5.6	203.4	64.0	3.3	7.0
1927	56.2	19.8	37.9	3.6	4.2	207.6	64.4	4.0	6.7
1928	57.3	21.1	39.2	3.7	3.0	210.6	64.5	4.2	4.2
1929	57.8	21.7	41.3	4.0	5.1	215.7	67.0	4.1	4.0
1930	55.0	15.6	37.9	4.2	1.0	216.7	61.2	5.2	7.7
1931	50.9	11.4	34.5	4.8	-3.4	219.3	53.4	5.9	7.5
1932	45.6	7.0	29.0	5.3	-6.2	207.1	44.3	4.9	8.3
1934	48.7	12.3	30.6	6.0	-3.0	199.0	49.7	4.0	6.8
1935	51.3	14.0	33.2	6.1	-1.3	197.3	54.4	4.4	7.2
1936	57.7	17.6	36.8	7.4	2.1	199.8	62.7	2.9	8.3
1937	58.7	17.3	41.0	6.7	2.0	201.8	65.0	4.3	6.7
1938	57.5	15.3	38.2	7.7	-1.9	199.9	60.9	5.3	7.4
1939	61.6	19.0	41.6	7.8	1.3	201.2	60.5	6.6	8.9
1940	65.0	21.1	45.0	8.0	3.3	204.5	75.7	7.4	9.6
1941	69.7	23.5	53.3	8.5	4.9	209.4	88.4	13.8	11.6

Проанализировав систему регрессионных уравнений, обнаруживаем, что в данном примере нарушена первая исходная предпосылка, согласно которой случайные возмущения должны относятся только к результативным признакам и не должны коррелировать с объясняющими переменными. Но в системе одновременных уравнений получается, что совместно зависимые переменные в конечном итоге коррелируют с возмущениями того же уравнения (например, в цепочке уравнений относительно C_t, X_t, P_t следует $e_t \to C_t \to X_t \to P_t$, то есть коррелируют между собой e_t и P_t, чего не должно быть; аналогично, цепочка уравнений относительно I_t, X_t, P_t приводит к непредусмотренной корреляции между e_t' и P_t).

Для оценки параметров системы одновременных уравнений предложена процедура двухшагового регрессионного анализа.

Заменим все объясняющие зависимые переменные (в правых частях уравнений системы) на их расчетные значения. Расчетные значения, как известно, не коррелируют со случайными ошибками, поэтому соответствующая предпосылка регрессионного анализа уже не будет нарушена.

В модели имеются следующие переменные: случайные возмущения e_t, e_t', e_t''; экзогенные переменные, известные в каждый момент времени t, W_t', G_t, T_t; эндогенные переменные, на которые влияют внесистемные переменные C_t, I_t, W_t, X_t, P_t, K_t.

Всего таких переменных 6, для их определения имеется 6 уравнений, именно эти показатели необходимо назвать зависимыми объясняющими переменными, когда они входят в уравнение в качестве аргументов. Еще имеются лаговые эндогенные переменные: X_{t-1}, P_{t-1}, K_{t-1}, значения лаговых переменных известны к текущему моменту времени. Внесистемные (экзогенные) и лаговые переменные называются предопределенными, они известны к текущему моменту времени. К ним относятся t, W_t', G_t, T_t, X_{t-1}, P_{t-1}, K_{t-1}.

Для получения расчетных значений C^R, I^R, W^R составляются линейные модели относительно всех предопределенных переменных; расчетные значения X^R и P^R получаются из соответствующих тождеств:

$$C_t^R = a_0 + a_1 t + a_2 W_t' + a_2 G' + a_3 T_t + a_4 X_{t-1} + a_5 P_{t-1} + a_6 K_{t-1};$$

$$I_t^R = b_0 + b_1 t + b_2 W_t' + b_2 G' + b_3 T_t + b_4 X_{t-1} + b_5 P_{t-1} + b_6 K_{t-1};$$

$$W_t^R = c_0 + c_1 t + c_2 W_t' + c_2 G' + c_3 T_t + c_4 X_{t-1} + c_5 P_{t-1} + c_6 K_{t-1};$$

$$X_t^R = C_t^R + I_t^R + G_t;$$

$$P_t^R = X_t^R - W_t^R - T_t;$$
$$K_t^R = K_{t-1} + I_t^R.$$

Следует вычислить расчетные значения и запомнить их в виде новых переменных. На втором шаге оцениваются параметры структурной модели:

$$\begin{cases} C_t = 16{,}55 + 0{,}017 \cdot P_t^R + 0{,}216 \cdot P_{t-1} + 0{,}810 \cdot \ W_t^R + W_t' \ + e_t; \\ I_t = 20{,}28 + 0{,}150 \cdot P_t^R + 0{,}616 \cdot P_{t-1} - 0{,}158 \cdot K_{t-1} + e_t'; \\ W_t = 1{,}50 + 0{,}439 \cdot X_t^R + 0{,}147 \cdot X_{t-1} + 0{,}130 \cdot \ t - 1931 \ + e_t''. \end{cases}$$

Двухшаговым методом наименьших квадратов получены объяснимые с точки зрения экономики оценки спорных коэффициентов регрессии (сравни их с ранее найденными). Большие изменения в решении получены, в основном, за счет замены P на P^R. На рис. 10.1 показано соответствие между реальными и сглаженными по МНК данными P на P^R.

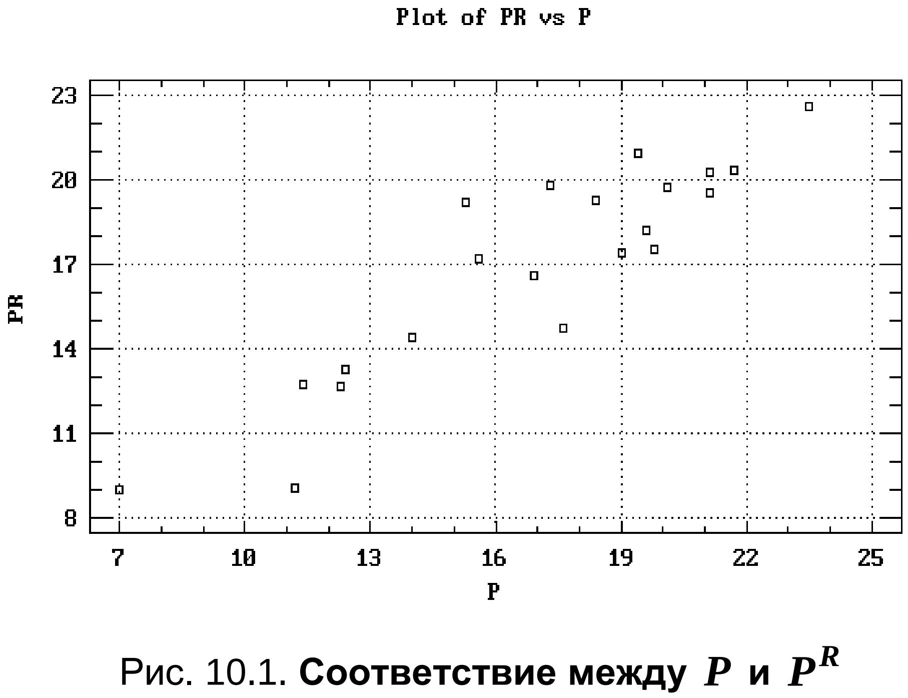


Рис. 10.1. **Соответствие между P и P^R**

Коэффициент корреляции между ними равен 0,909. Следовательно, незначительные поправки в исходных данных привели к таким большим смещениям в параметрах модели.

10.3. Задачи

Задача 10.1. Проверить следующую структурную модель на идентификацию:

$$\begin{cases} y_1 = b_{13} \cdot y_3 + a_{11} \cdot x_1 + a_{13} \cdot x_3, \\ y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + a_{22} \cdot x_2, \\ y_3 = b_{32} \cdot y_2 + a_{31} \cdot x_1 + a_{33} \cdot x_3. \end{cases}$$

Исходя из приведенной формы модели уравнений

$$\begin{cases} y_1 = 2 \cdot x_1 + 4 \cdot x_2 + 10 \cdot x_3, \\ y_2 = 3 \cdot x_1 - 6 \cdot x_2 + 2 \cdot x_3, \\ y_3 = -5 \cdot x_1 + 8 \cdot x_2 + 5 \cdot x_3. \end{cases}$$

Найти структурные коэффициенты модели.

Задача 10.2. Проверить на идентификацию модель денежного рынка:

$$\begin{cases} r_t = \beta_0 + \beta_1 y_t + \beta_2 m_t + \varepsilon_t, \\ y_t = \alpha_0 + \alpha_1 r_t + \nu_t. \end{cases}$$

В этой модели r_t, y_t, m_t – процентная ставка, ВВП и денежная масса в году t соответственно, ε_t и ν_t – случайные составляющие. Эндогенные переменные: r_t и y_t $N = 2$. Экзогенная переменная m_t $M = 1$.

Задача 10.3. Рассматривается кейнсианская модель формирования доходов:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ y_t = c_t + i_t. \end{cases}$$

Представить ее в виде системы приведенных уравнений.

10.4. Рассматривается модифицированная модель "доход-потребление":

$$
\begin{cases}
c_t = \beta_0 + \beta_1 y_t + \beta_2 c_{t-1} + \varepsilon_t, \\
i_t = \gamma_0 + \gamma_1 y_t + \nu_t, \\
y_t = c_t + i_t + g_t.
\end{cases}
$$

Указать эндогенные и экзогенные переменные, определить идентифицируемость структурных уравнений, составить приведенную систему.

Задача 10.5. Модель "спрос-предложение" содержит функции спроса, предложения:

$$
\begin{cases}
q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{t1}, \\
q_t = \beta_0 + \beta_1 p_t + \varepsilon_{t2}.
\end{cases}
$$

Здесь первое уравнение – функция спроса,

второе – функция предложения,

q_t – количество товара,

p_t – цена товара,

y_t – доход потребителей в момент времени t,

ε_{t1} и ε_{t2} – случайные составляющие.

Эндогенные переменные q_t, p_t.

Экзогенная переменная y_t.

Имеются результаты наблюдений (табл. 10.3).

Таблица 10.3

Результаты наблюдений

p	1	2	3	4	5
q	8	10	7	5	1
y	2	4	3	5	2

Найти оценки параметров этой системы уравнений с помощью косвенного метода наименьших квадратов.

Задача 10.6. Есть данные за пять лет (табл. 10.4).

Таблица 10.4

Исходные данные

Номер года	Годовое потребление свинины на душу населения, фунтов, y_1	Оптовая цена за фунт, долл., y_2	Доход на душу населения, долл., x_1	Затраты на обработку мяса, % к цене, x_2
1	60	5,0	1 300	60
2	62	4,0	1 300	56
3	65	4,2	1 500	56
4	62	5,0	1 600	63
5	66	3,8	1 800	50

Построить модель вида:

$$\begin{cases} y_1 = f\left(y_2, x_1\right), \\ y_2 = f\left(y_1, x_2\right). \end{cases}$$

Рассчитать соответствующие структурные коэффициенты.

Задача 10.7. Применив необходимое и достаточное условие идентификации, необходимо определить идентифицировано ли каждое из уравнений модели, определить метод оценки параметров модели и записать приведенную форму модели.

1. Модель денежного и товарного рынков:

$$\begin{cases} R_t = a_1 + b_{12}Y_t + b_{14}M_t + e_1, & (1) \\ Y_t = a_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + e_2, & (2) \\ I_t = a_3 + b_{31}R_t + e_3, & (3) \end{cases}$$

где уравнение (1) – функция денежного рынка,

уравнение (2) – функция товарного рынка,

уравнение (3) – функция инвестиций, а также:

R – процентные ставки;

Y – реальный ВВП;

M – денежная масса;

I – внутренние инвестиции;

G – реальные государственные расходы.

2. Макроэкономическая модель:

$$\begin{cases} C_t = a_1 + b_{11} D_t + e_{1t}, \\ I_t = a_2 + b_{22} Y_t + b_{23} Y_{t-1} + e_{2t}, \\ Y_t = D_t + T_t, \\ D_t = C_t + I_t + G_t, \end{cases}$$

где C – расходы на потребление;

Y – чистый национальный продукт;

D – чистый национальный доход;

I – инвестиции;

T – косвенные налоги;

G – государственные расходы;

t – текущий период;

$t-1$ – предыдущий период.

Задача 10.8. Рассматривается некоторая эконометрическая модель, структурная форма которой имеет вид:

$$\begin{cases} Y_1 = -4 + ???Y_2 - 9,4X_2 + e_1, \\ Y_2 = 12,83 - 2,67Y_1 + ???\,X_1 + e_2, \\ Y_3 = 1,36 - 1,76Y_1 + 0,828Y_2 + e_3. \end{cases}$$

Приведенная форма модели:

$$\begin{cases} Y_1 = 2 + 4X_1 - 3X_2 + \nu_1, \\ Y_2 = 7,5 + 5X_1 + 8X_2 + \nu_2, \\ Y_3 = 4 + ???\,X_1 + ???\,X_2 + \nu_3. \end{cases}$$

Какими методами получены параметры структурной и приведенной форм модели? Обоснуйте возможность применения косвенного МНК для расчета структурных параметров модели. Восстановите пропущенные характеристики.

Задача 10.9. Рассматривается модель потребления мяса на душу населения в некотором регионе:

$$\begin{cases} y_1 = a_0 + a_1 y_2 + a_2 x_1 + a_3 x_2 + e_1, \\ y_2 = b_0 + b_1 y_1 + b_2 x_3 + e_2, \end{cases}$$

где y_1 – годовое потребление мяса на душу населения, кг;

y_2 – цена за 1 кг мяса, грн;

x_1 – доход на душу населения, тыс. грн;

x_2 – годовое потребление рыбы на душу населения, кг;

x_3 – цена за 1 кг рыбы, грн.

Приведенная форма модели имеет вид:

$$\begin{cases} y_1 = 45 + 4x_1 - 1{,}2x_2 + 0{,}5x_3 + u_1, \\ y_2 = -52 + 3x_1 + 2x_2 + 0{,}8x_3 + u_2. \end{cases}$$

Необходимо:

1. Провести идентификацию модели.
2. Указать способ оценки параметров каждого уравнения структурной модели.
3. Найти структурные коэффициенты для одного из уравнений системы, используя косвенный МНК.

Задача 10.10. Для прогнозирования спроса на свою продукцию предприятие использует следующую модель, характеризующую общую экономическую ситуацию в регионе:

$$\begin{cases} Q_t = a_1 + b_{11}Y_t + e_{1t}, \\ C_t = a_2 + b_{21}Y_t + e_{2t}, \\ I_t = a_3 + b_{32} \quad Y_{t-1} - K_{t-1} \quad + e_{3t}, \\ Y_t = C_t + I_t, \end{cases}$$

где Q – реализованная продукция за период времени t;

Y – валовая добавленная стоимость региона;

C – конечное потребление;

I – инвестиции;

K – запас капитала;

t – текущий период;

$t-1$ – предыдущий период.

Необходимо:

1. Провести идентификацию модели, определив, идентифицировано ли каждое из уравнений модели.
2. Определить метод оценки параметров модели.
3. Записать приведенную форму модели.

11. Динамические эконометрические модели

11.1. Основные теоретические сведения

Модели, связывающие состояния экономических явлений в последовательные моменты (периоды) времени, принято называть динамическими. Такие модели позволяют изучать явления в динамике, в развитии. Аналитическое представление динамических моделей включает значения переменных, относящиеся как к текущему, так и к предыдущим моментам (периодам) времени.

Выделяют два типа динамических эконометрических моделей. К моделям первого типа относятся модели авторегрессии и модели с распределенным лагом, в которых значения переменной за прошлые периоды времени (лаговые переменные) непосредственно включены в модель. Лаговые модели с распределенным лагом имеют вид:

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \ldots + b_p \cdot x_{t-p} + \varepsilon_t .$$

Коэффициент регрессии a характеризует среднее абсолютное изменение в некоторый фиксированный период времени t, без учета воздействия лаговых значений фактора x. Этот коэффициент называют краткосрочным мультипликатором.

Любую сумму коэффициентов $\sum\limits_{j=0}^{h} b_j, \quad h < p$ называют промежуточным мультипликатором: в момент $t+1$ совокупное воздействие факторной переменной x_t на результат y_t составит $b_0 + b_1$; в момент времени $t+2$ соответственно $b_0 + b_1 + b_2$ и т. д.

Сумму всех b_j называют долгосрочным мультипликатором $b_0 + b_1 + b_2 + ... + b_l = b$ и он характеризует изменение y под воздействием единичного изменения x в каждом из рассматриваемых периодов времени. Рассчитываются относительные коэффициенты модели с распределенным лагом: $\beta_j = \dfrac{b_j}{b}$. Если все коэффициенты b_j имеют одинаковые знаки, то для любого i-го значения $0 < \beta_j < 1; \quad \sum\limits_{j=0}^{l} \beta_j = 1.$

Относительные коэффициенты β_j являются весами для соответствующих коэффициентов b_j, каждый из них измеряет долю общего изменения результативного признака в момент времени $t+j$. Зная величину β_j можно определить средний лаг и медианный лаг.

Средний лаг: $\bar{l} = \sum_{j=0}^{l} j \cdot \beta_j$ представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора в момент времени t. Небольшая величина среднего лага свидетельствует об относительно быстром реагировании результатов на изменение фактора, а большая величина – о том, что воздействие фактора на результат состоится через длительный период времени.

Медианный лаг – это величина лага, для которого $\sum_{j=0}^{l_{Me}} \beta_j \approx 0{,}5$. Это период, в течение которого с момента времени t будет реализована половина общего воздействия фактора на результат.

Модели второго типа учитывают динамическую информацию в неявном виде. В эти модели включены переменные, характеризующие ожидаемый или желаемый уровень результата, или одного из факторов в момент времени t. Модели авторегрессии имеют вид:

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + c_2 \cdot y_{t-2} + \ldots + c_q \cdot y_{t-q} + \varepsilon_t.$$

В модели авторегрессии b_0 характеризует краткосрочное изменение y_t под воздействием изменения x_t на 1 ед. К моменту времени $t+1$ результат y_t изменится под воздействием изменения фактора в момент времени t на b_0, а y_{t-1} под воздействием своего изменения в непосредственно предшествующий момент времени на c_1. Абсолютное изменение результата в момент времени $t+1$ составит $b_0 c_1$, а в момент времени $t+2$ абсолютное изменение составит $b_0 c_1^2$ и т. д.

Долгосрочный мультипликатор в модели авторегрессии рассчитывается как сумма краткосрочного и промежуточного мультипликатора:

$$b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \dots .$$

Учитывая $|c_1| < 1$, имеем $b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \ldots = \dfrac{b_0}{1 - c_1}$.

Такая интерпретация коэффициентов модели авторегрессии и расчет долгосрочного мультипликатора основаны на предпосылке о наличии бесконечного лага в воздействии текущего значения зависимой переменной на ее будущие значения.

Важным этапом при построении моделей с распределенным лагом и моделей авторегрессии является выбор оптимальной величины лага и определение его структуры.

Коэффициенты модели с конечным числом лагов оцениваются с помощью сведения ее к уравнению множественной регрессии. Оценка модели с распределенным лагом зависит от количества лагов.

Для оценки моделей с бесконечным числом лагов разработаны специальные методы, а именно метод последовательного увеличения количества лагов, метод Алмон, преобразование Койка.

Метод Ш. Алмон

Рассмотрим дистрибутивно-лаговую модель с конечным лагом в k периодов:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \ldots + b_k x_{t-k} + \varepsilon_t .$$

В основе модели Алмон лежит предположение о том, что весовые коэффициенты этой зависимости подчиняются полиноминальному закону распределения и они могут быть выражены функциями от продолжительности лага:

$$b_i = a_0 + a_1 i + a_2 i^2 + \ldots + a_m i^m, \ i = \overline{0,k} .$$

Причем, предполагают, что старшая степень полинома m меньшая, чем максимальная длина лага k: $m < k$.

Зависимость коэффициентов b_i от величины лага i лучше всего выбирать по внешнему виду графика зависимости величины параметра от лага.

Приведем простые примеры об ожидаемом виде распределения весовых коэффициентов (рис. 11.1).

