МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Государственное образовательное учреждение
высшего профессионального образования
«Оренбургский государственный университет»

А.Г. РЕННЕР, О.И. СТЕБУНОВА, Л.М. ТУКТАМЫШЕВА

ОСНОВЫ ЭКОНОМЕТРИКИ

Рекомендовано Ученым советом государственного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет» в качестве учебного пособия для студентов, обучающихся по программам высшего профессионального образования по экономическому направлению

Оренбург 2009

УДК 330.43 (075.8)
ББК 65в631я73
Р 39

Рецензент
доктор экономических наук, профессор Е.М. Дусаева

Реннер, А.Г.
Р 39 Основы эконометрики: учебное пособие/ А.Г. Реннер, О.И. Стебунова, Л.М. Туктамышева. – Оренбург: ГОУ ОГУ, 2009. – 156 с.
ISBN

В пособии рассмотрены основные разделы курса «Эконометрика», представлены контрольные задания и ответы к ним, индивидуальные задания по всем разделам курса и методические указания для их выполнения, контрольные задания, к которым в конце пособии приведены ответы. Изложение материала направлено на то, чтобы обеспечить понимание сущности каждой процедуры эконометрического моделирования.

Пособие предназначено для студентов экономических специальностей, изучающих дисциплину «Эконометрика».

ББК 65в631я73

Р $\frac{0601000000}{}$

ISBN

Содержание

Введение. Основные понятия и этапы эконометрического моделирования

Предметом изучения эконометрики является количественное описание (или выявление) закономерностей, обусловленных экономической теорией, методами математической статистики, на основе данных экономической статистики.

Роль экономической теории в рамках эконометрики заключается в том, что она, характеризуя на качественном уровне объективно существующие взаимосвязи между экономическими показателями, «ставит» задачу их формализации.

Роль экономической статистики - в информационном обеспечении эконометрических моделей, т.е. в разработке и измерении показателей необходимых для описания закономерностей, связей.

Роль математической статистики – в предоставлении математического инструментария (многомерного корреляционного анализа, многомерного регрессионного анализа, методов многомерной классификации, методов снижения размерности признакового пространства, методов анализа случайных процессов) для обработки информации, предоставляемой экономической статистикой, с целью математической формализации (описания) объективно существующих и возможно, предлагаемых экономической теорией закономерностей или связей.

Основная цель начального курса эконометрики – ознакомить студентов с математическими и инструментальными средствами решения ее задач, дать навыки решения типичных из них.

К примеру, сформулировано следующее положение экономической теории: объем инвестиций есть возрастающая функция национального дохода и убывающая функция характеристики государственного регулирования (норма процента). Наша первая задача перевести это положение на математический язык. Первоначально выбирают форму взаимосвязи - допустим, что она простейшая (линейная)

$$y_{1,t} = \beta_0 y_{2,t-1} + \beta_1 x_{1,t} + \delta_t, \tag{1}$$

где $y_{1,t}$ – объем инвестиций в момент времени t;

$y_{2,t\text{-}1}$ – национальный доход, естественно, в предшествующий момент времени (t-1);

$x_{1,t}$ – норма процента в момент времени t;

δ_t – случайная величина, характеризующая влияние на объем инвестиций других неучтенных факторов.

Дополнительная информация, вытекающая из формулировки положений экономической теории: $\beta_0>0$, т.к. инвестиции возрастающая функция национального дохода; $\beta_1<0$, т.к. инвестиции убывающая функция нормы процента.

В задачу экономической статистики входит измерение показателей $y_{1,t}$, $y_{2,t}$, $x_{1,t}$ за некоторый промежуток времени и приведение полученных данных к сопоставимому виду.

И, наконец, методами математической статистики мы должны исследовать возможность получения оценки неизвестных коэффициентов, с учетом наложенных ограничений и предложить процедуры (способы) оценивания.

Вопросы, возникающие после оценки параметров и свойств δ_t, обсудим позднее, а сейчас сделаем важное замечание, касающееся локальных свойств предложенной модели. Очевидно, что доля национального дохода направляемого на инвестиции, характеризуемая β_0, не может не зависеть от величины D_t, поэтому любая оценка модели пригодна только для определенного диапазона изменения дохода.

Рассмотрим пример более сложной модели, которая постулируется следующими положениями экономической теории:

- потребление есть возрастающая функция от имеющегося в наличии дохода, но возрастающая не быстрее, чем рост дохода;

- объем инвестиций есть возрастающая функция от национального дохода и убывающая функция государственного регулирования (нормы процента);

- национальный доход есть сумма потребительских, инвестиционных и государственных закупок товаров и услуг.

Предложим математическую формализацию этих положений, например, с помощью линейных соотношений:

$$\begin{aligned} y_{3,t} &= \beta_0^{(1)} + \beta_1^{(1)}\left(y_{2,t} - x_{2,t}\right) + \delta_{1,t}, \\ y_{1,t} &= \beta_1^{(2)} y_{2,t-1} + \beta_2^{(2)} x_{1,t} + \delta_{2,t}, \\ y_{2,t} &= y_{3,t} + y_{1,t} + x_{3,t}, \end{aligned} \tag{2}$$

где $y_{3,t}$ – потребление;

$x_{2,t}$ – подоходный налог;

$x_{3,t}$ – государственные закупки товаров и услуг;

δ_{1t}, δ_{2t} – случайные остатки, характеризующие влияние неучтенных факторов.

Априорные ограничения на параметры модели:

$$0 < \beta_1^{(1)} < 1,\ \beta_1^{(2)} > 0,\ \beta_2^{(2)} < 0.$$

В целом схема построения модели (2) та же, что и в случае (1), но процедура идентификации параметров неизмеримо сложнее и требует специально адаптированных методов математической статистики и линейной алгебры. Причем речь идет о простейшей модели, поскольку вряд ли линейные соотношения всегда справедливы.

Подход к идентификации параметров модели, как будет показано в данном курсе, зависит от вида модели – линейная, нелинейная, одно соотношение или система соотношений; от свойств случайных остатков δ_t, от типа (характера) переменных. Рассмотрим в связи с этим основные понятия дисциплины.

Все, участвующие в эконометрическом описании исследуемого процесса, переменные подразделяют на экзогенные, эндогенные и предопределенные [1].

Определение 1. Под экзогенными будем понимать переменные, задаваемые извне, автономно и можно сказать управляемые, планируемые. В рассмотренных выше примерах: $x_{1,t}$ – норма процента как инструмент государственного регулирования, $x_{2,t}$ – подоходный налог являются экзогенными переменными. В эконометрической модели экзогенные переменные выступают в роли объясняющих переменных.

Определение 2. Под эндогенными будем понимать такие переменные, значения которых формируются в процессе функционирования анализируемой системы под воздействием экзогенных переменных и во взаимодействии друг с другом. В приведенных примерах в качестве эндогенных переменных следует рассматривать $y_{1,t}$ – объем инвестиций, $y_{2,t}$ – национальный доход, $y_{3,t}$ – потребление.

В конечном итоге эндогенные переменные служат предметом объяснения, хотя формально могут присутствовать в форме «объясняющей» переменной (как $y_{2,t}$ в первом соотношении модели 2).

Определение 3. Предопределенными называют все переменные, роль которых в анализируемой системе (процессе) – объясняющие переменные.

В состав предопределенных переменных входят экзогенные и лаговые эндогенные переменные (эндогенные переменные измеренные в моменты времени предшествующие анализируемому). В наших примерах к предопределенным относятся $x_{1,t}$, $x_{2,t}$, $x_{3,t}$, $y_{2,t-1}$.

Эконометрические модели, предназначенные для объяснения поведения эндогенных переменных в зависимости от значений экзогенных и лаговых переменных, могут быть представлены как одним соотношением, так и системой, как в линейной, относительно оцениваемых параметров, форме, так и в нелинейной. Основное внимание, в предлагаемом начальном курсе эконометрики, будет уделено построению и анализу линейных моделей множественной регрессии, а в завершении подходам к исследованию систем линейных регрессионных уравнений. Среди нелинейных моделей будут рассмотрены только линеаризуемые.