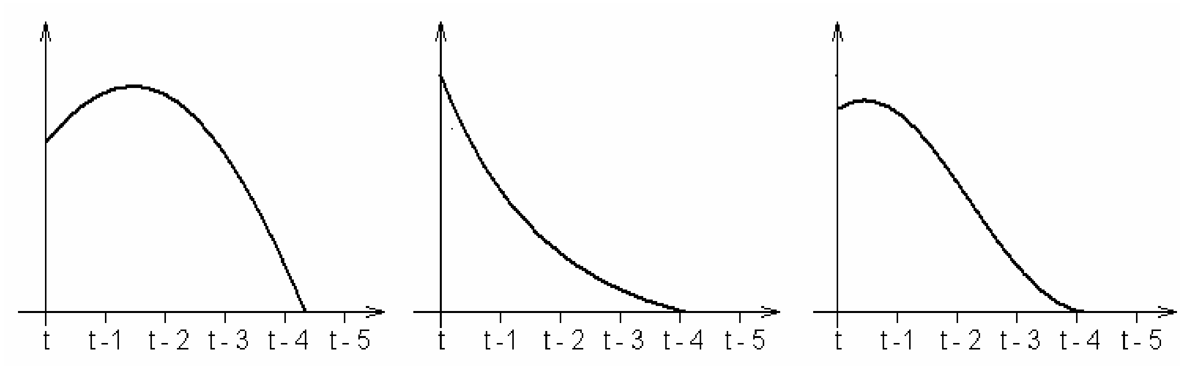


Рис. 11.1. **Априорные представления о распределении весовых коэффициентов лагов**

Алгоритм применения метода Алмон

1. Определяем продолжительность лага.
2. Определяем степень полинома и проверяем значимость коэффициентов в моделях.
3. Вычисляем значение новых переменных z_i, вычисляем параметры модели относительно новых переменных и возвращаемся к начальной модели с помощью формул перехода.

Применение метода Алмон для расчета параметров модели с распределенным лагом предполагает предварительное определение максимальной величины лага и степени полинома. Оптимальную величину можно приближенно определить на основе априорной информации экономической теории или проведенных ранее исследований. Приближенно в качестве величины лага можно взять значение максимального лага, для которого парный коэффициент корреляции между y и лаговыми переменными x_t, x_{t-1}, $x_{t-2,\dots}$ остается значимым. Также можно построить несколько уравнений с разной величиной лага и выбрать среди них лучшее. В качестве степени полинома на практике обычно выбирают $k=2$ или $k=3$. Величину k также можно определять путем сравнения моделей, построенных для различных значений k.

Метод Ш. Алмон достаточно универсален и может быть применен для моделирования процессов, которые характеризуются разнообразными структурами лагов и при относительно небольшом количестве переменных дает возможность построить модели с распределенным лагом любой длины. Однако, метод Алмон используется для построения модели с конечной величиной лага l. Для оценки же параметров модели с бесконечной величиной лага, а именно:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$$

рекомендуется применять метод Койка.

Метод Койка

На основании того, что существует геометрическая структура лага, то есть такая структура, когда влияние лаговых значений фактора на результат уменьшаются с увеличением величины лага в геометрической прогрессии принимается предположение о том, что существует некоторый постоянный темп λ $0 < \lambda < 1$ уменьшения во времени лагового влияния фактора на результат. Если, например, в период t результат менялся под влиянием изменения фактора в этот же период времени на b_0 единиц, то под влиянием изменения фактора в период $t-1$ результат изменится на $b_1 = b_0 \cdot \lambda$ единиц, в период $t-2$ на $b_1 = b_0 \cdot \lambda \cdot \lambda = b_0 \cdot \lambda^2$ единиц и так далее. Для некоторого периода $t-l$ это изменение составит $b_0 \cdot \lambda^{l}$ единиц. Бесконечная последовательность членов считается сходящейся, то есть коэффициенты b_j должны убывать достаточно быстро, начиная с некоторого номера. Койк предусмотрел убывание по геометрической прогрессии.

В общем виде имеем:

$$b_j = b_0 \cdot \lambda^j; \ j=0,\ 1,\ 2,\ \dots \quad 0 < \lambda < 1 \ . \qquad (11.1)$$

Ограничение на значение $\lambda > 0$ обеспечивает одинаковые знаки для всех коэффициентов $b_j > 0$, а ограничение $\lambda < 1$ означает, что с увеличением лага значения параметров модели

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$$

уменьшаются в геометрической прогрессии. Чем ближе λ к 0, тем больший темп снижения влияния фактора на результат во времени и тем большее частичное влияния на результат имеет текущее значение фактора x_t. При подстановке $b_j = b_0 \cdot \lambda^j$ имеем:

$$y_t = a + b_0 x_t + b_0 \cdot \lambda \cdot x_{t-1} + b_0 \cdot \lambda^2 \cdot x_{t-2} + ... + \varepsilon_t. \qquad (11.2)$$

Тогда для периода $t-1$ модель имеет такой вид:

$$y_{t-1} = a + b_0 x_{t-1} + b_0 \cdot \lambda \cdot x_{t-2} + b_0 \cdot \lambda^2 \cdot x_{t-3} + ... + \varepsilon_t. \qquad (11.3)$$

Умножим обе части уравнения на λ и получим:

$$\lambda \cdot y_{t-1} = \lambda \cdot a + b_0 \cdot \lambda \cdot x_{t-1} + b_0 \cdot \lambda^2 \cdot x_{t-2} + b_0 \cdot \lambda^3 \cdot x_{t-3} + ... + \lambda \cdot \varepsilon_{t-1}. \quad (11.4)$$

Вычтем соотношение (11.4) из соотношения (11.3):

$$y_t - \lambda \cdot y_{t-1} = a - \lambda \cdot a + b_0 \cdot x_t + \varepsilon_t - \lambda \cdot \varepsilon_{t-1} \qquad (11.5)$$

Преобразование выражения (11.5) приводит к модели Койка:

$$y_t = a \cdot (1 - \lambda) + b_0 \cdot x_t + \lambda \cdot y_{t-1} + u_t, \qquad (11.6)$$

где $u_t = \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$.

Таким образом, имеем модель авторегрессии, определив параметры которой найдем λ и b_0 начальной модели.

Далее с помощью соотношения (11.1) определяем параметры $b_1, b_2, ...$ модели $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$.

Следует отметить, что применение обычного МНК для оценки параметров модели (11.6) приводит к получению смещенных оценок ее параметров, что объясняется наличием в качестве фактора лагового результативного признака y_{t-1}.

Приведенная схема вычисления является преобразованием Койка, которое позволяет перейти от модели с бесконечными распределенными лагами к модели авторегрессии, которая содержит две независимых переменных x и y_{t-1}. Геометрическая структура лага позволяет определить значения среднего и медианного лагов в модели Койка. Учитывая, что сумма коэффициентов регрессии в модели

$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$ есть сумма геометрической прогрессии, то есть

$$\sum_{j=0}^{\infty} b_j = b_0 + b_0 \cdot \lambda + b_0 \cdot \lambda^2 + b_0 \cdot \lambda^3 + ... = b_0 \cdot \left(1 + \lambda + \lambda^2 + \lambda^3 + ...\right) = b_0 \cdot \frac{1}{1-\lambda},$$

то средний лаг определяется как:

$$T = \bar{l} = \frac{\sum_{j=0}^{\infty} j \cdot b_j}{\sum_{j=0}^{\infty} b_j} = \frac{b_0 \cdot \lambda \cdot \left(1 + 2 \cdot \lambda^2 + 3 \cdot \lambda^3 + ...\right)}{b_0 \cdot \frac{1}{1-\lambda}} = \frac{b_0 \cdot \lambda \cdot \frac{1}{(1-\lambda)^2}}{b_0 \cdot \frac{1}{1-\lambda}} = \frac{\lambda}{1-\lambda}.$$

Откуда при $\lambda = 0,5$ средний лаг $T = 1$, а при $l < 0,5$ средний лаг $T < 1$, то есть влияние фактора на результат в среднем осуществляется меньше одного периода. Величину $1-\lambda$ интерпретируют как скорость, с которой происходит адаптация результата во времени к изменению факторного признака. Медианный лаг вычисляется по помощи:

$$\sum_{j=0}^{l_{Me}-1} \beta_j = \sum_{j=0}^{l_{Me}-1} \frac{b_j}{\sum_{j=0}^{\infty} b_j} = \sum_{j=0}^{l_{Me}-1} \frac{b_j \cdot \lambda^j}{b_0 \cdot \frac{1}{1-\lambda}} = \sum_{j=0}^{l_{Me}-1} \lambda^j \cdot \left(1-\lambda\right) = 0,5.$$

Поэтому медианный лаг в модели Койка равен $l_{Me} = \frac{\ln 0,5}{\ln \lambda}$.

11.2. Примеры

Пример 11.1. В табл. 11.1 приведены объемы реализации (Y) и цены за единицу продукции X завода "Укрэлектромаш" (г. Харьков) по 24 периодам.

Таблица 11.1

Объемы реализации (Y) и цены за единицу продукции (X) "Укрелектромаш" (г. Харьков)

Период	Y_t	X_t	X_{t-1}	X_{t-2}	X_{t-3}	X_{t-4}	Z_0	Z_1	Z_2	Z_3
1	2	3	4	5	6	7	8	9	10	11
1	2 141	187,74								
2	2 934	180,63	187,74							

Окончание табл. 11.1

1	2	3	4	5	6	7	8	9	10	11
3	3 371	173,79	180,63	187,74						
4	5 448	158,34	173,79	180,63	187,74					
5	7 200	150,66	158,34	173,79	180,63	187,74	851,16	1 798,77	5 483,01	18 441,03
6	5 497	157,82	150,66	158,34	173,79	180,63	1 008,98	2 649,93	9 931,71	41 137,53
7	5 410	159,35	157,82	150,66	158,34	173,79	1 168,33	3 658,91	16 240,55	79 891,43
8	4 089	171,39	159,35	157,82	150,66	158,34	1 151,98	3 513,06	15 527,44	76 363,32
9	1 492	200,87	171,39	159,35	157,82	150,66	1 172,22	3 400,63	14 854,67	72 680,71
10	9 34	199,53	200,87	171,39	159,35	157,82	1 197,96	3 356,32	14 312,44	69 008,86
11	9 07	198,51	199,53	200,87	171,39	159,35	1 238,13	3 445,9	14 464,38	68 902,48
12	2 630	183,20	198,51	199,53	200,87	171,39	1 270,67	3 629,41	15 211,97	72 195,07
13	3 033	179,06	183,20	198,51	199,53	200,87	1 291,91	3 795,34	16 008,28	75 857,62
14	4 192	170,12	179,06	183,20	198,51	199,53	1 302,68	3 971,8	17 082,72	81 903,34
15	4 226	164,34	170,12	179,06	183,20	198,51	1 295,63	4 074,75	17 930,89	87 582,81
16	5 047	160,63	164,34	170,12	179,06	183,20	1 255,39	3 964,29	17 533,39	85 996,95
17	1 067	191,39	160,63	164,34	170,12	179,06	1 247,25	3 822,97	16 940,39	83 306,59
18	1 008	196,10	191,39	160,63	164,34	170,12	1 244,84	3 680,65	16 106,59	78 754,99
19	9 572	151,75	196,10	191,39	160,63	164,34	1 213,39	3 643,09	15 735,93	76 523,95
20	4 259	167,87	151,75	196,10	191,39	160,63	1 202,2	3 603,06	15 461,56	74 456,82
21	1 490	201,98	167,87	151,75	196,10	191,39	1 234,06	3 614,42	15 534	74 501,72
22	446	200,32	201,98	167,87	151,75	196,10	1 270,04	3 698,1	15 944,24	76 812,42
23	560	208,51	200,32	201,98	167,87	151,75	1 317,92	3 843,73	16 739,61	81 913,39
24	1 256	205,55	208,51	200,32	201,98	167,87	1 332,08	3 821,92	16 366,88	79 334,56

Составим дистрибутивно-лаговую модель

$$y_t = a + b_0 x_t + b_1 x_{t-1} + ... + b_k x_{t-k},$$

где $k \le 6$, так как максимальный лаг k не может быть большим 25 % от числа периодов.

Для иллюстрации обычным МНК были вычислены параметры моделей для $k = 4, 5, 6$ вместе со всеми их статистическими характеристиками.

На рис. 11.2 приведены графики поведения параметров модели b_i в зависимости от принятого максимального лага k.

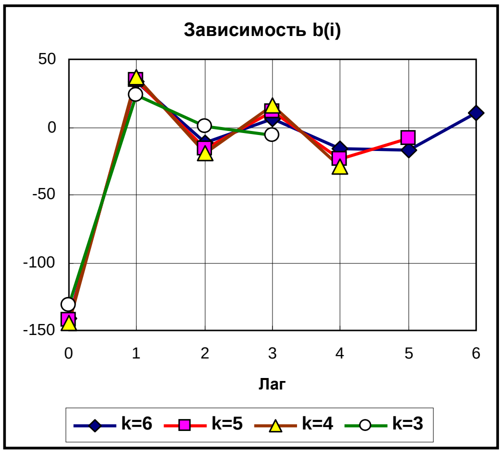


Рис. 11.2. **Зависимость параметров модели от лага**

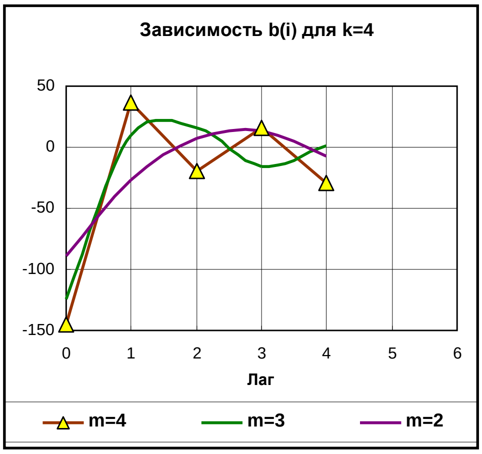


Рис. 11.3. **Качество аппроксимации параметров модели**

Как видим, в этих зависимостях нет ничего общего с априорными представлениями об ожидаемых распределениях, которые изображены на рис. 11.1.

Методом Алмон были вычислены все комбинации моделей для $k = 4,5,6$ и $2 \le m \le k$.

Во всех комбинациях, где $m = k$ МНК были получены обычные оценки параметров, правда, без статистических оценок их значимости.

Для примера рассмотрим результаты вычислений при $k = 4$ и $m = 2, 3, 4$ (рис. 11.3). Это наиболее распространенный вариант расчетов по методу Ш. Алмон.

В табл. 1 приведены значения лаговых переменных, на основании которых рассчитаны значения новых переменных z_0, z_1, z_2, z_3.

Потеряно $k=4$ наблюдений (17,7 %). Значения z_j очень быстро возрастают с номером *j*. Между этими переменными есть очень высокая корреляция, например, $r_{z_2 z_3}=0,999$.

На рис. 11. 3 показано качество аппроксимации зависимости b_i полиномами 2-й и 3-й степени (полином 4-й степени точно отображает 5 заданных точек). Как видим, аппроксимация не очень удачная. Для сравнения приведем уравнения регрессий для $k=4$ и $m=4, 3, 2$.

$$m=4: y_t = 28537 - 144,7x_t + 36,4x_{t-1} - 19,3x_{t-2} + 16,3x_{t-3} - 29,1x_{t-4};$$
$$m=3: y_t = 23887 - 124,9x_t + 9,5x_{t-1} + 16,3x_{t-2} - 16,1x_{t-3} + 0,9x_{t-4};$$
$$m=2: y_t = 21863 - 89,6x_t - 27,1x_{t-1} + 7,4x_{t-2} + 13,8x_{t-3} - 7,9x_{t-4}.$$

В этих моделях не совпадают даже знаки параметров.

Это еще раз подтверждает преувеличенное значение метода Алмон. С вычислительной точки зрения никаких преимуществ не имеем, поскольку полиномы Алмон также тесно коррелируют между собой, как и изначальные лаговые переменные.

Пример 11.2. Использовав данные экономики Канады за 1961 – 1981 гг., ученые Джеффри Сакс и Майкл Бруно построили модель:

$$U_t = \delta_0 + \delta_1 \cdot U_{t-1} + \delta_2 \cdot t + \delta_3 \cdot w_t + \varepsilon_t,$$

где t – время;

U_t, U_{t-1} – уровень безработицы в периоды t и $t-1$;

w_t – превышение реальной заработной платы сравнительно с ее уровнем в условиях полной занятости (значения, полученные с помощью вычислений);

$\delta_0, \delta_1, \delta_2, \delta_3$ – параметры модели;

ε_t – погрешность;

$$U_t = \delta_0 + 0,63 \cdot U_{t-1} + 0,07 \cdot t + 15,72 \cdot w_t,$$
$$R^2 = 0,85,\ t = 5,46,\ t = 2,82,\ t = 2,23.$$

Переменная w_t в модели определяет спрос на работу. Если предусмотреть, что переменная w_t осуществляет влияние на уровень безработицы с бесконечным временным лагом в условиях геометрической структуры лага, то согласно методу Койка получаем такую модель с распределенным лагом:

$$U_t = a + b_0 w_t + b_0 \cdot \lambda \cdot w_{t-1} + b_0 \cdot \lambda^2 \cdot w_{t-2} + ... + c \cdot t + \varepsilon_t .$$

Эта модель отличается от $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$ тем, что кроме текущего и лагового значений факторного признака, она учитывает фактор времени t. Соответствующие алгебраические преобразования соответственно методу Койка приводят к модели авторегрессии вида:

$$U_t = \left(a \cdot (1-\lambda) + \lambda \cdot c \right) + (1-\lambda) \cdot U_{t-1} + c \cdot (1-\lambda) \cdot t + b_0 w_t + u_t ,$$

то есть $\delta_0 = a \cdot (1-\lambda) + \lambda \cdot c;$ $\delta_1 = 1-\lambda;$ $\delta_2 = c \cdot (1-\lambda)$; $\delta_3 = b_0$.

В модели Джеффри Сакса и Майкла Бруно $\lambda = 0,63$.

При этом параметры модели Койка:

$$c = \frac{0,07}{1-0,63} = 0,189; \quad a = \frac{\delta_0}{1-0,63} + 0,189 \cdot 0,63 = 0,119 + 2,703 \cdot \delta_0 ;$$

$$b_0 = 15,72; \quad b_1 = 15,72 \cdot 0,63 = 9,904; \quad b_2 = 15,72 \cdot (0,63)^2 = 6,239 ;$$

$$b_3 = 15,72 \cdot (0,63)^3 = 3,931$$ и так далее.

Модель Койка приобретает вид:

$$U_t = (0,119 + 2,703 \cdot \delta_0) + 15,72 \cdot w_t + 9,904 \cdot w_{t-1} + 6,239 \cdot w_{t-2} + 3,931 \cdot w_{t-3} + \\ + ... + 0,189t.$$

Средний лаг равен $T = \frac{0,63}{1-0,63} = 1,703;$

величина медианного лага: $l_{Me} = \frac{\ln 0,5}{\ln 0,63} = \frac{-0,69314}{-0,46203} = 1,500203$.

Отсюда можно сделать выводы, что в среднем влияние отклонений между реальной заработной платой в экономике Канады и ее величиной в условиях полной занятости проявляется на протяжении относительно короткого промежутка времени, это 1,7 года, при этом половина влияния реализуется в течении первых 1,5 лет с момента изменения w_t.

Пример 11.3. Вычислить модель с распределенными лагами для описания затрат на жилье (y) в зависимости от уровня доходов (x) и относительных цен текущего (p) и нескольких предыдущих периодов (табл. 11.2); преобразовать эту модель в авторегрессионную методом Койка; оценить ее параметры и сделать выводы относительно кратко- и долгосрочного влияния объясняющих переменных.