Процесс эконометрического моделирования, включает, согласно [1], этапы:

1) постановочный – определение конечных целей моделирования, набора, участвующих в модели факторов и показателей;

2) априорный – предмодельный анализ экономической сущности изучаемого явления, формирование и формализация априорной информации, в частности, относящейся к природе исходных статистических данных;

3) параметризации – выбор общего вида модели, состав и формы входящих в нее связей;

4) информационный – сбор статистической информации об участвующих в модели факторах;

5) идентификации модели – статистическое оценивание неизвестных параметров модели, статистический анализ модели;

6) верификации модели – сопоставление реальных и модельных данных, оценка точности модельных данных.

В рамках данного курса основное внимание будет уделено этапам 3 – 6 эконометрического моделирования.

1 Классическая линейная модель множественной регрессии

1.1 Линейная модель множественной регрессии

Пусть социально-экономический процесс описывается показателями $y, x_1, x_2, ..., x_k$, при этом предполагается, что между ними существует объективная причинно-следственная связь, то есть на основе предварительного анализа установлено, что эндогенная переменная (результативный признак) y зависит от предопределенных (объясняющих переменных) $x_1, x_2, ..., x_k$. Взаимосвязи между экономическими переменными нецелесообразно искать в классе функциональных зависимостей, поскольку невозможно учесть влияние всех факторов на результативный признак.

Например, изучается объем выпускаемой продукции предприятий машиностроительной отрасли в зависимости от изменения ресурса производства. Как правило, объем выпускаемой продукции при одном и том же количестве затрачиваемого ресурса не будет одинаковым у различных предприятий, так как на него оказывает влияние совокупность факторов, которые не возможно измерить или предсказать, носящих случайный характер. Это означает, что одному значению объясняющей переменной (объему ресурса) соответствует несколько значений результативного признака (объема выпускаемой продукции) (рисунок 1.1).

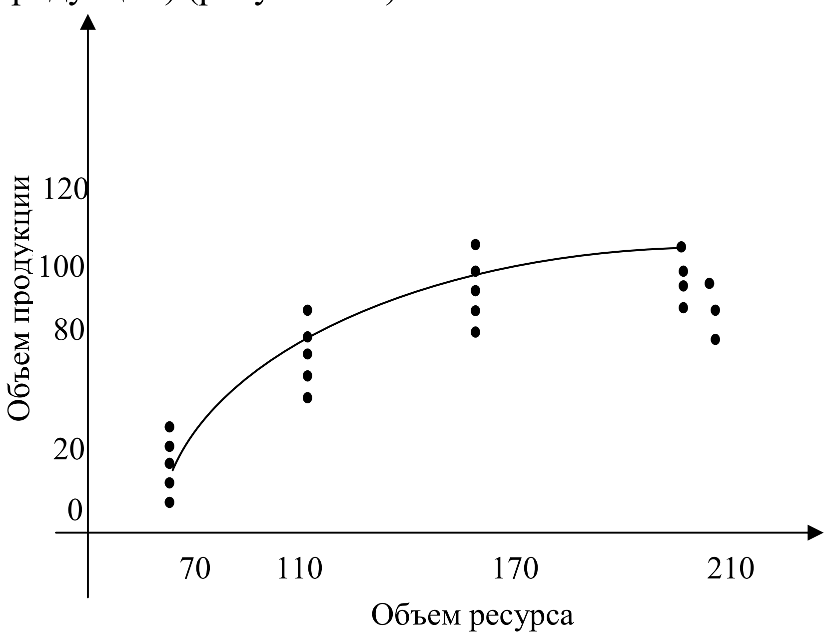


Рисунок 1.1 – Зависимость между объемом выпускаемой продукции (y) и объемом затрачиваемого ресурса

В связи с вышесказанным ставится задача построения зависимости между условными средними значениями результативного признака и текущими значениями объясняющих переменных $x_1, x_2, ..., x_k$, то есть функции регрессии [2, 3, 4, 5].

В общем случае регрессионная зависимость нелинейна, но, из соображений простоты, начнем с изучения линейной функции множественной регрессии (1.1, 1.1а):

$$\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k . \quad (1.1)$$

Если $x = (1, x_1, ..., x_k)^T$ - вектор объясняющих переменных, $\beta = (\beta_0, \beta_1, .., \beta_K)^T$ - вектор искомых коэффициентов, $\tilde{y}$ - условное среднее значение результативного признака, то функция множественной регрессии (1.1) примет вид:

$$\tilde{y} = x^T \beta , \quad (1.1а)$$

Для изучения взаимосвязи между исследуемыми показателями формируется выборка объемом "n". Результаты наблюдений над результативным признаком представлены вектором $Y = (y_1, ..., y_n)^T$ и матрицей X типа «объект-свойство» наблюдаемых значений признаков $x_1, ..., x_k$:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ . & . & . & ... & . \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix}$$

здесь x_{ij} – значение j-го признака на i-м объекте наблюдения; столбец из "1" можно считать столбцом "наблюденных" значений для признака $x_0^0 = 1$

Для оценки коэффициентов перейдем от (1.1) к так называемой линейной модели множественной регрессии, заменив в (1.1) (для каждого объекта) ненаблюдаемое $\tilde{y}_i$ на наблюдаемое значение y_i и вводя параметр ε_i, характеризующий величину расхождения $(y_i - \tilde{y}_i)$:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i , \ i = \overline{1, n} \quad (1.2)$$

где ε_i - регрессионные остатки, учитывающие влияние всех прочих факторов, не включенных в регрессионную модель. Обозначим $\bar{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^T$, тогда модель (1.2) можно представить в виде

$$Y = X\beta + \varepsilon . \quad (1.2а)$$

Система линейных уравнений (1.2)~(1.2а) называется линейной моделью множественной регрессии (ЛММР) (при $к=1$ говорят о парной (двумерной) модели регрессии).

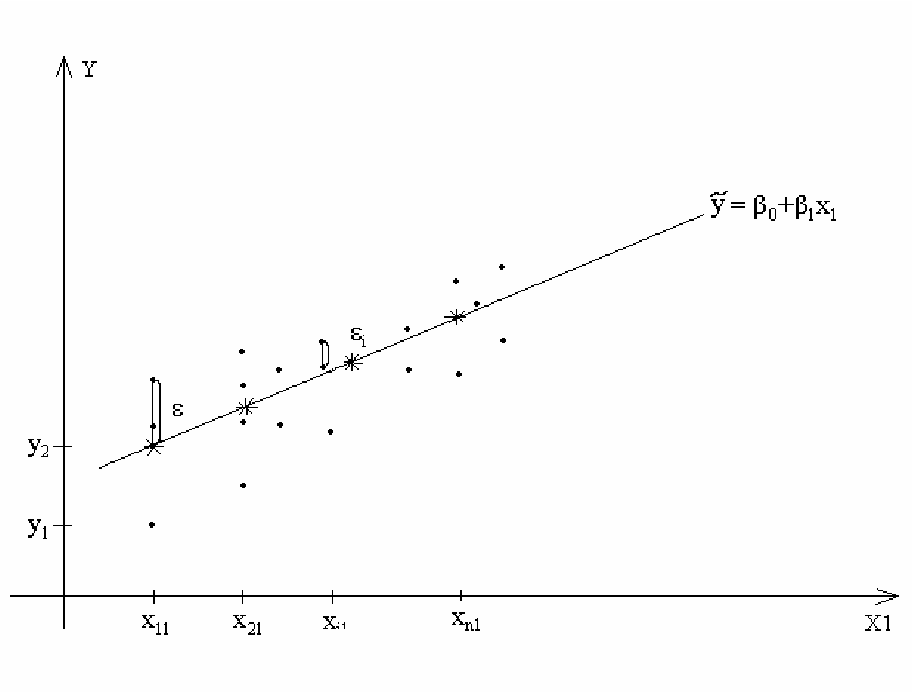


Рисунок 1.2 - Зависимость между результативным признаком y и объясняющей переменной x_1