Для демонстрации были вычислены модели, где x_{t-1}, p_{t-1}, x_{t-2}, p_{t-2} лаговые переменные, учитывающие сдвиг в один и два периода:

1) $\ln y_p = -1,51 + 1,81 \ln x - 0,35 \ln p; R^2 = 0,991;$

$$t_{b_0} = 0,8; t_{b_x} = 21; t_{b_p} = 1,1;$$

2) $\ln y_p = 0,86 + 1,09 \ln x_{t-1} - 0,73 \ln p_{t-1};\ R^2 = 0,993;$

$$t_{b_0} = 0,5; \quad t_{b_{xt-1}} = 19,4; \quad t_{b_{pt-1}} = 2,2;$$

3. $\ln y_p = 2,31 + 1,02 \ln x_{t-2} - 0,95 \ln p_{t-2};\ R^2 = 0,996;$

$$t_{b_0} = 1,4; \quad t_{b_{xt-2}} = 20,6; \quad t_{b_{pt-2}} = 3,3.$$

Таблица 11.2

Значения переменных

Год	y_t	x_t	p_t	$\ln y_t$	$\ln y_{t-1}$	$\ln x_t$	$\ln p_{t-1}$	$\ln x_{t-1}$	$\ln p_{t-1}$	$\ln x_{t-2}$	$\ln p_{t-2}$
1	2	3	4	5	6	7	8	9	10	11	12
1960	64,0	489,7	104,5	4,1589	4,1092	6,1938	4,6492	6,1732	4,6492	6,1327	4,6501
1961	67,0	503,8	105,1	4,2047	4,1589	6,2222	4,6549	6,1938	4,6492	6,1732	4,6492
1962	70,7	524,9	105,0	4,2584	4,2047	6,2632	4,6540	6,2222	4,6549	6,1938	4,6492
1963	74,0	542,3	104,8	4,3041	4,2584	6,2958	4,6521	6,2632	4,6540	6,2222	4,6549
1964	77,4	580,8	104,5	4,3490	4,3041	6,3644	4,6492	6,2958	4,6521	6,2632	4,6540

Окончание табл. 11.2

1	2	3	4	5	6	7	8	9	10	11	12
1965	81,6	616,3	104,0	4,4018	4,3490	6,4237	4,6444	6,3644	4,6492	6,2958	4,6521
1966	85,3	646,8	102,6	4,4462	4,4018	6,4720	4,6308	6,4237	4,6444	6,3644	4,6492
1967	89,1	673,5	102,2	4,4898	4,4462	6,5125	4,6269	6,4720	4,6308	6,4237	4,6444
1968	93,5	701,3	100,9	4,5380	4,4898	6,5529	4,6141	6,5125	4,6269	6,4720	4,6308
1969	98,4	722,5	100,0	4,5890	4,5380	6,5827	4,6052	6,5529	4,6141	6,5125	4,6269
1970	102,0	751,6	99,6	4,6250	4,5890	6,6222	4,6012	6,5827	4,6052	6,5529	4,6141
1971	106,4	779,2	100,0	4,6672	4,6250	6,6583	4,6052	6,6222	4,6012	6,5827	4,6052
1972	112,5	810,3	100,0	4,7230	4,6672	6,6974	4,6052	6,6583	4,6052	6,6222	4,6012
1973	118,2	865,3	99,1	4,7724	4,7230	6,7631	4,5961	6,6974	4,6052	6,6583	4,6052
1974	124,2	858,4	95,1	4,8219	4,7724	6,7551	4,5549	6,7631	4,5961	6,6974	4,6052
1975	128,3	875,8	93,3	4,8544	4,8219	6,7751	4,5358	6,7551	4,5549	6,7631	4,5961
1976	134,9	906,8	93,7	4,9045	4,8544	6,8099	4,5401	6,7751	4,5358	6,7551	4,5549
1977	141,3	942,9	94,5	4,9509	4,9045	6,8490	4,5486	6,8099	4,5401	6,7751	4,5358
1978	148,5	988,8	94,7	5,0006	4,9509	6,8965	4,5507	6,8490	4,5486	6,8099	4,5401
1979	154,8	1015,5	93,8	5,0421	5,0006	6,9231	4,5412	6,8965	4,5507	6,8490	4,5486
1980	159,8	1021,6	93,0	5,0739	5,0421	6,9291	4,5326	6,9231	4,5412	6,8965	4,5507
1981	164,8	1049,3	94,2	5,1047	5,0739	6,9559	4,5454	6,9291	4,5326	6,9231	4,5412
1982	167,5	1058,3	96,7	5,1210	5,1047	6,9644	4,5716	6,9559	4,5454	6,9291	4,5326
1983	171,3	1095,4	99,2	5,1434	5,1210	6,9989	4,5971	6,9644	4,5716	6,9559	4,5454

Таким образом, вычисленные модели почти не имеют отличий: значения эластичностей затрат по доходам почти одинаковые; все уравнения значимые в целом и имеют почти одинаковые коэффициенты детерминации.

Поэтому нельзя отдать предпочтение какой-то одной модели. Это не случайно, поскольку лаговые переменные обычно тесно коррелируют, а коэффициенты регрессии в любом из этих уравнений фактически отображают суммарные эффекты по нескольким временным периодам. Правильная модель должна учитывать эффекты каждого периода отдельно, то есть необходимо разработать модель с распределенным лагом:

4) $\underset{t_b}{\ln y_p} = \underset{0,8}{-1,51} + \underset{21,0}{1,81} \ln x_t - \underset{1,1}{0,35} \ln p_t$;

$$t_{b_0} = 0,8;\ t_{b_x} = 21;\ t_{b_p} = 1,1$$

5) $\ln y_p = 0,17 + 0,22 \ln x_t + 1.00 \ln p_t + 0,88 \ln x_{t-1} - 1,60 \ln p_{t-1}$;

$$t_{b_0} = 0,1; \quad t_{b_{\ln x}} = 0,8; \quad t_{b_{\ln p}} = 2,8; \quad t_{b_{\ln xt-1}} = 2,9; \quad t_{b_{\ln pt-1}} = 4,0;$$

6) $\ln y_p = 2,14 + 0,31 \ln x_t - 0,16 \ln p_t + 0,18 \ln x_{t-1} - 0,65 \ln p_{t-1} +$

$$+0,55 \ln x_{t-2} - 1,44 \ln p_{t-2};$$

$$t_{b_0} = 1,2; \quad t_{b_{\ln x}} = 1,2; \quad t_{b_{\ln p}} = 0,3; \quad t_{b_{\ln xt-1}} = 0,4; \quad t_{b_{\ln pt-1}} = 0,7;$$

$$t_{b_{\ln xt-2}} = 1,6; \quad t_{b_{\ln pt-2}} = 2,5.$$

Все три модели значимые в целом и имеют почти одинаковые коэффициенты детерминации $R^2 = 0,991 \div 0,997$. Однако с увеличением членов коэффициенты регрессии становятся все менее значимыми и все более неустойчивыми. Это объясняется мультиколлинеарностью – тесными корреляционными связями между лаговыми переменными.

С помощью преобразований Л. Койка получим модель:

$$\ln y_p = 0,50 + 0,15 \ln x_t - 0,16 \ln p_t + 0,845 \ln y_{t-1}; \ R^2 = 0,9996$$

$$t_{b_0} = 1,3; \quad t_{b_{\ln x}} = 3,1; \quad t_{b_{\ln p}} = 2,4; \quad t_{b_{\ln yt-1}} = 22,4.$$

С помощью данной модели можно оценить кратко- и долгосрочные эффекты.

В краткосрочном аспекте (для текущего периода) значения y_{t-1} необходимо рассматривать как фиксированное, тогда эластичность затрат по доходу и цены будут равняться коэффициентам регрессии: $b = 0{,}15;\ c = -0{,}16$.

В долгосрочном аспекте, когда x, p, y приравниваются к своим весовым значениям (тогда $y_{t-1} = y_t$), оказывается, что долгосрочное влияние x на y равняется величине:

$$\frac{b}{1-q} = \frac{0{,}15}{1-0{,}845} = 0{,}96,$$

а долгосрочное влияние p на y равняется:

$$\frac{c}{1-q} = \frac{-0{,}16}{1-0{,}845} = -1{,}02.$$

Эти числа близки к значимым коэффициентам регрессии модели (11.1) – (11.3).

Преобразование Койка основывается на нескольких отягощающих предположениях, что коэффициенты регрессии при лаговых объясняющих переменных экспоненциально убывают, начиная с первого члена. Иногда следует предусмотреть, что изменение зависимой переменной в ответ на изменение объясняющей переменной сначала небольшое, а потом возрастает со временем и только потом начинает уменьшаться. Чтобы учесть этот возможный эффект, примем гипотезу об экспоненциальном убывании весовых коэффициентов не с первого, а со второго члена. Тогда преобразование Койка приведет к модели:

$$\ln y = a + \ b\ln x + c\ln p \ + \ b_0 \ln x_t + c_0 \ln p_t \ + q y_t + u,$$

где $b_0 = b_1 - bq;\ c_0 = c_1 - cq$.

Весовые коэффициенты для цен прошлых лет вычисляются аналогично.

Параметры b_0, c_0 уточняют поведение первого члена b_1, c_1.

Пример 11.4. Имеются данные за 24 квартала об объеме продукции (млн грн) y и инвестициях в основной капитал (млн грн) x, представленные в табл. 11.3.

Таблица 11.3

Исходные данные

y	x	y	x
3,5	1,51	4,5	1,64
3,6	1,5	4,6	1,69
3,6	1,5	4,7	1,74
3,7	1,53	4,9	1,8
3,7	1,53	4,8	1,75
3,8	1,55	4,8	1,65
3,9	1,58	5	1,73
4,1	1,62	5,1	1,81
4,2	1,65	5,3	1,87
4,3	1,63	5,4	1,88
4,4	1,65	5,4	1,8
4,5	1,67	5,4	1,84

Необходимо построить модель с распределенными лагами вида

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3}$$

двумя способами:

1. Применить к исходным данным обычный МНК и сделать выводы.
2. Использовать лаги Алмон, предполагая структуру лага линейной.

Сравнить оба варианта.

Решение

1. Составим расчетную таблицу (табл. 11.4).

Таблица 11.4

Расчетная таблица

№ п/п	y_t	x_t	x_{t-1}	x_{t-2}	x_{t-3}
1	2	3	4	5	6
1	3,5	1,51	–	–	–
2	3,6	1,5	1,51	–	–
3	3,6	1,5	1,5	1,51	–
4	3,7	1,53	1,5	1,5	1,51
5	3,7	1,53	1,53	1,5	1,5
6	3,8	1,55	1,53	1,53	1,5

Окончание табл. 11.4

1	2	3	4	5	6
7	3,9	1,58	1,55	1,53	1,53
8	4,1	1,62	1,58	1,55	1,53
9	4,2	1,65	1,62	1,58	1,55
10	4,3	1,63	1,65	1,62	1,58
11	4,4	1,65	1,63	1,65	1,62
12	4,5	1,67	1,665	1,63	1,65
13	4,5	1,64	1,67	1,665	1,63
14	4,6	1,69	1,64	1,67	1,665
15	4,7	1,74	1,69	1,64	1,67
16	4,9	1,8	1,74	1,69	1,64
17	4,8	1,75	1,8	1,74	1,69
18	4,8	1,65	1,75	1,8	1,74
19	5	1,73	1,65	1,75	1,8
20	5,1	1,81	1,73	1,65	1,75
21	5,3	1,87	1,81	1,73	1,65
22	5,4	1,88	1,87	1,81	1,73
23	5,4	1,8	1,88	1,87	1,81
24	5,4	1,84	1,8	1,88	1,87

По наблюдениям $4-24$ строим модель, используя МНК, то есть используя в каждом столбце одинаковое число наблюдений – 21.

Результаты регрессии оказались следующими:

$$\hat{y}_t = -4,493 + 2,703x_t + 0,585x_{t-1} + 0,757x_{t-2} + 1,373x_{t-3}$$

$$t \qquad (-18,3) \quad (7,4) \qquad (1,1) \qquad\qquad (1,3) \qquad\qquad (3,2)$$

$$R^2 = 0,989 \qquad F = 353,6.$$

Хотя уравнение регрессии статистически значимо по величине F-критерия и соответственно R^2, отдельные параметры незначимы:

$$b_1 = 0,585; \quad b_2 = 0,757; \quad b_3 = 1,373.$$

2. Теперь построим эконометрическую модель, используя лаги Алмон.

Предполагая линейную структуру лага, то есть линейную зависимость коэффициентов регрессии (b_j) от величины лага j, а именно: $b_j = c_0 + c_1 j$, получим следующие выражения для коэффициентов регрессии b_j:

$$j = 0 \qquad b_0 = c_0$$
$$j = 1 \qquad b_1 = c_0 + c_1$$
$$j = 2 \qquad b_2 = c_0 + 2c_1$$
$$j = 3 \qquad b_3 = c_0 + 3c_1$$

Подставив эти выражения в модель с распределенными лагами, получим:

$$y_t = a + c_0 x_t + (c_0 + c_1)x_{t-1} + (c_0 + 2c_1)x_{t-2} + (c_0 + 3c_1)x_{t-3}.$$

Перегруппируем слагаемые относительно коэффициентов "c_j":

$$y_t = a + c_0(x_t + x_{t-1} + x_{t-2} + x_{t-3}) + c_1(x_{t-1} + 2x_{t-2} + 3x_{t-3}).$$

Выражения в скобках представляют собой новые переменные z, которые находятся по исходным данным (табл. 11.5).

Таблица 11.5

Расчетная таблица

№ п/п	y_t	x_t	x_{t-1}	x_{t-2}	x_{t-3}	z_0	z_1
1	2	3	4	5	6	7	8
1	3,5	1,51	–	–	–	–	–
2	3,6	1,5	1,51	–	–	–	–
3	3,6	1,5	1,5	1,51	–	–	–
4	3,7	1,53	1,5	1,5	1,51	6,04	9,03
5	3,7	1,53	1,53	1,5	1,5	6,06	9,03
6	3,8	1,55	1,53	1,53	1,5	6,11	9,09
7	3,9	1,58	1,55	1,53	1,53	6,19	9,2
8	4,1	1,62	1,58	1,55	1,53	6,28	9,27
9	4,2	1,65	1,62	1,58	1,55	6,4	9,43
10	4,3	1,63	1,65	1,62	1,58	6,48	9,63
11	4,4	1,65	1,63	1,65	1,62	6,55	9,79
12	4,5	1,67	1,665	1,63	1,65	6,615	9,875

Окончание табл. 11.5

1	2	3	4	5	6	7	8
13	4,5	1,64	1,67	1,665	1,63	6,605	9,89
14	4,6	1,69	1,64	1,67	1,665	6,665	9,975
15	4,7	1,74	1,69	1,64	1,67	6,74	9,98
16	4,9	1,8	1,74	1,69	1,64	6,87	10,04
17	4,8	1,75	1,8	1,74	1,69	6,98	10,35
18	4,8	1,65	1,75	1,8	1,74	6,94	10,57
19	5	1,73	1,65	1,75	1,8	6,93	10,55
20	5,1	1,81	1,73	1,65	1,75	6,94	10,28
21	5,3	1,87	1,81	1,73	1,65	7,06	10,22
22	5,4	1,88	1,87	1,81	1,73	7,29	10,68
23	5,4	1,8	1,88	1,87	1,81	7,36	11,05
24	5,4	1,84	1,8	1,88	1,87	7,39	11,17

Далее применяем МНК относительно y, z_0 и z_1 по позициям $4-24$, то есть строим уравнение $\hat{y} = a + c_0 z_0 + c_1 z_1$.

Результаты регрессии записываем в виде:

$$
\begin{array}{llll}
y = -4,325 + & 1,869 z_0 & - 0,36 z_1 & \\
t \quad (-14) & (8,1) & (-2,4) & \\
R^2 = 0,979 & F = 421,4. & &
\end{array}
$$

Построенное уравнение регрессии статистически значимо и значимы все коэффициенты регрессии при новых переменных. Найдем теперь коэффициенты регрессии b_j, используя ранее показанные выражения:

$$b_0 = c_0 = 1,869;$$

$$b_1 = c_0 + c_1 = 1,869 - 0,36 = 1,509;$$

$$b_2 = c_0 + 2c_1 = 1,869 - 2 \cdot 0,36 = 1,15;$$

$$b_3 = c_0 + 3c_1 = 1,869 - 3 \cdot 0,36 = 0,79.$$

Тогда модель с распределенными лагами примет вид:

$$\begin{array}{llllll} \hat{y} = -4,325 & +1,869x_t & +1,509x_{t-1} & +1,15x_{t-2} & +0,79x_{t-3} \\ t \quad (-14) & (8,1) & (17,1) & (13,1) & (3,5) \end{array}$$

Подставим в это уравнение исходные данные $x_t, x_{t-1}, x_{t-2}, x_{t-3}$ и получим теоретические значения y, то есть $\hat{y}$, которые совпадают с $\hat{y}$ при подстановке в уравнение $\hat{y} = a + c_0 z_0 + c_1 z_1$ значений z_0 и z_1.

Так для позиции 4, подставив z_0 и z_1 имеем:

$$\hat{y} = -4,325 + 1,869 \cdot 6,04 - 0,36 \cdot 9,03 = 3,716.$$

Подставим теперь значения $x_t, x_{t-1}, x_{t-2}, x_{t-3}$ в модель с распределенными лагами:

$$\hat{y} = -4,325 + 1,869 \cdot 1,53 + 1,509 \cdot 1,5 + 1,15 \cdot 1,5 + 0,79 \cdot 1,51 = 3,716.$$

В модели с распределенными лагами (с учетом лагов Алмон) коэффициент детерминации несколько ниже: $0,979$ против $0,989$ при обычном МНК. Однако при этом все параметры модели оказываются статистически значимыми: как коэффициенты c_j, так и b_j, определяемые через c_j. Поэтому модель с распределенными лагами с применением процедуры лаги Алмон более предпочтительна.

Анализ модели показывает, что рост инвестиций в основной капитал на 1 млн грн в текущем периоде приведет через три года к росту продукции в среднем на $1,869 + 1,509 + 1,15 + 0,79 = 5,318$, $5,318$ млн грн. При этом в текущем периоде реализуется $35,1\,\%$ этого воздействия (1,869/5,318), а через год – еще 28,4 % (1,509/5,318). Иными словами, в текущем году рост инвестиций на 1 млн грн ведет к росту объема продукции на $1,869$ млн грн (краткосрочный мультипликатор), а через год прирост составит $1,869 + 1,509 = 3,378$ млн грн. Через два года прирост продукции составит уже $3,378 + 1,15 = 4,528$ млн грн, и через три года долгосрочный мультипликатор составит $5,318$ млн грн.

Пример 11.5. Оценить параметры конечной дистрибутивно-лаговой модели, используя лаги Алмон на основании данных о величине месячного дохода и расходов на потребление (табл. 11.6).

Таблица 11.6

Выборочная совокупность данных о величине месячного дохода и расходов на потребление

№ наблюдения (помесячно)	Затраты на потребление, грн y_t	Доход семьи x_t
1	332,16	564,62
2	301,72	57,88
3	318,06	674,02
4	380,29	779,88
5	375,64	835,76
6	386,6	843,68
7	394,4	867,88
8	425,24	898,2
9	434,68	900,24
10	438,12	906
11	449,72	910,64
12	458,64	919,52
13	463,2	956,6
14	471,26	906,71
15	481,85	926,06
16	525,58	1 056,52
17	605,95	1 210,51
18	705,63	1 357,01
19	861,34	1 539,23
20	1 008,36	1 650,98
21	1 136,29	1 797,91
22	1 083,67	1 700,6
23	1 043,44	1 822,61
24	1 246,57	2 027,29

Решение

Допустим, что на величину затрат на потребление влияет величина дохода семьи за текущий и три прошедшие месяцы. Тогда конечная дистрибутивно-лаговая модель имеет вид:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t ,$$

где x_t – величина дохода семьи, грн;

y_t – затраты на потребление, грн;

ε_t – случайная величина.

На первом шаге предположим, что для оценивания параметров конечной дистрибутивно-лаговой модели наилучшим является полином второй степени. Тогда параметры модели можно изобразить таким образом:

$$\beta_i = a_0 + a_1 i + a_2 i^2 , i = 0, 1, 2, 3.$$

По схеме Алмон дистрибутивно-лаговая модель зависимости расходов на потребление от дохода семьи сводится к множественной линейно-корреляционно-регрессионной модели вида:

$$y_t = \alpha + a_0 z_{0t} + a_1 z_{1t} + a_2 z_{2t} + \varepsilon_t ,$$

где

$z_{0t} = x_t + x_{t-1} + x_{t-2} + x_{t-3}$; $z_{1t} = x_{t-1} + 2x_{t-2} + 3x_{t-3}$; $z_{2t} = x_{t-1} + 4x_{t-2} + 9x_{t-3}$.

Вычисление значений вспомогательных переменных z_{0t}, z_{1t}, z_{2t}, z_{3t} приведено в табл. 11.7.

Отметим, что вспомогательные переменные рассчитаны, начиная с четвертого периода, так как они содержат текущую величину дохода и три лаговых значения этого показателя.

Используя метод наименьших квадратов, получает следующую выборочную множественную корреляционно-регрессионную модель:

$$\tilde{y}_t = -133,3727 - 0,0501 z_{0t} + 0,7594 z_{1t} - 0,2621 z_{2t} .$$

Коэффициент множественной корреляции равен $0,9864$ (коэффициент детерминации составляет $R^2 = 0,973$), что говорит о тесной связи результативной переменной y и вспомогательных переменных z_{0t}, z_{1t}, z_{2t}.

Таблица 11.7

Значение вспомогательных переменных $z_{0t}, z_{1t}, z_{2t}, z_{3t}$

№ наблюдения (помесячно)	Значение переменных для оценки параметров модели			
	z_{0t}	z_{1t}	z_{2t}	z_{3t}
4	2 616,4	3 563,64	8 147,12	20 701,8
5	2 887,54	3 921,56	8 856,88	22 314,8
6	3 133,34	4 417,58	10 021,46	25 273,34
7	3 327,2	4 854,84	11 205,64	28 586,52
8	3 445,52	5 062,52	11 764,44	30 182,84
9	3 510	5 165	11 962,84	30 620,6
10	3 572,32	5 300,28	12 303,96	31 518,6
11	3 615,08	5 401,08	12 590,76	32 359,32
12	3 636,4	5 423,36	12 636,8	32 465,12
13	3 692,76	5 458,8	12 716,08	32 666,64
14	3 693,47	5 527,56	12 830,44	32 900,04
15	3 708,89	5 578,47	13 008,79	33 386,55
16	3 845,89	5 609,28	13 162,3	34 007,94
17	4 099,8	5 628,77	12 921,15	32 946,17
18	4 550,1	6 101,73	13 771,13	34 666,29
19	5 163,27	6 947,59	15 707,73	39 567,13
20	5 757,73	7 884,78	17 861,86	45 079,08
21	6 345,13	8 800,47	20 020,99	50 604,09
22	6 688,72	9 717,56	22 254,9	56 564,96
23	6 972,1	10 249,36	23 751,06	60 660,34
24	7 348,41	10 617,54	24 806,2	63 970,98

Значение t-критерия Стьюдента для проверки статистической значимости коэффициентов регрессии $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2$ выборочной множественной корреляционно-регрессионной модели составляет:

$$t_{\tilde{a}_0} = 2{,}191;\ t_{\tilde{a}_1} = 4{,}7179;\ t_{\tilde{a}_2} = 3{,}6768.$$

Табличное значение распределение Стьюдента с $df = 24 - 4 = 20$ степенями свободы и заданным уровнем значимости $\alpha = 0{,}05$ равен

$t_{0,05} = 2,086$. Таким образом с вероятностью $p = 0,95$ можно утверждать, что оценки $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2$ параметров a_0, a_1, a_2 множественной корреляционно-регрессионной модели статистически значимы.

После этого предположим, что параметры конечной дистрибутивно-лаговой модели можно описать полиномом третей степени, то есть:

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3, \quad i = 0, 1, 2, 3.$$

Для полинома третей степени множественная линейная корреляционно-регрессионная модель имеет вид:

$$y_t = a + a_0 z_{0t} + a_1 z_{1t} + a_2 z_{2t} + a_3 z_{3t} + \varepsilon_t,$$

где $z_{0t} = x_t + x_{t-1} + x_{t-2} + x_{t-3}$; $z_{1t} = x_{t-1} + 2x_{t-2} + 3x_{t-3}$;

$z_{2t} = x_{t-1} + 4x_{t-2} + 9x_{t-3}$; $z_{3t} = x_{t-1} + 8x_{t-2} + 27x_{t-3}$.

Значения вспомогательных переменных z_{it}, $i = \overline{0,3}$ находятся в табл. 11.8.

Выборочная множественная корреляционно-регрессионная модель имеет вид:

$$\tilde{y}_t = -187,1837 - 0,0265 z_{0t} + 3,2984 z_{1t} - 2,8879 z_{2t} + 0,6115 z_{3t}.$$

Коэффициент множественной корреляции равен $0,9914$ (коэффициент детерминации составляет $R^2 = 0,9828$). Это свидетельствует о тесной связи результативной переменной y и вспомогательных переменных $z_{0t}, z_{1t}, z_{2t}, z_{3t}$.

Значение t-критерия Стьюдента для проверки статистической значимости коэффициентов регрессии $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3$ этой модели равно:

$$t_{\tilde{a}_0} = 1,3016; \; t_{\tilde{a}_1} = 3,9058; \; t_{\tilde{a}_2} = 3,3149; \; t_{\tilde{a}_3} = 3,0208.$$

Табличное значение распределение Стьюдента при заданном уровне значимости $\alpha = 0,05$ и числе степеней свободы, равным $df = 24 - 4 - 1 = 19$, составляет $t_{0,05} = 2,093$. А значит с вероятностью $p = 0,95$ можно утверждать статистическую значимость оценок $\tilde{a}_1, \tilde{a}_2, \tilde{a}_3$

параметров a_1, a_2, a_3 множественной корреляционно-регрессионной модели статистическую незначимость оценки $\tilde{a}_0$ и параметра a_0, для которого $t_{\tilde{a}_0} < t_{табл}$ $(1,3016 < 2,093)$.

Таким образом, для аппроксимации параметров конечной дистрибутивно-лаговой модели можно использовать полином второй степени.

Тогда оценки параметров выборочной множественной корреляционно-регрессионной модели составят:

$$\tilde{a}_0 = -0,0501;\ \tilde{a}_1 = 0,7594;\ ,\tilde{a}_2 = -0,2621.$$

Оценки b_0, b_1, b_2, b_3 для параметров $\beta_0, \beta_1, \beta_2, \beta_3$ конечной дистрибутивно-лаговой модели равны:

$b_0 = a_0 = -0,0501;$

$b_1 = a_0 + a_1 + a_2 = -0,0501 + 0,7594 - 0,2621 = 0,4472;$

$b_2 = a_0 + 2a_1 + 4a_2 = -0,0501 + 2 \cdot 0,7594 - 4 \cdot 0,2621 = 0,4201;$

$b_3 = a_0 + 3a_1 + 9a_2 = -0,0501 + 3 \cdot 0,7594 - 9 \cdot 0,2621 = -0,1311.$

Таким образом, выборочная конечная дистрибутивно-лаговая модель имеет вид:

$$\tilde{y}_t = -133,3727 - 0,0501x_t + 0,4472x_{t-1} + 0,4201x_{t-2} - 0.1311x_{t-3}.$$

Долгосрочный мультипликатор уровня потребления составляет $0,6861$, то есть на протяжении четырех периодов на потребление тратят $68,61\,\%$ от всей величины дохода семьи. При этом основной прирост затрат на потребление можно наблюдать с задержкой в один и два периода.

11.3. Задачи

Задача 11.1. На основании статистических данных о величинах дохода и затрат на потребление (табл. 11.8) для авторегрессионной модели

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \varepsilon_t$$

- найти вспомогательную переменную и обосновать ее выбор;
- оценить параметры авторегрессионной модели с помощью метода наименьших квадратов;

- проверить значимость оценок параметров модели;
- объяснить результаты.

Таблица 11.8

Статистические данные

Временной период	Затраты на потребление, грн, y_t	Доход грн, x_t
1	8,56	11,54
2	8,69	12,56
3	9,75	13,05
4	10,58	14,23
5	12,65	15,97
6	12,97	16,25
7	14,32	17,2
8	16,2	18,96
9	16,84	21,03
10	17,09	22,69
11	17,69	23,45
12	18,02	27,76
13	18,56	30,64
14	19,35	31,58
15	19,49	34,69

Задача 11.2. На основании статистических данных о величинах прибыли предприятий и затрат на обновление производства (табл. 11.9) построить дистрибутивно-лаговую модель, экономически интерпретировать переменные и объяснить причины наличия лагов в этой модели.

Таблица 11.9

Статистические данные

Временной период	Прибыль, тыс. грн, y_t	Затраты на обновление производства, тыс. грн, x_t
1	2	3
1	26,15	86,45
2	27,65	95,64
3	28,96	115,06
4	31,78	118,23

Окончание табл. 11.9

1	2	3
5	40,25	152,03
6	49,35	161,24
7	52,68	175,62
8	60,25	196,35
9	62,31	200,65
10	68,91	205,36
11	72,41	214,68
12	72,3	216,21
13	74,92	254,67
14	76,89	268,79
15	85,69	310,65

Задача 11.3. Имеются данные об объеме экспорта и импорта Украины (в млн дол. США) за 2010-2012 гг. (табл. 11.10).

Таблица 11.10

Статистические данные

Месяц	Экспорт			Импорт		
	2010	2011	2012	2010	2011	2012
Январь	3 008,0	4 652,8	5 324,8	3 262,3	5 082,1	5 385,3
Февраль	3 372,0	4 691,0	4 957,7	3 712,8	6 386,7	6 760
Март	3 944,8	6 004,2	5 875,7	4 716,3	7 047,7	6 933,7
Апрель	4 206,7	5 603,8	5 795,0	4 602,9	6 297,7	7 263,4
Май	4 194,2	5 697,4	6 239,1	4 412,8	6 766,5	7 582,3
Июнь	4 328,6	6 156,7	5 473,4	4 718,8	6 788,1	6 967,1
Июль	4 241,0	5 355,3	5 755,1	5 163,0	6 527,4	7 197,5
Август	4 249,5	5 781,2	5 835,7	5 423,0	7 213,8	7 277,6
Сентябрь	4 699,7	5 964,4	5 766,5	5 674,6	7 385,4	6 927,4
Октябрь	4 740,7	5 753,9	6 244,2	6 177,0	7 543,7	7 674,6
Ноябрь	5 128,8	6 262,3	5 951,1	6 232,4	7 678,5	6 879,7
Декабрь	5 291,2	6 471,2	5 591,5	6 646,3	7 890,6	7 809,5

Необходимо построить модель с распределенными лагами, используя лаги Алмон, предполагая структуру лага линейной.

Задача 11.4. Имеются данные о динамике товарооборота и доходов населения Украины за 2005 – 2007 гг. (табл. 11.11).

Таблица 11.11

Статистические данные

Месяц	Товарооборот, % к предыдущему году	Доходы населения, % к предыдущему году	Месяц	Товарооборот, % к предыдущему году	Доходы населения, % к предыдущему году
I	120,4	116,8	VII	124,8	117,3
II	119,0	126,6	VIII	124,6	117,4
III	117,5	125,1	IX	124,2	116,6
IV	118,3	123,4	X	124,4	111
V	118,8	124,8	XI	124,2	111
VI	119,8	126,2	XII	124,8	111,7
VII	120,6	126,8	I	124,0	116,1
VIII	121,7	125,8	II	125,9	112,3
IX	121,8	124,4	III	125,8	112,2
X	121,7	122,3	IV	125,9	109,2
XI	121,7	118,6	V	126,4	108,8
XII	122,4	119,8	VI	126,4	111,8
I	128,0	122,8	VII	128,4	110,6
II	125,6	122,8	VIII	128,4	111,8
III	124,7	123,0	IX	128,2	111,0
IV	126,1	126,6	X	114,0	112,8
V	125,2	115,0	XI	128,9	114,4
VI	125,5	120,6	XII	128,8	115,9

Необходимо построить модель с распределенными лагами вида

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3}$$

двумя способами:

1. Применить к исходным данным обычный МНК и сделать выводы.
2. Использовать лаги Алмон, предполагая структуру лага линейной.

Сравнить оба варианта.

12. Моделирование временных рядов

12.1. Основные теоретические сведения

Временным рядом называется упорядоченная во времени последовательность показателей, характеризующих уровни развития изучаемого явления в последовательные моменты или периоды времени. Отдельные наблюдения временного ряда называются **уровнями этого ряда.**

Временные ряды бывают моментные, интервальные и производные.

Моментные ряды характеризуют значения показателя на определенные моменты времени.

Интервальные ряды характеризуют значения показателя за определенные интервалы времени.

Производные ряды получаются из средних или относительных величин показателя.

Важным условием правильного отражения временным рядом реального процесса развития является сопоставимость уровней временного ряда. Несопоставимость чаще всего встречается в стоимостных характеристиках, изменениях цен, территориальных изменениях, при укрупнении предприятий. Для несопоставимых величин показателя неправомерно проводить прогнозирование.

Считают, что значения уровней временных рядов экономических показателей складываются из следующих компонент: тренда T, сезонной S, циклической C и случайной E.

Модель, в которой временный ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью временного ряда:* $Y_t = T_t + S_t + C_t + E_t$. Она применима в тех случаях, когда анализируемый временной ряд имеет приблизительно одинаковые изменения на протяжении всей длительности.

Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной моделью временного ряда*: $Y_t = T_t \cdot S_t \cdot C_t \cdot E_t$. Модель мультипликативных компонент эффективнее в тех случаях, когда изменение временной последовательности увеличивается с ростом уровня, то есть значения ряда расходятся как имеющие тренд, а наблюдаемая последовательность значений напоминает рупор или воронку.

Корреляционную зависимость между последовательными уровнями временного ряда называют **автокорреляцией уровней ряда**. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называется лагом. Считается, что максимальный лаг должен быть не больше $\frac{n}{4}$.

Коэффициент автокорреляции характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию, коэффициент автокорреляции может приближаться к нулю. По знаку коэффициент автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Последовательность коэффициентов автокорреляции уровней первого, второго и так далее порядков называют *автокорреляционной функцией временного ряда*. График зависимости ее значений от величины лага (порядка коэффициент автокорреляции) называется коррелограммой. При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

При анализе временного ряда существует основная задача – определения основной тенденции в развитии исследуемого явления.

На практике простым методом обнаружения общей тенденции является укрупнение интервалов. Например, ряд недельных данных можно преобразовать в ряд месячных данных, ряд квартальных данных – в годовые. Таким преобразованием может быть суммирование уровней исходного ряда или нахождение средних значений. Этот метод называется *сглаживанием временного ряда*.

Суть этого метода заключается в замене фактических уровней временного ряда расчетными, которые в меньшей степени подвержены колебаниям и способствует более четкому проявлению тенденции развития. Скользящие средние позволяют сгладить случайные и периодические колебания временного ряда.

Алгоритм сглаживания по простой скользящей средней.

1. Определяют длину интервала сглаживания l, включающего в себя l последовательных уровней ряда ($l < n$). Чем шире интервал сглаживания, тем в большей степени поглощаются колебания, и тенденция развития носит более плавный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

2. Весь период наблюдений разбивают на части и интервал сглаживания переходит по ряду с шагом равным 1.

3. Для каждого уровня временного ряда вычисляется его среднее арифметическое значение.

4. Заменяют фактические значения ряда, стоящие в центре каждой части, на соответствующие средние значения.

Удобно длину интервала сглаживания l брать в виде нечетного числа $l = 2p + 1$, поскольку в этом случае полученные значения скользящей средней приходятся на средний интервал ряда. Наблюдения, которые берутся для расчета среднего значения, называются активной частью. При нечетном значении $l = 2p + 1$ все уровни активной части могут быть представлены в виде:

$$y_{t-p},\, y_{t-p+1},\, \ldots,\, y_{t-1},\, y_t,\, y_{t+1},\, \ldots,\, y_{t+p-1},\, y_{t+p},$$

где y_t – центральный уровень активной части;

$y_{t-p}, y_{t-p+1}, \ldots, y_{t-1}$ – последовательность из p уровней активной части, предшествующей центральной;

$y_{t+1}, \ldots, y_{t+p-1}, y_{t+p}$ – последовательность из уровней активной части, следующей за центральной.

В этом случае скользящая средняя определяется по формуле:

$$\overline{y_t} = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p-1}, y_{t+p}}{2p+1},$$

где y_i – фактическое значение i-го уровня;

$\overline{y_t}$ – значение скользящей средней в момент t;

$2p + 1$ – длина интервала сглаживания.

Метод сглаживания приводит к устранению периодических колебаний во временном ряду, если длина интервала сглаживания равна или кратна периоду колебаний. Рекомендуется для устранения сезонных колебаний использовать скользящие средние с длиной интервала сглаживания, равной 4 или 12, однако при этом не будет выполняться условие нечетности. В этом случае скользящая средняя рассчитывается по формуле:

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + ... + y_{t-1} + y_t + y_{t+1} + ... + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} =$$

$$= \frac{\frac{1}{2}y_{t-p} + \sum\limits_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}.$$

Поэтому для сглаживания сезонных колебаний с квартальным временным рядом или месячным временным рядом используют 4-членную и 12-членную скользящую среднюю:

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4};$$

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-6} + y_{t-51} + ... + y_t + ... + y_{t+5} + \frac{1}{2}y_{t+6}}{12}.$$

Метод простой скользящей средней дает хорошие результаты в динамических рядах с линейной тенденцией развития. Если же данный процесс развивается нелинейно, то простая скользящая средняя приводит к существенным искажениям и в этом случае рекомендуют использовать взвешенную скользящую среднюю:

$$\overline{y_t} = \frac{\sum\limits_{i=t-p}^{t+p} y_i w_i}{\sum\limits_{i=t-p}^{t+p} w_i},$$

где w_i – весовые коэффициенты.

Для полиномов 2-го и 3-го порядков по 5-членной взвешенной скользящей средней центральное значение интервала определяется по формуле:

$$y_t = \frac{-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}}{35}.$$

Весовые коэффициенты обладают следующими свойствами:

1) они симметричны относительно центрального уровня;

2) сумма весов с учетом общего множителя, вынесенного за скобки, равна единице;

3) имеются как положительные, так и отрицательные веса, что позволяет сглаженной кривой сохранять различные изгибы кривой тренда.

Комплекс аналитических методов выравнивания сводится к выбору конкретных кривых роста и определению их параметров. Под *кривой роста* понимают некоторую функцию, аппроксимирующую заданный динамический ряд.

В процедуре прогнозирования на основе кривых роста выделяют такие этапы:

1) выбор одной или нескольких кривых, форма которых соответствует характеру изменения временного ряда;

2) оценка параметров выбранных кривых;

3) проверка адекватности выбранных кривых прогнозируемому процессу или явлению, оценка точности моделей и окончательный выбор кривой роста;

4) расчет точечного и интервального прогнозов.

Модели кривых роста рекомендуют разделять на три группы. К первой группе относятся функции, используемые для описания процессов с монотонным характером тенденции развития и отсутствия пределов роста. Это характерно для тенденций изменения многих экономических показателей промышленных предприятий. Ко второй группе относятся кривые, описывающий процесс, имеющий предел роста в исследуемом периоде. Такие процессы чаще всего демографические, хотя встречаются и в исследовании экономических процессов на промышленных предприятиях. Функции, относящиеся ко второму классу, называются кривыми с насыщением. Если кривые насыщения имеют точки перегиба, то они относятся к третьему классу – к S-образным кривым. По кривым третьей группы прогнозируют процессы научно-технического прогресса, нового производства продукции.

В прогнозировании экономических показателей с помощью кривых роста чаще всего применяются следующие функции:

$$y = a + bt,\ y = a + \frac{b}{t},\ y = e^{a+bt},\ y = a \cdot t^b,\ y = a + b_1 t + b_2 t^2 + \ldots + b_k t^k .$$

В подавляющем большинстве случаев расчет оценок параметров моделей осуществляется с помощью метода наименьших квадратов в форме регрессионных моделей, в которых в качестве зависимой переменной служат значения показателей, а фактором есть время. Для нелинейных трендовых моделей применяют процедуру линеаризации.

Наиболее простым методом выбора кривых роста считается визуальный, основный на графическом изображении временного ряда. Если на графическом изображении четко не прослеживается тенденция, то рекомендуют провести некоторые стандартные преобразования ряда (например, сглаживание) и подобрать соответствующую функцию. С помощью современных программных средств, например, статистических пакетов процедуру выбора кривой роста осуществить очень легко.

Выбор наилучшего уравнения для построения тренда можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации R^2 и выбора уравнения тренда с максимальным значением скорректированного коэффициента детерминации.