Относительно регрессионных остатков и объясняющих переменных в дальнейшем будем предполагать, что выполняются 5 условий:

1) $x_1, \ldots, x_к$ – детерминированные переменные;

2) ранг матрицы X равен "$к+1$" – среди признаков нет линейно зависимых;

3) $M\varepsilon_i = 0$, $i = \overline{1,n}$ - нет систематических ошибок в измерении y;

4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2$, $i = \overline{1,n}$ - гомоскедастичность регрессионных остатков (равноточные измерения);

5) $\text{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0$, $i \neq j$, $i = \overline{1,n}$ $j = \overline{1,n}$ - условие некоррелированных регрессионных остатков.

Условия 4 – 5 можно заменить одним условием в векторной форме

4') $M\bar{\varepsilon}\bar{\varepsilon}^T = \sigma^2 E_n$, где E_n *– единичная матрица* .

Условия (1 – 5) известны как условия Гаусса – Маркова. Их суть и назначение станут ясными в процессе исследования. Линейная модель множественной регрессии (1.1)~(1.1а), удовлетворяющая требованиям (1–5), называется классической линейной моделью множественной регрессии (КЛММР).

1.2 Оценка неизвестных коэффициентов классической линейной модели множественной регрессии: метод наименьших квадратов

Оценку коэффициентов β уравнения регрессии можно искать исходя из требований минимума модуля отклонения наблюдаемых значений y_i от "значений" функции регрессии ($\sum_{i=1}^{n} /\varepsilon_i/ \to \min$), либо (обычно) из критерия минимума суммы квадратов отклонений наблюдаемых значений y_i от "значений" функции регрессии (метод наименьших квадратов), более удобного с позиции технической реализации [1, 4, 5].

Перейдем к оценке коэффициентов методом наименьших квадратов (МНК). Выпишем квадратичный функционал, обозначив через $b = (b_0, b_1, ..., b_k)^T$ оценку вектора β:

$$F = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - ... - b_k x_{in})^2 = (Y - Xb)^T (Y - Xb) =$$
$$= Y^T Y - b^T X^T Y - Y^T Xb + b^T X^T Xb = Y^T Y - 2b^T X^T Y + b^T X^T Xb \to \min. \quad (1.3)$$

Из положительной определенности матрицы квадратичной функции (1.3), следует, что достаточно воспользоваться необходимым условием существования экстремума. В итоге получим систему линейных алгебраических уравнений относительно коэффициентов $b_0, ..., b_k$ (нормальную систему) [1]:

$$2X^T Xb - 2X^T Y = 0. \quad (1.4)$$

В силу второго условия Гаусса-Маркова (rang $X=k+1$), матрица $X^T X$ – невырождена и из (1.4) получим МНК - оценки для вектора β[1]:

$$b_{\text{МНК}} \equiv b = (X^T X)^{-1} X^T Y \quad (1.5)$$

и, следовательно, оценку уравнения регрессии

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_k x_k. \quad (1.6)$$

1.3 Анализ вариации результативного признака *y*. Выборочный коэффициент детерминации

Можно показать, что общая вариация (дисперсия) результативного признака складывается из вариации функции регрессии, обусловленной варьированием значений объясняющих переменных $x_1, .., x_k$, (факторной дис-

[1] Здесь и далее в зависимости от контекста результаты оценки будем рассматривать либо как случайную величину, либо как значение случайной величины.

персии) и из вариации результативной переменной относительно функции регрессии (остаточной дисперсии) [1, 4, 5]:

$$Вар(y) = Q_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (Y - \bar{Y})^T (Y - \bar{Y}) = (Y - \hat{Y})^T (Y - \hat{Y}) + \\ + (\hat{Y} - \bar{Y})^T (\hat{Y} - \bar{Y}) \equiv Q_{ост} + Q_{факт} \qquad (1.7)$$

где $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ - значение выборочной средней;

$Y = (y_1, ..., y_n)^T$ - вектор наблюдаенных значений результативного признака;

$\bar{Y}_{n*1} = (\bar{y}, \bar{y}, ..., \bar{y})^T$ - вектор средних значений результативного признака;

$\hat{Y} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_n)^T$ - вектор модельных значений результативного признака, $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + ... + b_k x_{ik}$, $i = 1,2,...,n$;

$$Q_{ост} = (Y - \hat{Y})^T (Y - \hat{Y}) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2, \; e_i = y_i - \hat{y}_i;$$

$$Q_{факт} = (\hat{Y} - \bar{Y})^T (\hat{Y} - \bar{Y}) = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 .$$

Из (1.7), разделив обе части на $Q_{общ}$, получим $1 = \frac{Q_{ост}}{Q_{общ}} + \frac{Q_{факт}}{Q_{общ}}$. В качестве выборочной оценки коэффициента детерминации возьмем:

$$\hat{R}^2_{y/x_1,...,x_k} \equiv \frac{Q_{факт}}{Q_{общ}} \equiv \frac{(\hat{Y} - \bar{Y})^T (\hat{Y} - \bar{Y})}{(Y - \bar{Y})^T (Y - \bar{Y})} \equiv 1 - \frac{Q_{ост}}{Q_{общ}} \equiv 1 - \frac{(Y - \hat{Y})^T (Y - \hat{Y})}{(Y - \vec{Y})^T (Y - \bar{Y})} \equiv \\ \equiv 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (1.8)$$

Очевидно, что выборочный коэффициент детерминации принимает значения в промежутке [0;1] и характеризует долю общей вариации результативного признака *y*, объясняемую вариацией выборочной функции регрессии $\hat{f}(\bar{X})$. Чем меньше разброс наблюдаемых значений результативной переменной относительно функции регрессии, тем меньше остаточная дисперсия, и тем ближе $\hat{R}^2_{y/x_1,...,x_k}$ к единице.

Подправленная на несмещенность оценка $\hat{R}^{*2}_{y/x_1,...,x_к}$ коэффициента детерминации $\hat{R}^2_{y/x_1,...,x_к}$ имеет вид [1, 2, 5]:

$$\hat{R}^{*2}_{y/x_1,...,x_к} \approx 1-(1-\hat{R}^2_{y/x_1,...,x_к})\frac{n-1}{n-k-1}. \qquad (1.9)$$

1.4 Статистические свойства МНК – оценок КЛММР

1) МНК – оценка b является несмещенной оценкой вектора β [1, 5].

В самом деле из (1.2а), первого и третьего условий Гаусса – Маркова следует:

$$b=(X^TX)^{-1}X^TY=(X^TX)^{-1}X^T(X\beta+\varepsilon)=\beta+(X^TX)^{-1}X^T\varepsilon \qquad (1.10)$$

$$M(b)=M(\beta+(X^TX)^{-1}X^T\varepsilon)=\beta+(X^TX)^{-1}X^TM\varepsilon=\beta, \qquad (1.10а)$$

что доказывает несмещенность $b_{МНК}$.

2) Найдем ковариационную матрицу случайного вектора b, воспользовавшись условиями Гаусса-Маркова

$$\begin{aligned}\Sigma_{\bar{b}} &= M[(b-Mb)(b-Mb)^T]=M[((X^TX)^{-1}X^T\varepsilon)((X^TX)^{-1}X^T\varepsilon)^T)]=\\ &=M[(X^TX)^{-1}X^T\varepsilon\varepsilon^T(X^TX)^{-1}]=(X^TX)^{-1}X^TM(\varepsilon\varepsilon^T)X(X^TX)^{-1}=\\ &=(X^TX)^{-1}X^T\sigma^2\varepsilon_nX(X^TX)^{-1}=\sigma^2(X^TX)^{-1}.\end{aligned} \qquad (1.11)$$

Откуда, в частности, $D_{bj}=\sigma^2[(X^TX)^{-1}]_{jj}$, $j=\overline{0,k}$.