Статистическое качество построенных моделей кривых роста для прогноза проверяется по критериям проверки качества построенных регрессионных моделей: критерия Стьюдента, критерия Фишера, критерия Дарбина – Уотсона. Существование автокорреляции остатков может существенно искажать прогнозные значения.

Статистические методы оценки уровня сезонности

В настоящее время при описании и прогнозировании тренд-сезонных процессов используются комбинированные подходы, связанные с применением индексов сезонности вместе с кривыми роста, процедуры, опирающиеся на адаптивные модели, сезонный вариант модели ARIMA, специализированные подходы, учитывающие особенности конкретных временных рядов.

Процедура расчета сезонной составляющей зависит от принятой модели временного ряда – аддитивной или мультипликативной. Выбор формы модели зависит от структуры сезонных колебаний. Если амплитуда колебаний почти постоянна, то разрабатывают аддитивную модель временного ряда, в которой значения сезонной компоненты считаются постоянными для различных циклов. Если же амплитуда сезонных колебаний возрастает или уменьшается, то разрабатывают мультипликативную модель временного ряда. В аддитивной модели характеристики сезонности измеряются в абсолютных величинах, в мультипликативной – в относительных.

Алгоритм расчета для аддитивной сезонности

1. На первом этапе предполагается описание тенденции с помощью скользящей средней при четной длине интервалов сглаживания $l = 2p$. Скользящая средняя для временных рядов месячной динамики определяется по формуле:

$$y_t^{'} = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + ... + y_t + ... + y_{t+5} + \frac{1}{2}y_{t+6}}{12};$$

для рядов квартальной динамики:

$$y_t^{'} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

2. На втором этапе рассчитывается отклонение фактических значений от уровней сглаженного ряда: $x_t = y_t - y_t^{'}$. Уровни вновь полученного ряда отражают эффект сезонности и случайности.

3. На третьем этапе для элиминирования влияния случайных факторов определяется предварительные значения сезонной составляющей как средние значения из уровней x_t для одноименных месяцев (кварталов).

4. На четвертом этапе проводится корректировка первоначальных значений сезонной составляющей. Для аддитивного случая сумма значений сезонной составляющей для полного сезонного цикла равняется нулю. Поэтому скорректированные оценки сезонной компоненты определяются по формуле: $S_i = \overline{x_i} - \bar{x}$, где $\bar{x} = \frac{1}{m}\sum_{i=1}^{m}\overline{x_i}$, m – число фаз в сезонном цикле (для месячной динамики $m = 12$, для квартальных данных $m = 4$).

Алгоритм расчета для мультипликативной сезонности

1. На первом этапе предполагается описание тенденции с помощью скользящей средней при четной длине интервалов сглаживания $l = 2p$. Скользящая средняя для временных рядов месячной динамики определяется по формуле:

$$y_t' = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12};$$

для рядов квартальной динамики:

$$y_t' = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

2. На втором этапе рассчитывается отклонение фактических значений от уровней сглаженного ряда как $x_t = \frac{y_t}{y_t'}$.

Взаимопогашаемость сезонных колебаний в мультипликативной форме выражается в том, что средняя арифметическая значений коэффициентов сезонности для сезонного цикла равняется единице.

3. На третьем этапе для элиминирования влияния случайных факторов определяется предварительные значения сезонной составляющей как средние значения из уровней x_t для одноименных месяцев (кварталов).

4. Окончательные оценки коэффициентов сезонности будут:

$$S_i = \overline{x_i}\,\overline{x}, \quad i = \overline{1, m},$$

где $\overline{x} = \frac{m}{\sum_{i=1}^{m} \overline{x_i}}$, m – число фаз в сезонном цикле.

Алгоритм построения тренд-сезонных моделей

1. Оценивание сезонной составляющей проводится по рассмотренным алгоритмам с учетом характера сезонности: аддитивной или мультипликативной.

2. Сезонная корректировка исходных данных.

3. Расчет параметров тренда на основе временного ряда, полученного на предыдущем шаге.

4. Моделирование динамики исходного ряда с учетом трендовой и сезонной составляющих.

5. Оценка точности и адекватности полученной модели.

6. Применение вычисленной модели для прогнозирования.

12.2. Примеры

Пример 12.1. Расчет коэффициентов автокорреляции уровней для временного ряда расходов на конечное потребление.

Пусть имеются следующие условные данные (табл. 12.1) о средних расходах на конечное потребление (y_t , ден. ед.) за 8 лет.

Таблица 12.1

Расчет коэффициента автокорреляции первого порядка

t	y_t	y_{t-1}	$y_t-\overline{y_1}$	$y_{t-1}-\overline{y_2}$	$y_t-\overline{y_1} \quad y_{t-1}-\overline{y_2}$	$y_t-\overline{y_1}^{\,2}$	$y_{t-1}-\overline{y_2}^{\,2}$
1	7	–	–	–	–	–	–
2	8	7	-3,29	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2,58	1,6641	4
5	11	10	-0,29	0	0	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1
7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16
Итого	86	70	-0,03	0	44,0	53,4287	38

$$r_1 = \frac{\sum_{t=2}^{n} y_t-\overline{y_1} \quad y_{t-1}-\overline{y_2}}{\sqrt{\sum_{t=2}^{n} y_t-\overline{y_1}^{\,2} \quad y_{t-1}-\overline{y_2}^{\,2}}} = \frac{44}{\sqrt{53,42\cdot 38}} = 0,976,$$

$$\overline{y_1} = \frac{1}{n-1}\sum_{t=2}^{n} y_t = \frac{8+8+10+11+12+14+16}{7} = 11,29,$$

$$\overline{y_2} = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1} = \frac{7+8+8+10+11+12+14}{7} = 10.$$

Полученное значение свидетельствует об очень тесной зависимости между расходами на конечное потребление текущего и непосредственно предшествующего годов и, следовательно, о наличии во временном ряде расходов на конечное потребление сильной линейной тенденции.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-1} (табл. 12.2).

$$r_2 = \frac{\sum_{t=2}^{n}(y_t - \overline{y_3})(y_{t-2} - \overline{y_4})}{\sqrt{\sum_{t=2}^{n}(y_t - \overline{y_3})^2 (y_{t-2} - \overline{y_4})^2}} = \frac{27,3334}{\sqrt{40,8334 \cdot 19,3334}} = 0,973,$$

$$\overline{y_3} = \frac{\sum_{t=3}^{n} y_t}{n-2} = \frac{8+10+11+12+14+16}{6} = 11,83, \quad \overline{y_4} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+8+8+10+11+12}{6} = 9,33.$$

Таблица 12.2

Расчет коэффициента автокорреляции второго порядка

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3) \cdot (y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	7	–	–	–	–	–	–
2	8	–	–	–	–	–	–
3	8	7	-3,83	-2,33	8,9239	14,6689	5,4289
4	10	8	-1,83	-1,33	2,4339	3,3489	1,7689
5	11	8	-0,83	-1,33	1,1039	0,6889	1,7689
6	12	10	0,17	0,67	0,1139	0,0289	0,4489
7	14	11	2,17	1,67	3,6239	4,7089	2,7889
8	16	12	4,17	2,67	11,1339	17,3889	7,1289
Σ	86	56	0,02	0,02	27,3334	40,8334	19,3334

Полученные результаты подтверждают вывод о том, что ряд расходов на конечное потребление содержит линейную тенденцию.

Пример 12.2. По данным об урожайности пшеницы (y_t) за 16 лет рассчитать трехлетние и семилетние скользящие средние и графически сравнить результаты (табл. 12.3).

Таблица 12.3

Урожайность пшеницы, ц/га

t	1	2	3	4	5	6	7	8
y_t	10,3	14,3	7,7	15,8	14,4	16,7	15,3	20,2
t	9	10	11	12	13	14	15	16
y_t	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

Решение. Результаты расчетов приведены в табл. 12.4

Таблица 12.4

Расчет скользящих средних

t	y_t	$l=3$	$l=7$
1	10,3	–	–
2	14,3	10,8	–
3	7,7	12,6	–
4	15,8	12,6	13,5
5	14,4	15,6	14,9
6	16,7	15,5	15,3
7	15,3	17,4	15,3
8	20,2	17,5	15,2
9	17,1	15,0	15,5
10	7,7	13,4	16,0
11	15,3	13,1	15,8
12	16,3	17,2	15,6
13	19,9	16,9	16,1
14	14,4	17,7	–
15	18,7	17,9	–
16	20,7	–	–

При трехлетней скользящей средней:

$\bar{y}_2 = \dfrac{10,3+14,3+7,7}{3} = 10,8$; $\bar{y}_3 = \dfrac{14,3+7,7+15,8}{3} = 12,6$;

$\bar{y}_4 = \dfrac{7,7+15,8+14,4}{3} = 12,6$; $\bar{y}_5 = \dfrac{15,8+14,4+16,7}{3} = 15,6$;

$\bar{y}_6 = \dfrac{14,4+16,7+15,3}{3} = 15,5$; $\bar{y}_7 = \dfrac{16,7+15,3+20,2}{3} = 17,4$;

$\bar{y}_8 = \dfrac{15,3+20,2+17,1}{3} = 17,5$; $\bar{y}_9 = \dfrac{20,2+17,1+7,7}{3} = 15,0$;

$\bar{y}_{10} = \dfrac{17,1+7,7+15,3}{3} = 13,4$; $\bar{y}_{11} = \dfrac{7,7+15,3+16,3}{3} = 13,1$;

$\bar{y}_{12} = \dfrac{15,3+16,3+19,9}{3} = 17,2$; $\bar{y}_{13} = \dfrac{16,3+19,9+14,4}{3} = 16,9$;

$\bar{y}_{14} = \dfrac{19,9+14,4+18,7}{3} = 17,7$; $\bar{y}_{15} = \dfrac{14,4+18,7+20,7}{3} = 17,9$.

При семилетней скользящей средней:

$$\bar{y}_4 = \frac{10,3+14,3+7,7+15,8+14,4+16,7+15,3}{7} = 13,5;$$

$$\bar{y}_5 = \frac{14,3+7,7+15,8+14,4+16,7+15,3+20,2}{7} = 14,9;$$

$$\bar{y}_6 = \frac{7,7+15,8+14,4+16,7+15,3+20,2+17,1}{7} = 15,3;$$

$$\bar{y}_7 = \frac{15,8+14,4+16,7+15,3+20,2+17,1+7,7}{7} = 15,3;$$

$$\bar{y}_8 = \frac{14,4+16,7+15,3+20,2+17,1+7,7+15,3}{7} = 15,2;$$

$$\bar{y}_9 = \frac{16,7+15,3+20,2+17,1+7,7+15,3+16,3}{7} = 15,5;$$

$$\bar{y}_{10} = \frac{15,3+20,2+17,1+7,7+15,3+16,3+19,9}{7} = 16,0;$$

$$\bar{y}_{11} = \frac{20,2+17,1+7,7+15,3+16,3+19,9+14,4}{7} = 15,8;$$

$$\bar{y}_{12} = \frac{17,1 + 7,7 + 15,3 + 16,3 + 19,9 + 14,4 + 18,7}{7} = 15,6;$$

$$\bar{y}_{13} = \frac{7,7 + 15,3 + 16,3 + 19,9 + 14,4 + 18,7 + 20,7}{7} = 16,1.$$

На рис. 12.1 представлены фактический и сглаженные ряды урожайности.

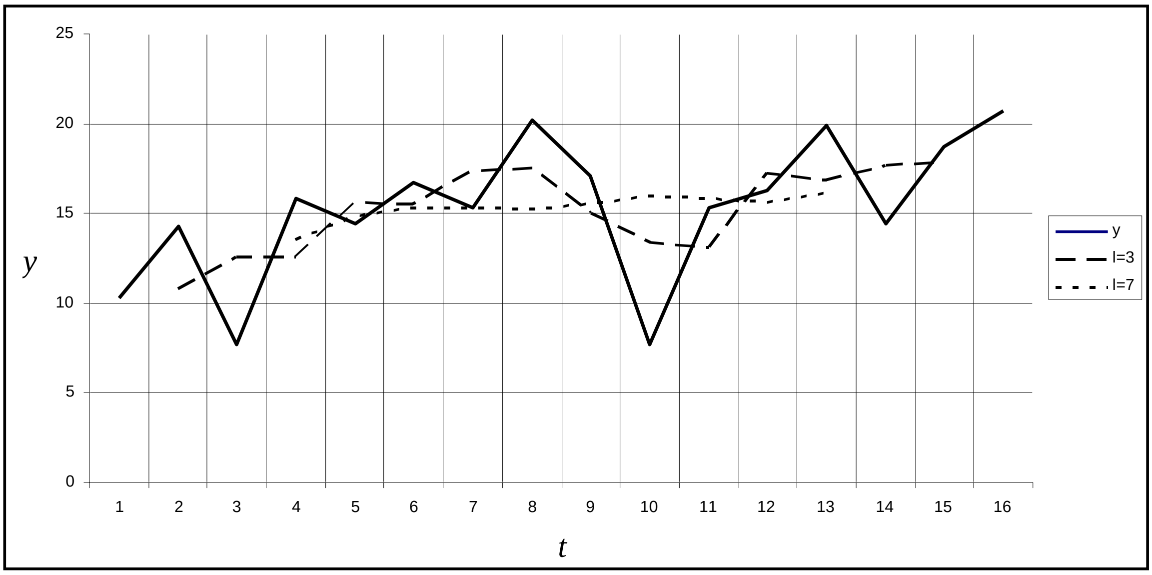


Рис. 12.1. **Сглаживание ряда урожайности с помощью скользящих средних**

Таким образом, ряд, сглаженный по семилетней скользящей средней ($l = 7$), имеет более гладкий характер по сравнению с рядом, сглаженным по трехлетней скользящей средней ($l = 3$). Чем больше длина интервала сглаживания, тем более гладким получается ряд.

Пример 12.3. Для проведения экономического анализа внешнеэкономической деятельности промышленного предприятия важно прогнозировать значения показателя эффективности реализации продукции на внутреннем рынке (y). В табл. 12.5 приведены квартальные значения показателя в течение 5 лет.

Построить модель кривой роста показателя эффективности реализации продукции промышленного предприятия на внутреннем рынке и прогнозировать его значение на два последующих квартала.

Таблица 12.5

Квартальные значения показателя

Y	1,225	1,242	1,264	1,205	1,277	1,289	1,241	1,235	1,236	1,234
t	1	2	3	4	5	6	7	8	9	10
Y	1,193	1,182	1,236	1,201	1,199	1,236	1,182	1,184	1,168	1,201
t	11	12	13	14	15	16	17	18	19	20

Решение. С помощью статистического пакета *Statgraphics Plus V5.1 International Professional* сначала была вычислена модель кривой роста в линейной форме: $y = 1,2605 - 0,0037t$, при этом статическое качество модели было проверено с помощью критериев Стьюдента, Фишера, скорректированного коэффициента детерминации и критерия Дарбина – Уотсона:

$$t_a = 105,9;\ t_b = -3,74;\ F = 14,0;\ R^2 = 0,4062;\ DW = 1,999.$$

Модель относительно качественная. На последующие кварталы прогноз значения показателя эффективности реализации продукции на внутреннем рынке такой: 1,18245; 1,17873. Видим, что эффективность реализации продукции на внутреннем рынке у промышленного предприятия будет немного снижаться в последующих двух кварталах.

В табл. 12.6 представлены рассчитанные альтернативные модели по основным кривым роста по рейтингу значений коэффициентов детерминации.

Таблица 12.6

Альтернативные модели по основным кривым роста

Альтернативные модели	Скорректированный коэффициент детерминации
1	2
$y = \dfrac{1}{0,7931 + 0,0025t}$	0,4124
$y = e^{0,2317 - 0,003t}$	0,4094

Окончание табл. 12.6

1	2
$y = 1{,}1228 - 0{,}0017t^{2}$	0,4078
$y = 1{,}2605 - 0{,}0037t$	0,4062
$y = 1{,}248 - 0{,}00029t - 0{,}000163t^2$	0,3961
$y = 1{,}2829 - 0{,}0199\sqrt{t}$	0,3407
$y = 1{,}2672t^{-0{,}0175}$	0,2348
$y = 1{,}2667 - 0{,}0213\ln t$	0,2320
$y = \dfrac{1}{0{,}8242 - \dfrac{0{,}0275}{t}}$	0,0256
$y = e^{0{,}1937 + \frac{0{,}033}{7}}$	0,0238
$y = 1{,}2142 + \dfrac{0{,}0403}{t}$	0,0219

Из табл. 12.6 видно, что учитывая значения скорректированного коэффициента детерминации можно сделать вывод, что наилучшей моделью является $y = \dfrac{1}{0{,}7931 + 0{,}0025t}$.

Прогноз по этой модели эффективности реализации продукции на внутреннем рынке такой: 1,18282; 1, 17935. Расхождение прогнозных значений, рассчитанных по двум моделям различаются в тысячных единицах.

Пример 12.4. Приведен анализ финансовых показателей предприятия (табл. 12.7). При анализе финансовых показателей была выявлена сезонность: на протяжении 1-го и 2-го кварталов двух лет наблюдалось негативное изменение показателей, в 3-м, 4-м кварталах положительная тенденция в динамике показателей. Построить прогноз работы предприятия.

Таблица 12.7

Фактические значения прибыли от реализации продукции

№ квартала	Квартал	Выручка, тыс. грн	№ квартала	Квартал	Выручка, тыс. грн
1	IV/2008	10 596	5	IV/2009	15 759
2	I/2009	7 311	6	I/2010	9 640
3	II/2009	6 893	7	II/2010	10 814
4	III/2009	12 346	8	III/2010	22 183

Решение

Сначала для прогнозирования рассмотрим аддитивную модель:

$$F_t = T + S + E,$$

где F_t – прогнозное значение;

T – тренд;

S – сезонная компонента;

E – ошибка прогноза.

По исходным данным (табл. 12.7) проследим динамику выручки от реализации продукции предприятием (рис. 12.2).

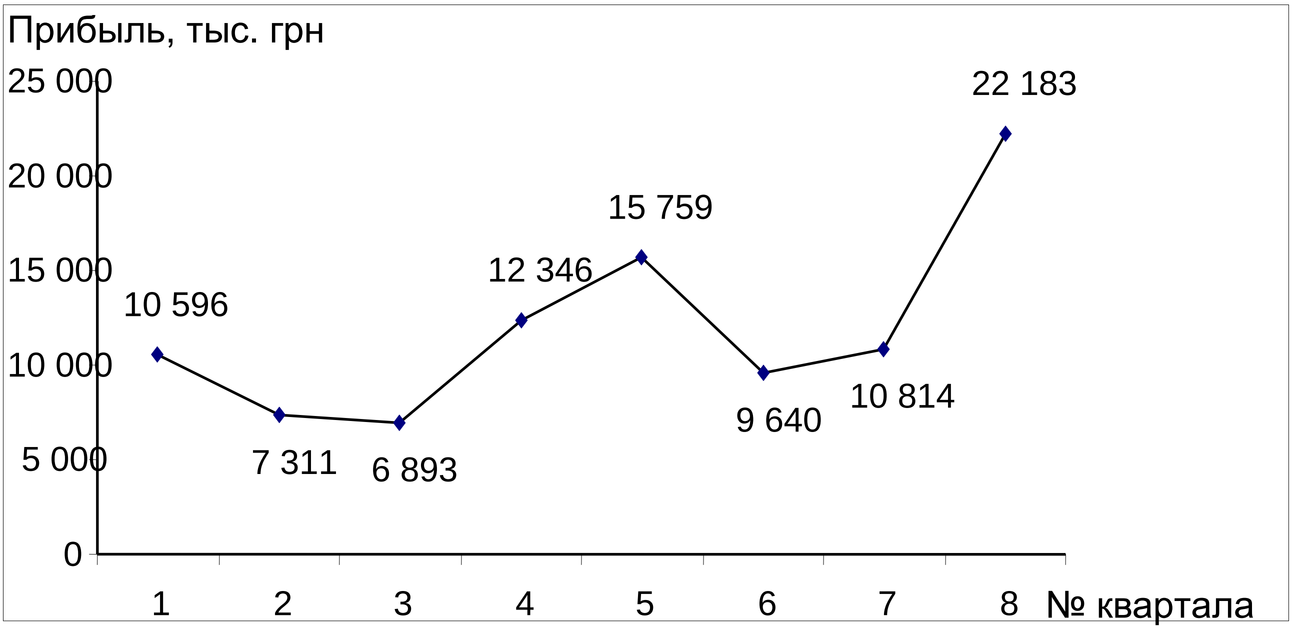


Рис. 12.2. **Динамика прибыли от реализации продукции**

Первый шаг при построении модели – определение тренда.