Несмещенная оценка для σ^2 и $\hat{\Sigma}_{\bar{b}}$ определяется по формулам [1, 4, 5]:

$$\hat{S}^2=\frac{1}{n-k-1}(Y-Xb)^T(Y-Xb), \qquad (1.12)$$

$$\hat{\Sigma}_b=\hat{S}^2(X^TX)^{-1}. \qquad (1.13)$$

Несмещенная оценка факторной дисперсии:

$$\hat{S}^2_{факт}=\frac{Q_{факт}}{к}. \qquad (1.14)$$

Приведем без доказательства одно из условий состоятельности оценок $b\ и\ \hat{S}^2$ (см. Себер. Дж.): оценки $b\ и\ \hat{S}^2$ являются состоятельными тогда и только тогда, когда наименьшее собственное число $\lambda_{\min}$ матрицы X^TX

стремится к бесконечности при $n \to \infty$. Отметим, что МНК - оценки коэффициентов КЛММР являются также эффективными в классе оценок линейных, относительно, компонент Y [1, 2, 3]. Дальнейшее изучение свойств оценок классической линейной модели множественной регрессии (КЛММР) проводится при дополнительном предположении о нормальном характере распределения регрессионных остатков: $\varepsilon_i \in N(0,\sigma^2)$, $i = \overline{1,n}$ или $\varepsilon \in N(0,\sigma^2 E_n)$.

1.4.1 Проверка гипотезы о незначимости линейной модели регрессии

Для проверки значимости построенного уравнения регрессии выдвигается гипотеза: H_0 $\beta_1=\beta_2=\ldots=\beta_к=0$, то есть ни одна из объясняющих переменных не оказывает существенного влияния на результативный признак (линейная модель множественной регрессии незначима). Альтернативная гипотеза H_1: $\exists j \in [\overline{1,\kappa}]: \beta_j \neq 0$, хотя бы одна из объясняющих переменных оказывает существенное влияние на результативный признак (ЛММР значима).

Для проверки гипотезы H_0 используется статистика [1, 2, 3, 4, 5]:

$$F = \frac{Q_{факт}/k}{Q_{ост}/(n-k-1)} = \frac{\hat{R}^2_{y/x_1,\ldots,x_n}/k}{(1-\hat{R}^2_{y/x_1,\ldots,x_n})/(n-k-1)}, \qquad (1.15)$$

которая в случае справедливости H_0 имеет распределение Фишера – Снедекора с числом степеней свободы $\nu_1 = k$ *и* $\nu_2 = n-k-1$.

По таблице распределения Фишера определяется критическое значение статистического критерия $(F_{КР}(\alpha;\nu_1;\nu_2))$ для заданного уровня значимости α, числа степеней свободы $\nu_1 = k$ *и* $\nu_2 = n-k-1$, и сравнивается с полученным по выборочным данным значением $(F_{наб})$. Если $F_{наб} > F_{КР}(\alpha;\nu_1;\nu_2)$, то нулевая гипотеза отвергается, в противном случае принимается[2].

1.4.2 Проверка гипотез о незначимости коэффициентов КЛММР

В случае, если нулевая гипотеза о незначимости уравнения регрессии отвергнута, проверяем гипотезы о значимости коэффициентов уравнения регрессии. Выдвигаются гипотезы вида:

---

[2] В современных пакетах прикладных программ при проверке любой гипотезы выдается информация о значимости нулевой гипотезы (α), которая сопоставляется с заданным уровнем значимости $(\alpha_{зад})$, в большинстве пакетах он по умолчанию принимается равным 0,05. Если $\alpha < \alpha_{зад}$, то нулевая гипотеза отвергается, в противном случае принимается.