Для этого используем возможности табличного процессора Microsoft Excel – опцию "Линия тренда".

Построим линейную, логарифмическую, степенную, полиномиальную 2-й степени, экспоненциальную линии тренда. (см. рис. 12.3 – 12.7).

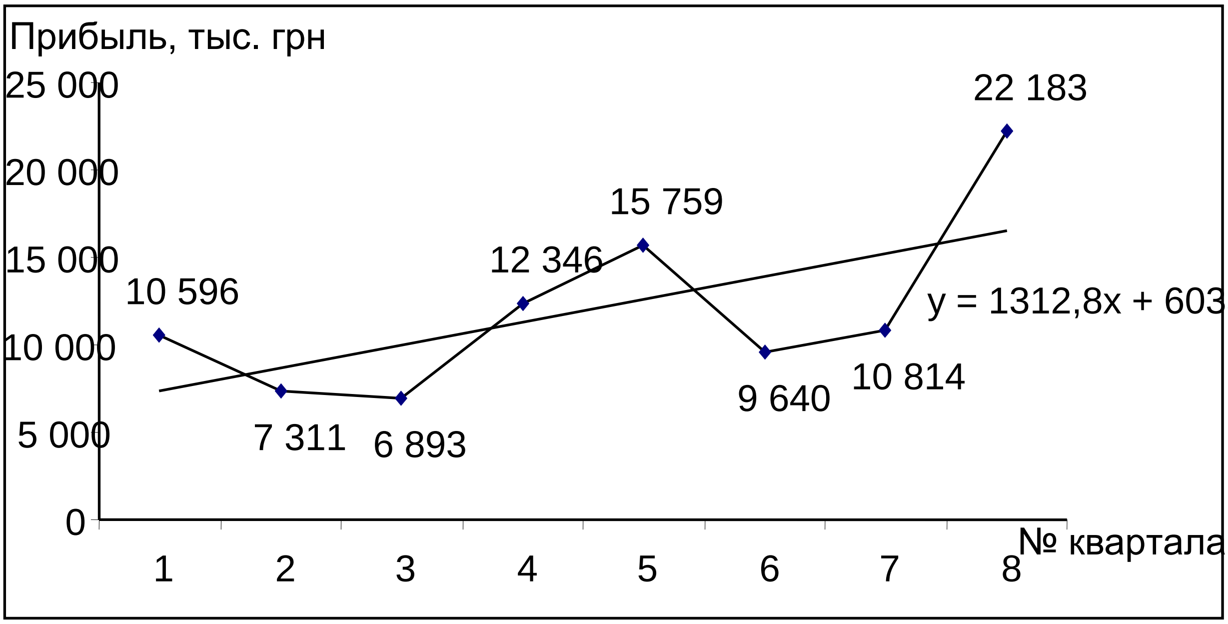


Рис. 12.3. **Линейная линия тренда**

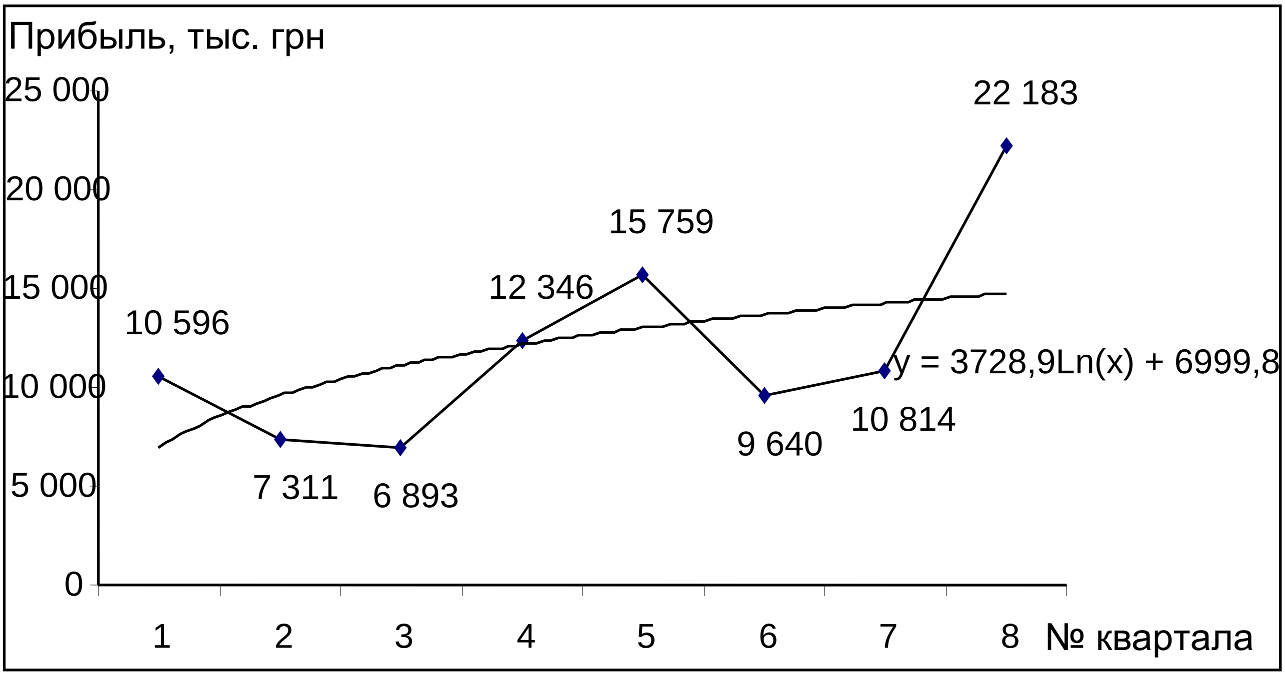


Рис. 12.4. **Логарифмическая линия тренда**

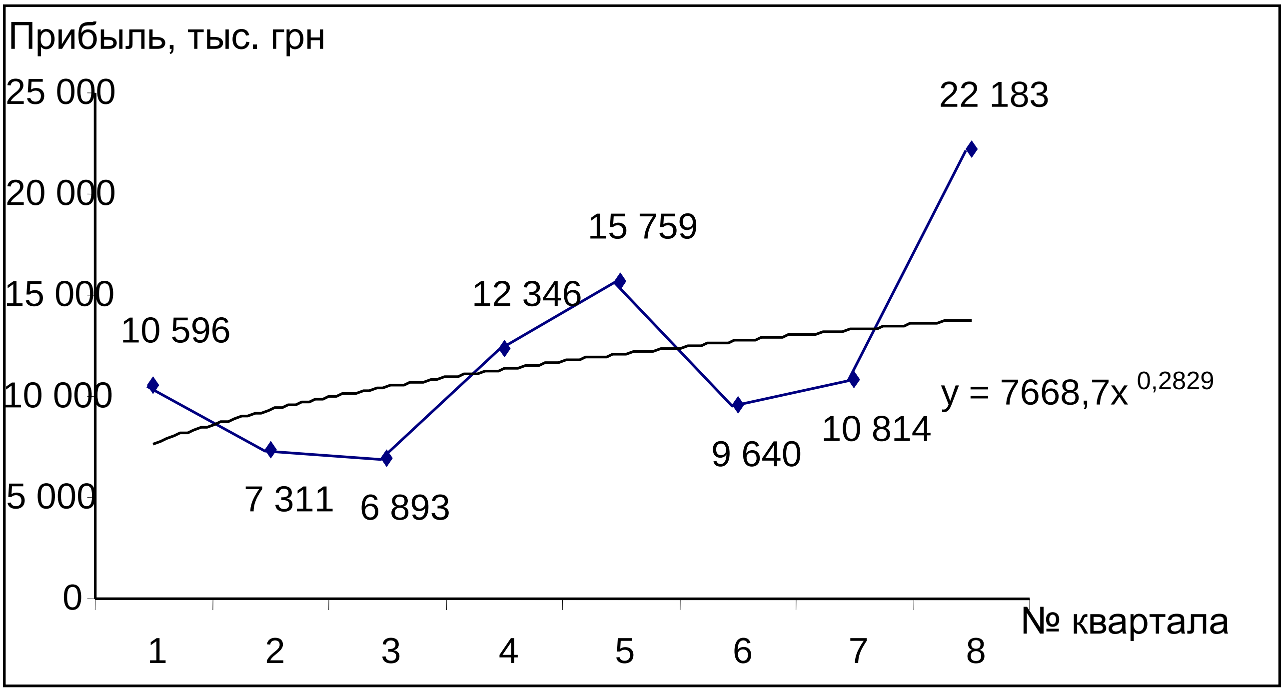


Рис. 12.5. **Степенная линия тренда**

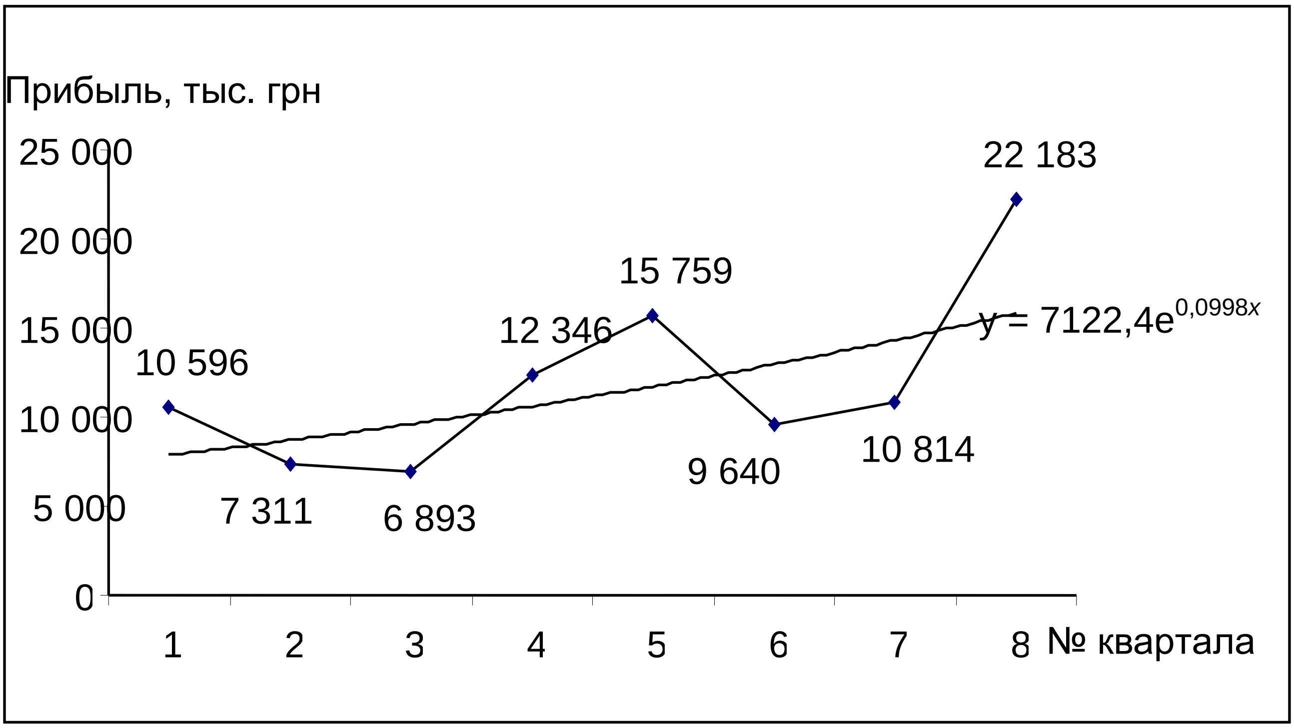


Рис. 12.6. **Экспоненциальная линия тренда**

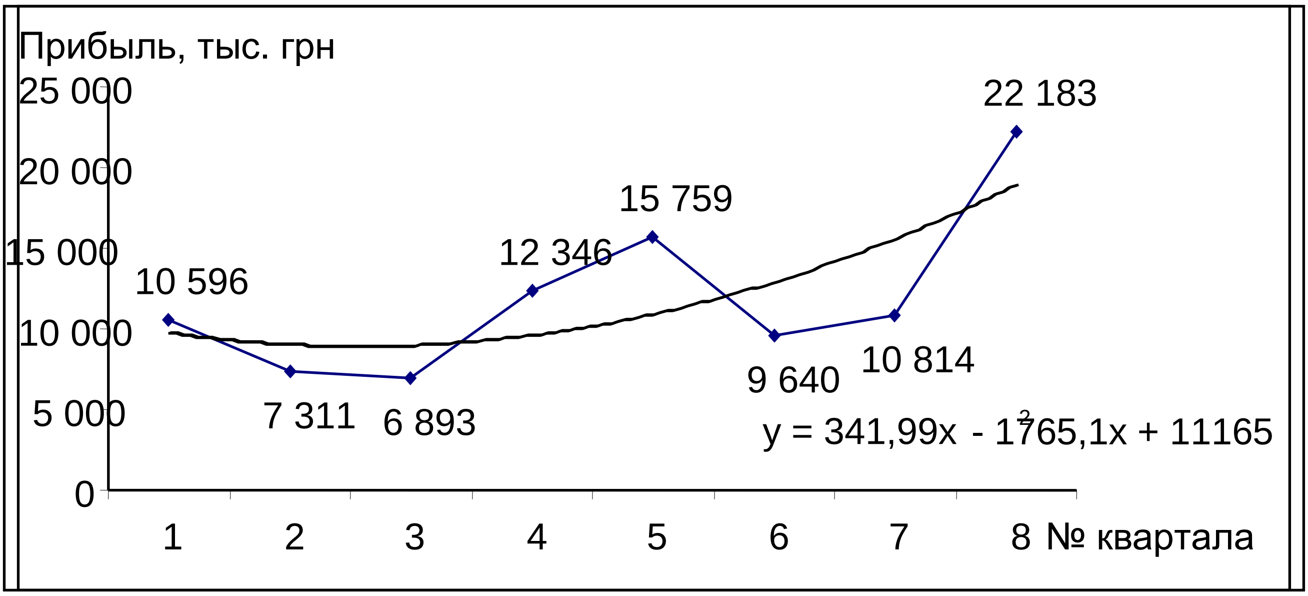


Рис. 12.7. **Полиномиальная 2-й степени линия тренда**

На этом этапе невозможно определить, на основании какого тренда будет наиболее точная модель, поэтому будем создавать пять моделей.

Используя уравнения, которые приведены на графиках, рассчитаем значение тренда для каждого квартала.

Расчеты представлены в табл. 12.8.

Таблица 12.8

Значение тренда (*T*) для каждого из кварталов

№ квартала	Квартал	Тренд (*T*)				
		Полино-миальный	Линей-ный	Логариф-мический	Степен-ной	Экспонен-циальный
1	IV/2008	9 741,89	7 347,8	6 999,8	7 668,7	7 869,895
2	I/2009	9 002,76	8 660,6	9 584,477	9 330,032	8 695,84
3	II/2009	8 947,61	9 973,4	11 096,42	10 464,04	9 608,468
4	III/2009	9 576,44	11 286,2	12 169,15	11 351,27	10 616,88
5	IV/2009	10 889,25	12 599	13 001,23	12 090,95	11 731,12
6	I/2010	12 886,04	13 911,8	13 681,09	12 730,95	12 962,29
7	II/2010	15 566,81	15 224,6	14 255,9	13 298,42	14 322,69
8	III/2010	18 931,56	16 537,4	14 753,83	13810,39	15825,85

Для расчета сезонных компонент найдем разность фактических значений прибыли и значений тренда (табл. 12.9.).

Таблица 12.9

Разности фактических значений прибыли и значений тренда

№ квартала	Квартал	($F - T$) для модели с трендом				
		Полиномиальным	Линейным	Логарифмическим	Степенным	Экспоненциальным
1 сезон						
1	IV/2008	854,11	3 248,2	3 596,2	2 927,3	2 726,105
2	I/2009	–1 691,76	–1 349,6	–2 273,48	–2 019,03	–1 384,84
3	II/2009	–2 054,61	–3 080,4	–4 203,42	–3 571,04	–2 715,47
4	III/2009	2 769,56	1 059,8	176,847	994,7283	1 729,125
2 сезон						
5	IV/2009	4 869,75	3 160	2 757,767	3 668,052	4 027,885
6	I/2010	–3 246,04	–4 271,8	–4 041,09	–3 090,95	–3 322,29
7	II/2010	–4 752,81	–4 410,6	–3 441,9	–2 484,42	–3 508,69
8	III/2010	3 251,44	5 645,6	7 429,17	8 372,611	6 357,149

Фактические данные используются только за два предыдущих сезона, поэтому логически верным будет найти среднее значение этих разностей для каждого квартала сезона.

Чтобы привести сумму к 0, необходимо сумму средних сезонных колебаний поделить на количество периодов в сезоне (это 4) и результат отнять от значения среднего по каждому периоду.

Таким образом, найдем сезонную компоненту. Рассчитанные значения приведены в табл. 12.10.

На этом этапе можно сделать вывод, что модели с полиномиальным, линейным и логарифмическим трендом лучшие, потому что сумма средних сезонных колебаний в них близкая к нулю.

Таблица 12.10

Значение сезонных компонент

Квартал	Для модели с трендом:				
	Полиноми-альным	Линей-ным	Логариф-мическим	Степен-ным	Экспонен-циальным
Средние сезонные колебания					
IV	2 861,93	3 204,1	3 176,983	3 297,676	3 376,995
I	–2 468,9	–2 810,7	–3 157,28	–2 554,99	–2 353,57
II	–3 403,71	–3 745,5	–3 822,66	–3 027,73	–3 112,08
III	3 010,5	3 352,7	3 803,009	4 683,67	4 043,137
Сумма	–0,18	0,6	0,048118	2 398,627	1 954,487
Сезонная компонента (S)					
IV	2 861,975	3 203,95	3 176,971	2 698,019	2 888,373
I	–2 468,86	–2 810,85	–3 157,3	–3 154,65	–2 842,19
II	–3 403,67	–3 745,65	–3 822,67	–3 627,39	–3 600,7
III	3 010,545	3 352,55	3 802,997	4 084,013	3 554,515
Сумма	0	0	0	0	0

В табл. 12.11 рассчитаны значения моделей $F_t = T + S$ для каждого из кварталов.

Для определения качества построенных моделей найдем среднюю ошибку аппроксимации по формуле:

$$\overline{A} = \frac{1}{n}\sum A_i \text{, где } A_i = \left|\frac{F - F_t}{F}\right| \cdot 100\ \%.$$

Если средняя ошибка аппроксимации не превышает 8 – 10 %, то построенная модель качественная.

Таблица 12.11

Значение моделей

№ квар-тала	Квартал	F_t для модели с трендом:				
		Полино-миальным	Линей-ным	Логариф-мическим	Степен-ным	Экспонен-циальным
1	IV/2008	12 603,87	10 551,75	10 176,77	10 366,72	10 758,27
2	I/2009	6 533,905	5 849,75	6 427,18	6 175,385	5 853,651
3	II/2009	5 543,945	6 227,75	7 273,743	6 836,654	6 007,769
4	III/2009	12 586,99	14 638,75	15 972,15	15 435,28	14 171,39
5	IV/2009	13 751,23	15 802,95	16 178,2	14 788,97	14 619,49
6	I/2010	10 417,19	11 100,95	10 523,8	9 576,301	10 120,11
7	II/2010	12 163,15	11 478,95	10 433,23	9 671,033	10 721,99
8	III/2010	21 942,11	19 889,95	18 556,83	17 894,4	19 380,37

Значения, представленные в табл. 12.11 показывают, что качество построенных моделей не очень хорошее. Поэтому нельзя использовать эти модели для прогнозирования (табл. 12.12).

Таблица 12.12

Относительные отклонения и средняя ошибка аппроксимации (%)

№ квар-тала	Квартал	A_i для модели с трендом				
		Полино-миальным	Линей-ным	Логариф-мическим	Степен-ным	Экспонен-циальным
1	2	3	4	5	6	7
1	IV/2008	18,94927	0,41761	3,956479	2,163844	1,531409
2	I/2009	10,62912	19,98701	12,0889	15,53296	19,93365

Окончание табл. 12.12

1	2	3	4	5	6	7
3	II/2009	19,57138	9,651095	5,523625	0,817443	12,84247
4	III/2009	1,951928	18,57079	29,37105	25,02256	14,78528
5	IV/2009	12,7405	0,278888	2,660096	6,15542	7,230863
6	I/2010	8,062085	15,15508	9,168005	0,660777	4,980349
7	II/2010	12,47591	6,148974	3,521061	10,56933	0,850863
8	III/2010	1,085944	10,33697	16,34663	19,33281	12,63415
$\bar{A}$		10,68327	10,0683	10,32948	10,03189	9,348629

Построим модели с использованием тех же самых трендов, но не по аддитивной, а по мультипликативной формуле:

$$F_t = T \cdot S \cdot E .$$

Для расчета сезонных компонент сначала найдем отношение фактических значений прибыли к значениям тренда $\frac{F}{T}$ (табл. 12.13).

Таблица 12.13

Отношение фактических значений прибыли к значениям тренда

№ квар-тала	Квартал	Для модели с трендом				
		Полиноми-альным	Линей-ным	Логариф-мическим	Степен-ным	Экспонен-циальным
1	2	3	4	5	6	7
1 сезон						
1	IV/2008	1,087674	1,442064	1,513758	1,381721	1,346397

Окончание табл. 12.13

1	2	3	4	5	6	7
2	I/2009	0,812084	0,844168	0,762796	0,783599	0,840747
3	II/2009	0,770373	0,691138	0,621192	0,658732	0,717388
4	III/2009	1,289206	1,093902	1,014532	1,087631	1,162866
2 сезон						
5	IV/2009	1,447207	1,250814	1,212116	1,303372	1,343351
6	I/2010	0,748096	0,692937	0,704622	0,75721	0,743695
7	II/2010	0,694683	0,710298	0,758563	0,813179	0,755026
8	III/2010	1,171747	1,341384	1,503542	1,606255	1,401694

В табл. 12.14 рассчитаны средние значения отношений для каждого квартала сезона. Сумма сезонных компонент в данном случае должна равняться 4.

Чтобы свести сумму к 4, необходимо значения за квартал умножить на 4 и поделить на сумму средних сезонных отклонений.

Таким образом, найдем сезонную компоненту.

Дальше рассчитываем значение моделей $F_t = T \cdot S$ для каждого из кварталов. Значения представлены в табл. 12.15.

Таблица 12.14

Значение сезонных компонент

Квартал	Для модели с трендом				
	Полиномиальным	Линейным	Логарифмическим	Степенным	Экспоненциальным
1	2	3	4	5	6
Средние сезонные колебания					
IV	1,267441	1,346439	1,362937	1,342546	1,344874
I	0,78009	0,768552	0,733709	0,770404	0,792221

Окончание табл. 12.14

1	2	3	4	5	6
II	0,732528	0,700718	0,689877	0,735956	0,736207
III	1,230476	1,217643	1,259037	1,346943	1,28228
Сумма	4,010535	4,033352	4,04556	4,195849	4,155582
Сезонная компонента (S)					
IV	1,264111	1,335305	1,347588	1,27988	1,294523
I	0,778041	0,762197	0,725446	0,734444	0,762561
II	0,730604	0,694924	0,682108	0,701604	0,708644
III	1,227244	1,207574	1,244858	1,284072	1,234272
Сумма	4	4	4	4	4

Таблица 12.15

Значение моделей (F_t)

№ квар-тала	Квартал	F_t для модели с трендом				
		Полино-миальным	Линей-ным	Логариф-мическим	Степен-ным	Экспонен-циальным
1	IV/2008	12 314,83	9 811,554	9 432,844	9 815,017	10 187,76
2	I/2009	7 004,517	6 601,084	6 953,022	6 852,388	6 631,109
3	II/2009	6 537,159	6 930,753	7 568,954	7 341,609	6 808,983
4	III/2009	11 752,63	13 628,92	15 148,87	14 575,85	13 104,12
5	IV/2009	13 765,22	16 823,51	17 520,3	15 474,97	15 186,19
6	I/2010	10 025,87	10 603,53	9 924,896	9 350,171	9 884,541
7	II/2010	11 373,17	10 579,94	9 724,067	9 330,219	10 149,69
8	III/2010	23 233,64	19 970,14	18 366,43	17 733,53	19 533,41

Средняя ошибка аппроксимации созданных мультипликативных моделей представлена в табл. 12.16.

Как можем видеть из расчетов, которые приведены в табл. 12.16, для прогнозирования лучше всего взять модель с экспоненциальным трендом.

Таблица 12.16

Относительные отклонения и средняя ошибка аппроксимации (%)

№ квар-тала	Квартал	A_i для модели с трендом				
		Полино-миальным	Линейным	Логариф-мическим	Степен-ным	Экспонен-циальным
1	IV/2008	16,22151	7,403227	10,97731	7,370541	3,852801
2	I/2009	4,192077	9,710242	4,896429	6,272907	9,299564
3	II/2009	5,162355	0,547696	9,806382	6,508176	1,218872
4	III/2009	4,806188	10,39141	22,70265	18,06131	6,140575
5	IV/2009	12,65168	6,754919	11,17648	1,802363	3,634787
6	I/2010	4,002787	9,995162	2,955354	3,00653	2,536734
7	II/2010	5,170818	2,16445	10,07891	13,72092	6,14309
8	III/2010	4,736251	9,975492	17,20495	20,05801	11,94424
$\bar{A}$		7,117958	7,117825	11,22481	9,600094	5,596333

В табл. 12.17 приведены расчеты ошибки E в мультипликативной модели с экспоненциальным трендом.

Таблица 12.17

Ошибка E для модели с экспоненциальным трендом

№ квар-тала	Квартал	Фактические значения F	Трендовые значения T	Сезонная компонента S	Рассчитанные значения $T \cdot S$	Ошибка $E = \dfrac{F}{T \cdot S}$
1	2	3	4	5	6	7
1	IV/2008	10596	7869,895	1,294523	10187,76	1,040072
2	I/2009	7311	8695,84	0,762561	6631,109	1,102531

Окончание табл. 12.17

1	2	3	4	5	6	7
3	II/2009	6893	9608,468	0,708644	6808,983	1,012339
4	III/2009	12346	10616,88	1,234272	13104,12	0,942147
5	IV/2009	15759	11731,12	1,294523	15186,19	1,037719
6	I/2010	9640	12962,29	0,762561	9884,541	0,97526
7	II/2010	10814	14322,69	0,708644	10149,69	1,065452
8	III/2010	22183	15825,85	1,234272	19533,41	1,135644

В табл. 12.18 рассчитаем значение экспоненциального тренда и значение модели для 9 – 12 кварталов.

Таблица 12.18

Значение тренда и модели на 9 – 12 кварталы

№ квартала	Квартал	T	F_t
9	IV/2010	17486,77	22637,02
10	I/2011	19322,01	14734,21
11	II/2011	21349,85	15129,44
12	III/2011	23590,52	29117,12

На рис. 12.8 представлены фактические значения прибыли от реализации за восемь кварталов и значения, рассчитанные с помощью мультипликативной модели с экспоненциальным трендом на 12 кварталов.

Рассчитанные значения модели на 9 – 12 кварталы являются точным прогнозом.

Чтобы предоставить интервальный прогноз, нужно определить доверительные границы. Для этого используем упрощенную формулу.

$F_t \pm S \cdot d$, где $d = 2$, что обеспечит 95 %-ую надежность,

$S=\sqrt{\frac{\sum (F-F_t)^2}{n-2}}$, где n – количество исходных данных.

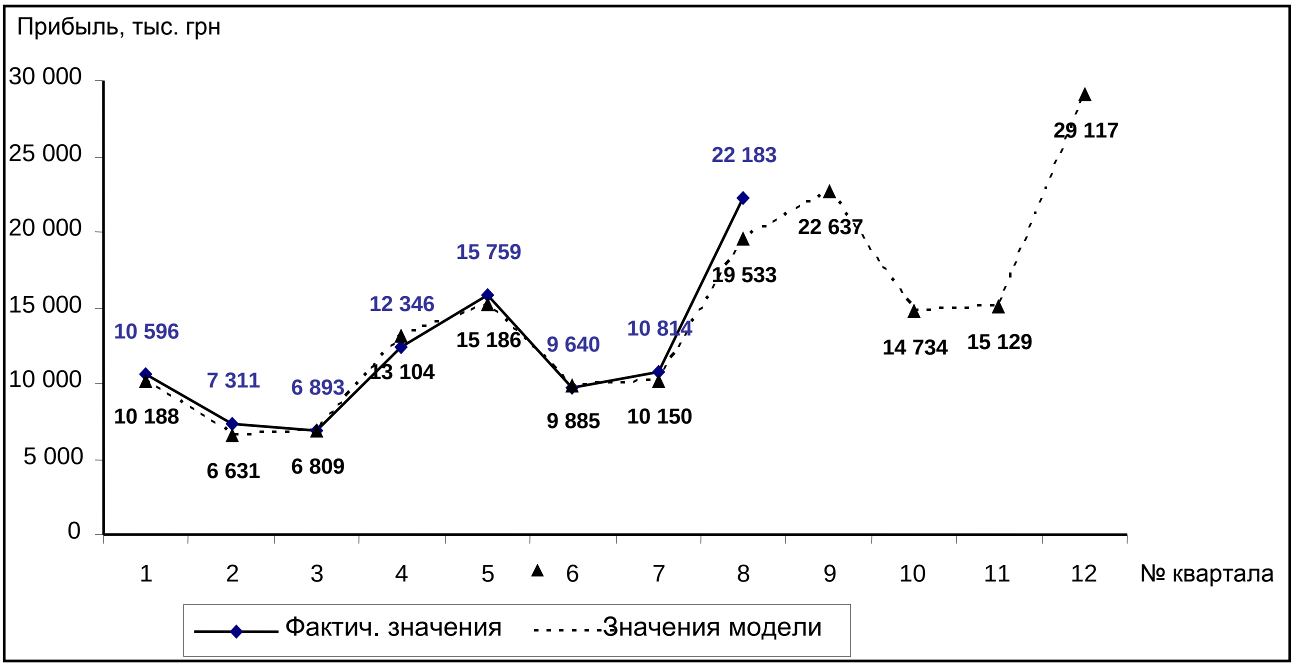


Рис. 12.8. **Фактические значения прибыли и значение модели**

Расчет отклонений значений модели от фактических значений прибыли, их квадратов, S и произведения $S \cdot d$ представлены в табл. 12.19.

Таблица 12.19

Расчет S, $S \cdot d$

№ квартала	Квартал	$F - F_t$	$(F - F_t)^2$
1	IV/2008	408,2428	166 662,2
2	I/2009	679,8911	462 252
3	II/2009	84,01685	7 058,831
4	III/2009	-758,1154	574 739
5	IV/2009	572,8061	328 106,8
6	I/2010	-244,5412	59 800,39
7	II/2010	664,3138	441 312,8
8	III/2010	2 649,59	7 020 327
$\sum (F - F_t)^2$			9 060 259
S			1 228,838
$S \cdot d$			2 457,676

Верхняя и нижняя границы рассчитаны в табл. 12.20.

Таблица 12.20

Расчет нижней и верхней границы

№ квартала	Квартал	Значение модели F_t	Нижняя граница $F_t - S \cdot d$	Верхняя граница $F_t + S \cdot d$
1	IV/2008	10 187,76	7 730,081	12 645,43
2	I/2009	6 631,109	4 173,433	9 088,785
3	II/2009	6 808,983	4 351,307	9 266,659
4	III/2009	13 104,12	10 646,44	15 561,79
5	IV/2009	15 186,19	12 728,52	17 643,87
6	I/2010	9 884,541	7 426,865	12 342,22
7	II/2010	10 149,69	7 692,01	12 607,36
8	III/2010	19 533,41	17 075,73	21 991,09
9	IV/2010	22 637,02	20 179,35	25 094,7
10	I/2011	14 734,21	12 276,53	17 191,89
11	II/2011	15 129,44	12 671,77	17 587,12
12	III/2011	29 117,12	26 659,44	31 574,8

На рис. 12.9 можно увидеть фактические значения выручки за 8 кварталов, рассчитанные значения за 12 кварталов, верхнюю и нижнюю границы. Значение верхней и нижней границы за 9 – 12 кварталы являются интервальным прогнозом.

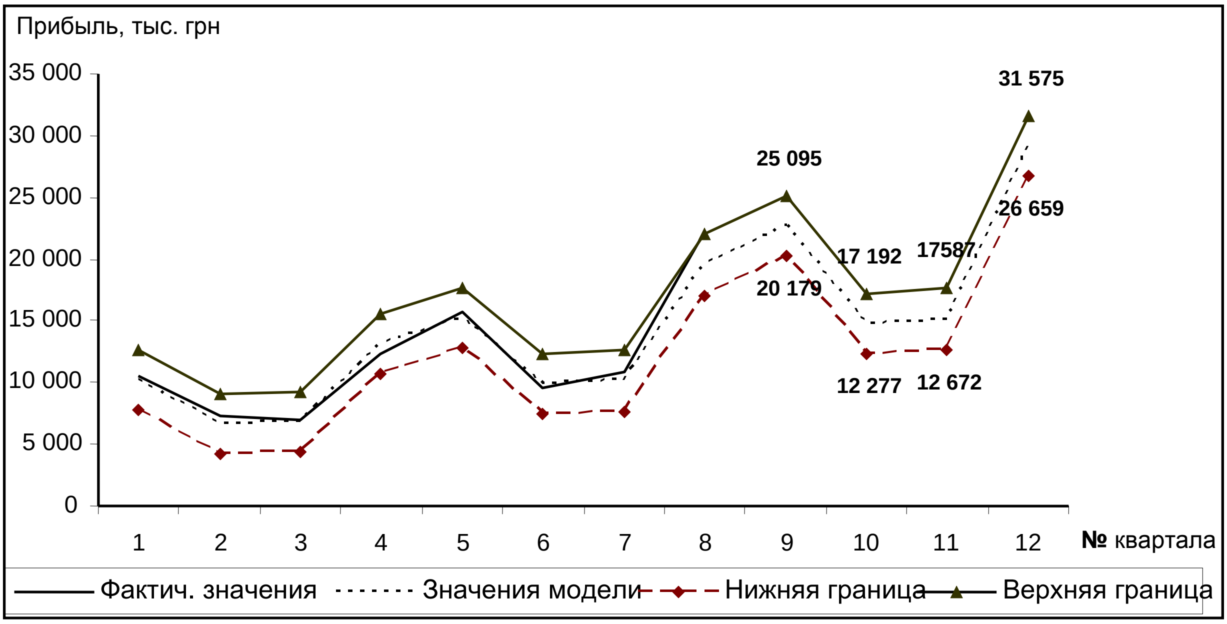


Рис. 12.9. **Фактические значения прибыли и прогноз с 95 %-ми границами**

Таким образом, прогнозные значения составляют:

- на IV/2010 – 20 179 ÷ 25 095 тыс. грн;
- на I/2011 – 12 277 ÷ 17 192 тыс. грн;
- на II/2011 – 12 672 ÷ 17 587 тыс. грн;
- на III/2011 – 26 659 ÷ 31 575 тыс. грн.

По результатам проведенного анализа видно, что в ближайшем периоде предприятие может улучшить результативные показатели, но при этом сезонный характер работы вносит сложность в ритмичную деятельность предприятия.

12.3. Задачи

Задача 12.1. В табл. 12.21 приведены данные, которые характеризуют динамику еженедельного объема реализации продукции предприятия (y_t , тис. грн):

Таблица 12.21

Исходные данные

№ недели	1	2	3	4	5	6	7	8	9	10
y_t	16	10	18	25	17	14	20	22	27	13

Вычислить скользящие средние для значений $m = 3; 5$ и определить общую тенденцию изменения временного ряда графически.

Определить аналитическое выражение тренда (подобрать функцию y_t^*).

Задача 12.2. Изучается зависимость объемов продажи товара A (y_t) от динамики потребительских цен (x_t). Получены данные по 12 кварталам (табл. 12.22).

Известно, что

$$\sum_{t-1}^{12} x_t = 1360; \quad \sum_{t=1}^{12} y_t = 470; \quad \sum_{t=1}^{12} x_t y_t = 54\ 945; \quad \sum_{t-1}^{12} x_t^2 = 20\ 320.$$

Построить модель зависимости объемов продажи товара A от индекса потребительских цен с включением фактора времени ($Y = a_1 x_t + a_2 t$) и интерпретировать параметры модели.

Таблица 12.22

Исходные данные

Квартал	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Индекс потребитель-ских цен, в % к I кварталу	100	102	104	110	112	112	114	117	120	124	124	130
Средний за день объем продажи товара A, кг	20	24	22	30	35	37	45	47	46	50	54	60

Задача 12.3. Имеются данные об уровне безработицы (% к населению работоспособного возраста) в Украине за 12 месяцев 2013 года (табл. 12.23).

Таблица 12.23

Исходные данные

Месяц	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
y_t	2,0	2,1	2,0	1,9	1,8	1,7	1,6	1,6	1,5	1,4	1,5	1,8

Необходимо:

1. Определить коэффициенты автокорреляции уровней этого ряда первого и второго порядка.
2. Обосновать выбор уравнения тренда и определить его параметры.
3. Интерпретировать полученные результаты.

Задача 12.4. По значениям ряда, представленного в табл. 12.24, построить модель декомпозиции временного ряда, выделить такие составляющие: тренд, циклическую, сезонную и случайную.

Построить прогноз на 2009 год по кварталам с учетом трендовой, циклической и случайной составляющей.

Таблица 12.24

Исходные данные

1997				1998				1999				2000			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
23,4	33,7	44,1	20,4	34,4	44,2	61,5	33,5	44,9	64,4	82,1	38,4	34,4	44,2	61,5	33,5

2001				2002				2003				2004			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
44,9	64,4	82,1	38,4	60,5	85,5	108	50,8	76	116	223	102	147	218	273	120

2005				2006				2007			
1	2	3	4	1	2	3	4	1	2	3	4
114	177	223	102	147	218	273	120	170	264	332	153

Задача 12.5. В эконометрической модели изучается зависимость заработной платы W_t (тыс. \$) от прибыли X_t (тыс. \$) по данным за 40 лет. Оценка параметров модели дала такие результаты:

$$W_t = -4{,}79 + 0{,}237 X_t + 0{,}201 X_{t-1} + 0{,}229 X_{t-2} - 0{,}512 t + e_t;$$
$$(-3{,}87)(2{,}10) \qquad (1{,}49) \qquad (2{,}32) \qquad (-4{,}79)$$
$$R^2 = 0{,}9689; \quad DW = 0{,}4919.$$

В скобках указаны фактические значения критерия Стьюдента для коэффициентов регрессии, t – переменная фактора времени.

Проанализировать полученные результаты регрессионного анализа: определить краткосрочный и долгосрочный мультипликаторы, охарактеризовать структуру лага. Перечислить основные эконометрические проблемы, которые возникают при построении моделей с распределенным лагом.

Задача 12.6. С целью прогнозирования объемов экспорта страны на будущие периоды были собраны данные по 30 годам таких показателей: Y_t – объемы экспорта (млрд грн);

x_t – индекс объемов промышленного производства (в % к предыдущему году).

Получены такие результаты обработки исходных данных:
уравнение линейных трендов:

$$Y_t = 3{,}1 + 1{,}35t + e_t; \quad R^2 = 0{,}91; \quad DW = 2{,}31;$$

$$X_t = -8{,}4 + 4{,}8t + e_t; \quad R^2 = 0{,}89; \quad DW = 2{,}08;$$

уравнение регрессии по уровням временных рядов:

$$Y_t = -10{,}5 + 0{,}5x_t + e_t; \quad R^2 = 0{,}95; \quad DW = 2{,}21;$$

уравнение регрессии по первым разностям уровней временных рядов:

$$\Delta Y_t = 1{,}4 + 0{,}03\Delta x_t + e_t; \quad R^2 = 0{,}86; \quad DW = 2{,}25;$$

уравнение регрессии по вторым разностям уровней временных рядов:

$$\Delta^2 Y_t = 0{,}7 + 0{,}012\,\Delta^2 x_t + e_t; \quad R^2 = 0{,}47; \quad DW = 2{,}69;$$

уравнение регрессии по уровням временных рядов с включением фактора времени:

$$Y_t = 4{,}23 + 0{,}24x_t + 0{,}78t + e_t; \quad R^2 = 0{,}97; \quad DW = 0{,}9.$$

Проверить гипотезы относительно величины коэффициента автокорреляции. Обосновать ответ. Выбрать лучшее уравнение регрессии, которое можно использовать для прогнозирования объемов экспорта, и дать интерпретацию его параметров.

Задача 12.7. В табл. 12.25 приведены статистические данные процентов изменений заработной платы (Y), рост производительности работы (x_1) и уровня инфляции (x_2) за 20 лет.

Таблица 12.25

Исходные данные

Y	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,4
x_1	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	3,5
x_2	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	6,3
Y	12,0	12,5	8,5	5,9	6,8	5,6	4,8	6,7	5,5	4,0
x_1	5,0	2,3	1,5	6,0	2,9	2,8	2,6	0,9	0,6	0,7
x_2	6,1	6,9	7,1	3,1	3,7	3,9	3,9	4,8	4,3	4,8

С помощью МНК построить уравнение регрессии:

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + e_t \text{ и } y_t = c_0 + c_1 x_{1t-1} + c_2 x_{2t-1} + \upsilon_t,$$

($x_{10} = 3{,}5;\ \ x_{20} = 4{,}5$).

Оценить качество уравнений. Сравнить построенные модели. Какая из них имеет преимущества и почему?

Задача 12.8. В табл. 12.26 приведены данные о прибылях хозяйств (X) и затраты хозяйств на розничные закупки (Y) за 18 лет.

Таблица 12.26

Исходные данные

X	9,098	9,137	9,095	9,280	9,230	9,348	9,525	9,755	10,28
Y	5,49	5,54	5,305	5,505	5,42	5,32	5,54	5,69	5,87
X	11,02	12,055	12,088	12,215	12,495	11,305	11,43	11,45	11,697
Y	6,342	6,48	6,395	6,555	6,755	5,905	6,125	6,185	6,225

Оценить параметры уравнений регрессии $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ и $y_t = \beta_0 + \beta_1 x_t + \gamma y_{t-1} + \varepsilon_t$. Оценить качество построенных моделей и сравнить результаты. Какая из моделей имеет преимущества?

Задача 12.9. Получены данные, которые отображают динамику изменений объемов продажи промышленной продукции за период 15-ти месяцев (табл. 12.27). При условии линейности найти уравнение тренда и оценить его значимость на уровне $\alpha = 0{,}05$. Установить с помощью критерия Дарбин – Уотсона наличие или отсутствие автокорреляции остатков на 5 %-м уровне значимости.

Таблица 12.27

Исходные данные

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y_i	187	331	408	408	395	426	455	506	505	528	544	590	605	629	753

Задача 12.10. Имеются поквартальные данные по розничному товарообороту Украины за 2009 – 2013 гг. (табл. 12.28).

Необходимо:

1. Построить график временного ряда.
2. Построить аддитивную и мультипликативную модели временного ряда.
3. Оценить качество модели через показатели средней абсолютной ошибки и среднего относительного отклонения.

Таблица 12.28

Статистические данные

Номер квартала	Товарооборот, % к предыдущему периоду	Номер квартала	Товарооборот, % к предыдущему периоду
1	86	11	114,6
2	80,3	12	113,7
3	79,3	13	112,6
4	79,4	14	114,3
5	96,9	15	114,2
6	102,5	16	113,7
7	105,2	17	110,8
8	107,8	18	107,4
9	112,3	19	106,2
10	115,0	20	105,6

Задача 12.11. Имеются данные об экспорте и импорте Германии млрд долл.) за 12 лет (табл. 12.29).

Таблица 12.29

Статистические данные

Номер года	Экспорт	Импорт	Номер года	Экспорт	Импорт
1	184	158	7	403	390
2	243	191	8	422	402
3	294	228	9	382	346
4	323	280	10	430	385
5	341	270	11	524	464
6	410	346	12	521	456

1. По каждому ряду постройте тренды и выберите лучший из них.
2. Постройте уравнение регрессии и оцените тесноту и силу связи двух рядов (по отклонениям от тренда и по множественной регрессионной модели с включенным в нее фактора времени).
3. Выполните прогноз уровней одного ряда, исходя из его связи с уровнями другого ряда.

Глоссарий

А

Автокорреляция – явление взаимосвязи между рядами данных: первоначальным и этим же рядом, сдвинутым относительно первоначального положения на несколько моментов времени.

Авторегрессия – разновидность динамической эконометрической модели, которая содержит в качестве факторных переменных лаговые значения эндогенных переменных.

Аддитивная модель временного ряда – модель, в которой все компоненты ряда динамики представлены как суммы этих составляющих.

Б

Бета-коэффициент – коэффициент эконометрической модели в стандартизованных переменных, который показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднего квадратического отклонения.

В

Верификация модели – проверка истинности модели, определение соответствия построенной модели реальному экономическому явлению.

Временной лаг – сдвиг уровней временного ряда относительно первоначального положения на несколько моментов времени.

Временной ряд – упорядоченная во времени последовательность показателей, характеризующих уровни развития изучаемого явления в последовательные моменты или периоды времени.

Временные данные – набор сведений, характеризующих один и тот же объект за разные периоды времени.

Г

Гетероскедастичность – нарушение условия постоянства дисперсий случайных отклонений.

Гомоскедастичность – условие постоянства дисперсий случайных отклонений.

Д

Двухшаговый метод наименьших квадратов – способ решения систем одновременных уравнений.

Доверительный интервал – интервал, в который истинное значение параметра попадает с определенной вероятностью.

И

Идентификация модели – проведение статистического анализа модели и оценивания качества ее параметров; установление соответствия межу приведенной и структурной формами модели.

Идентифицируемая модель – разновидность структурной модели системы одновременных уравнений, в которой все структурные коэффициенты однозначно определяются через приведенные коэффициенты.

К

Ковариация – характеристика сопряженности вариации двух признаков, статистическая мера взаимодействия двух случайных величин.

Корреляционная зависимость – связь, при которой каждому значению независимой переменной соответствует определенное среднее значение зависимой переменной.

Корреляционный анализ – количественное определение тесноты связи между результативным и множеством факторных признаков.

Корреляция – статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Косвенный метод наименьших квадратов – метод решения системы одновременных уравнений, основанный на получении состоятельных и несмещенных оценок параметров структурной формы модели по оценкам параметров приведенной формы.

Коэффициент детерминации – показывает, какая доля вариации результативного признака учтена в модели и обусловлена влиянием на нее факторных признаков.

Коэффициент линейной регрессии – показывает, на какую величину в среднем изменится результативный признак, если соответствующую факторную переменную увеличить на единицу измерения.

Л

Лаги Алмон – вид модели с распределенным лагом, который характеризуется полиномиальной структурой и конечной величиной лага.

Лаговые переменные – переменные модели, которые датируются предыдущими моментами времени и находятся в уравнении модели с текущими переменными.

М

Метод наименьших квадратов (МНК) – метод, позволяющий получить такие оценки параметров линейной регрессии, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}_x$ минимальна.

Множественная зависимость – связь между несколькими факторными признаками и результативным признаком.

Множественная корреляция – зависимость между результативным признаком и двумя или более факторными признаками, включенными в исследование.

Мультиколлинеарность – тесная зависимость между факторными признаками, включенными в модель.

Н

Неидентифицируемая модель – разновидность структурной модели системы одновременных уравнений, в которой структурные коэффициенты невозможно найти по приведенным коэффициентам.

О

Обобщенный метод наименьших квадратов – метод, позволяющий получить такие оценки параметров линейной регрессии, при которых минимальна обобщенная остаточная дисперсия.

П

Параметризация – выбор общего вида модели, в том числе состава и формы входящих в нее связей, формулирование исходных предпосылок и ограничений модели.

Парная регрессия – связь между двумя признаками: результативным и факторным.

Приведенная форма модели – способ записи системы одновременных уравнений, в котором каждая эндогенная переменная определена в виде линейной функции.

Пространственные данные – набор сведений по разным объектам, взятым за один и тот же период времени.

Р

Регрессионный анализ – определение аналитической формы связи, в которой изменение результативного признака обусловлено влиянием факторных признаков.

Ряд динамики – ряд последовательно расположенных статистических показателей, изменение которых имеет определенную тенденцию развития изучаемого явления.

С

Сверхидентифицируемая модель – разновидность структурной модели системы одновременных уравнений, в которой структурные коэффициенты, выраженные через приведенные коэффициенты, имеют два и более числовых значений.

Система независимых уравнений – разновидность систем эконометрических уравнений, в которой каждый результативный признак является функцией одной и той же совокупности факторов.

Система одновременных уравнений – разновидность эконометрических моделей, в которой результативный признак одного уравнения системы в каждом последующем уравнении является фактором наряду с одной и той же совокупностью факторов.

Структурная форма модели – способ записи системы одновременных уравнений, который отражает реальный экономический объект или явление и показывает, как изменение любой экзогенной переменной определяет значения эндогенной переменной модели.

У

Условия Гаусса – Маркова – предпосылки, выполнение которых необходимо для использования МНК.

Ф

Функциональная зависимость – связь, при которой каждому значению независимой переменной соответствует точно определенное значение зависимой переменной.

Ч

Частная корреляция – зависимость между результативным и одним или двумя факторными признаками при фиксированном значении других факторных признаков.

Э

Экзогенные переменные – признаки или показатели, значения которых задаются извне модели и они являются управляемыми или планируемыми (независимые).

Эконометрика – наука, позволяющая на базе положений экономической теории и результатов экономических измерений придавать конкретные количественные выражения общим (количественным) закономерностям, обусловленным экономической теорией.

Эндогенные переменные – переменные, которые формируются в процессе функционирования социально-экономической системы и зависят от экзогенных переменных или от их взаимодействия (зависимые).

Рекомендованная литература

Айвазян С. А. Прикладная статистика и основы эконометрики : учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998. – 1022 с.

Бородич С. А. Эконометрика : учебн. пособ. / С. А. Бородич. – Мн. : Новое знание, 2001. – 408 с.

Доугерти К. Введение в эконометрику / К. Доугерти ; пер. с англ. – М. : ИНФРА-М, 1999. – 402 с.

Елисеева И. И. Практикум по эконометрике : учебн. пособ. / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко. – М. : Финансы и статистика, 2006. – 344 с.

Магнус Я. Р. Економетрика: начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 1997. – 248 с.

Малярец Л. М. Экономико-математические методы и модели : учебн. пособ. для иностр. студентов / Л. М. Малярец. – Х. : Изд. ХНЭУ, 2013. – 288 с.

Методические рекомендации к решению индивидуальных заданий по дисциплине "Экономико-математические методы и модели" для иностранных студентов отрасли знаний 0305 "Экономика и предпринимательство" всех форм обучения / Э. Ю. Железнякова, Л. А. Норик – Х. : ХНЭУ, 2012. – 80 с.

Новиков А. И. Эконометрика : учебн. пособ. / А. И. Новиков. – М. : ИНФРА-М, 2007. – 144 с.

Практикум по эконометрике : учебн. пособ. / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко и др. – М. : Финансы и статистика, 2007. – 344 с.

Практикум по эконометрии в Excel : учебн. пособ. для экономических вузов / А. А. Егоршин, Л. М. Малярец – Х. : ИД "ИНЖЕК", 2005. – 100 с.

Просветов Г. И. Эконометрика. Задачи и решения : учебн.-метод. пособ. / Г. И. Просветов. – М. : Издательство РДЛ, 2004. – 104 с.

Тихомиров Н. П. Эконометрика : учебник / Н. П. Тихомиров, Е. Ю. Дорохина . – М. : Изд. "Экзамен", 2003. – 512 с.

Христиановский В. В. Прикладная эконометрия : учебн. для экон. вузов / В. В. Христиановский, Н. Г. Гузь, О. Г. Кривенчуг. – Донецк : ДонГУ, 1998. – 172 с.

Шанченко Н. И. Лекции по эконометрике : учебн. пособ. для студентов высших учебных заведений, обучающихся по специальности "Прикладная информатика (в экономике) / Н. И. Шанченко. – Ульяновск : УлГТУ, 2008. – 139с.

Ферстер Э. Методы корреляционного и регрессионного анализа : руководство для экономистов / Э. Ферстер, Б. Ренц ; пер. с нем. и предисл. В. М. Ивановой. – М. : Финансы и статистика, 1983. – 302 с.

Эконометрика : учебник / под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2004. – 344 с.

Экономико-математические методы и модели : учебн. пособ. / под общ. ред. А. В. Кузнецова. – Мн. : БГЭУ, 1999. – 416 с.

Экономико-математические методы и прикладные модели : учебн. пособ. для вузов / под ред. В. В. Федосеева. – М. : ЮНИТИ, 1999. – 392 с.

Приложения

Приложение А

Критические точки распределения *F* Фишера – Снедекора

(k_1 – число степеней свободы большей дисперсии, k_2 – число степеней свободы меньшей дисперсии) при уровне значимости $\alpha = 0{,}05$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,13	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,16	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,!0	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18
30	4,!7	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,!2	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,!8	2,12	2,07	2,04	2,00
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,93	1,89
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,83	1,80
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,78
∞	3,84	2,99	2,60	2,37	2,21	1,09	2,01	1,94	1,88	1,83	1,79	1,75

Критические точки распределения *F* Фишера – Снедекора

(k_1 – число степеней свободы большей дисперсии, k_2 – число степеней свободы меньшей дисперсии) при уровне значимости $\alpha = 0{,}01$

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,4	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,6	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,4	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	9,9	9,5	9,6
6	13,7	10,9	9,8	9,2	8,8	8,5	8,1	7,7	7,3	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,5	6,1	5,7
8	11,3	8,7	7,6	7,0	6,6	6,5	6,0	5,7	5,3	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,1	4,7	4,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,1	4,7	4,3	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,4	4,0	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,2	3,8	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,0	3,6	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,8	3,4	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,7	3,3	2,9
16	8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,6	3,2	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,5	3,1	2,7
18	8,3	6,0	5,1	4,6	4,3	4,0	3,7	3,4	3,0	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,3	2,9	2,4
20	8,1	5,9	4,9	4,4	4,1	3,9	3,6	3,2	2,9	2,4
22	7,9	5,7	4,8	4,3	4,0	3,8	3,5	3,1	2,8	2,3
24	7,8	5,6	4,7	4,2	3,9	3,7	3,3	3,0	2,7	2,2
26	7,7	5,5	4,6	4,1	3,8	3,6	3,3	3,0	2,6	2,1
28	7,6	5,5	4,6	4,1	3,8	3,5	3,2	2,9	2,5	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,2	2,8	2,5	2,0
40	7,3	5,2	4,3	3,8	3.5	3,3	3,0	2,7	2,3	1,8
60	7,1	5,0	4,1	3,7	3,3	3,1	2,8	2,5	2,1	1,6
120	6,9	4,8	4,0	3,5	3,2	3,0	2,7	2,3	2,0	1,4
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,2	1,8	1,0

Критические точки распределения Стьюдента

Уровень значимости α (двухсторонняя критическая область)				
Число степеней свободы k	$\alpha = 0{,}01$	$\alpha = 0{,}02$	$\alpha = 0{,}05$	$\alpha = 0{,}1$
1	63,7	31,82	12,7	6,31
2	9,92	6,97	4,30	2,92
3	5,84	4,54	3,18	2,35
4	4,60	3,75	2,78	2,13
5	4,03	3,37	2,57	2,01
6	3,71	3,14	2,45	1,94
7	3,50	3,00	2,36	1,89
8	3,36	2,90	2,31	1,86
9	3,25	2,82	2,26	1,83
10	3,17	2,76	2,23	1,81
11	3,11	2,72	2,20	1,80
12	3,05	2,68	2,18	1,78
13	3,01	2,65	2,16	1,77
14	2,98	2,62	2,14	1,76
15	2,95	2,60	2,13	1,75
16	2,92	2,58	2,12	1,75
17	2,90	2,57	2,11	1,74
18	2,88	2,55	2,10	1,73
19	2,86	2,54	2,09	1,73
20	2,85	2,53	2,09	1,73
21	2,83	2,52	2,08	1,72
22	2,82	2,51	2,07	1,72
23	2,81	2,50	2,07	1,71
24	2,80	2,49	2,06	1,71
25	2,79	2,49	2,06	1,71
26	2,78	2,48	2,06	1,71
27	2,77	2,47	2,05	1,71
28	2,76	2,46	2,05	1,70
30	2,75	2,46	2,04	1,70
40	2,70	2,42	2,02	1,68
60	2,66	2,39	2,00	1,67
120	2,62	2,36	1,98	1,66
∞	2,58	2,33	1,96	1,64
k	$\alpha = 0{,}005$	$\alpha = 0{,}01$	$\alpha = 0{,}025$	$\alpha = 0{,}05$
Уровень значимости α (односторонняя критическая область)				

Критические значения критерия Дарбина – Уотсона DW
(при уровне значимости $\alpha = 0{,}05$)

n	Количество независимых переменных											
	m=1		m=2		m=3		m=4		m=5		m=6	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40										
7	0,70	1,37	0,47	1,9								
8	0,76	1,33	0,56	1,78	0,37	2,29						
9	0,82	1,32	0,63	1,70	0,46	2,13	0,30	2,59				
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82		
11	0,93	1,32	0,76	1,60	0,60	1,93	0,44	2,28	0,32	2,65	0,12	2,89
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51	0,16	2,67
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,39	0,21	2,49
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30	0,26	2,35
15	1,08	1,36	0,95	1,54	0,81	1,75	0,69	1,98	0,56	2,22	0,30	2,24
16	1,11	1,37	0,98	1,54	0,86	1,73	0,73	1,94	0,62	2,16	0,35	2,15
18	1,16	1,39	1,05	1,54	0,93	1,70	0,82	1,87	0,71	2,06	0,44	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,89	1,83	0,79	1,99	0,52	1,92
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94	1,57	1,85
25	1,29	1,45	1,21	1,55	1,12	1,65	1,04	1,77	0,95	1,89	1,68	1,78
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85	0,76	1,73
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82	0,86	1,69
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80	0,91	1,67
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79	1,00	1,64
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78	1,07	1,64
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77	1,12	1,64
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77	1,21	1,64
70	1,58	1,64	1,55	1,67	1,53	1,70	1,49	1,74	1,46	1,77	1,28	1,65
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77	1,34	1,65
90	1,64	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78	1,38	1,66
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78	1,42	1,67
150	1,72	1,75	1,71	1,76	1,69	1,77	1,68	1,79	1,67	1,80	1,54	1,71
200	1,76	1,78	1,75	1,79	1,74	1,80	1,73	1,81	1,72	1,82	1,61	1,74

Приложение Е

Критические точки распределения χ^2 в зависимости от уровня значимости и числа степеней свободы

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,362	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,7	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,76	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,(1	7,63
20	37,6	34,2	31.4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

СОДЕРЖАНИЕ

НАВЧАЛЬНЕ ВИДАННЯ

Малярець Людмила Михайлівна
Железнякова Еліна Юріївна
Норік Лариса Олексіївна

ЕКОНОМЕТРИКА

в прикладах та задачах

Навчальний посібник для іноземних студентів

(рос. мовою)

Відповідальний за випуск **Малярець Л. М.**

Відповідальний редактор **Оленич М. М.**

Редактор **Бутенко В. О.**

Коректор **Маркова Т. А.**

Викладено матеріал навчальної дисципліни, яка викладається бакалаврам галузі знань 0305 "Економіка і підприємництво". Викладено матеріал, який супроводжується великою кількістю прикладів, що мають як навчальний, так і дослідницький характер. Наведено реальні приклади економіки суб'єктів господарювання різних рівнів управління.

Рекомендовано для студентів, магістрів, аспірантів, які вивчають економетричні моделі.

План 2014 р. Поз. № 22-П.
Підп. до друку 24.12.2014 р. Формат 60 х 90 1/16. Папір MultiCopy. Друк Riso.
Ум.-друк. арк. 16,75. Обл.-вид. арк. 20,94. Тираж 400 прим. Зам. № 337.

Видавець і виготівник – видавництво ХНЕУ ім. С. Кузнеця, 61166, м. Харків, пр. Леніна, 9-А

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
Дк № 481 від 13.06.2001 р.