H_0: $\beta_j=0$ – коэффициент β_j незначимо отличен от нуля, то есть объясняющая переменная x_j не оказывает существенного влияния на результативный признак.

H_1: $\beta_j \neq 0$ – коэффициент β_j значимо отличен от нуля, то есть объясняющая переменная x_j оказывает существенное влияние на результативный признак.

Для проверки гипотез H_0 строятся статистики [1, 2, 3, 4, 5]:

$$t_j = \frac{b_j}{S_{b_j}}, \quad j = 1,2,...,k, \quad S_{b_j} = \hat{S}\sqrt{[(X^T X)^{-1}]_{jj}}, \qquad (1.16)$$

которые в случае справедливости H_0, имеют распределение Стьюдента с $\nu = n - k - 1$ степенями свободы.

По таблице распределения Стьюдента определяется критическое значение статистического критерия $(t_{КР}(\alpha;\nu))$ для заданного уровня значимости α, числа степеней свободы $\nu = n - k - 1$, и сравнивается с полученными значениями по выборочным данным $(t_{j\,наб})$. Если $/t_{j\,наб}/ > t_{КР}(\alpha;\nu)$, то нулевая гипотеза отвергается, то есть объясняющая переменная x_j оказывает существенное влияние на результативный признак; в противном случае нулевая гипотеза принимается.

1.4.3 Построение доверительных интервалов для значимых коэффициентов КЛММР

Для коэффициентов уравнения регрессии, значимо отличных от нуля, строятся доверительные интервалы, используя статистику [1, 2, 3, 4, 5]

$$t = \frac{b_j - \beta_j}{S_{b_j}}, \qquad (1.17)$$

имеющую распределение Стьюдента с $\nu = n - k - 1$ степенями свободы. По определению доверительного интервала $P(|t| < \delta) = 1 - P(|t| \geq \gamma) = \gamma$, $P(|t| \geq \delta) = 1 - \gamma \equiv \alpha$, где γ - доверительная вероятность. В нашем случае из уравнения $\alpha = P(|t| \geq \delta) = St(\delta)$ определяется δ_α, в итоге доверительный интервал имеет вид:

$$b_j - \delta_\alpha S_{b_j} \leq \beta_j \leq b_j + \delta_\alpha S_{b_j} \qquad (1.18)$$

1.4.4 Построение доверительного интервала для $\tilde{y}(X_{n+1})$ и $y(X_{n+1})$

Так как при любом векторе $X_{n+1}=(1,x_{n+1,1},...,x_{n+1,k})^T$, $\hat{y}_{n+1}=X_{n+1}^T b$

$M\hat{y}_{n+1}=MX_{n+1}^T b=X_{n+1}^T\beta=\tilde{y}_{n+1}$

$D\hat{y}_{n+1}=D(X_{n+1}^T\bar{b})=X_{n+1}^T\sum_b X_{n+1}=\sigma^2 X_{n+1}^T(X^TX)^{-1}X_{n+1}$,

$\hat{S}_{\hat{y}_{n+1}}^2=\hat{S}^2X_{n+1}^T(X^TX)^{-1}X_{n+1}$,

то доверительный интервал для $\tilde{y}(X_{n+1})$ строится так же как и в предыдущем параграфе, используя статистику

$$t=\frac{\hat{y}_{n+1}-\tilde{y}_{n+1}}{\hat{S}_{\hat{y}_{n+1}}}, \tag{1.19}$$

имеющую распределение Стьюдента с $\nu=n-k-1$ степенями свободы.

В итоге доверительный интервал имеет вид:

$$\hat{y}_{n+1}-\delta_\alpha\hat{S}_{\hat{y}_{n+1}}\le\tilde{y}_{n+1}\le\hat{y}_{n+1}+\delta_\alpha\hat{S}_{\hat{y}_{n+1}} \tag{1.20}$$

При построении доверительного интервала для $y(X_{n+1})=y_{n+1}$ воспользуемся статистикой [1, 2, 3, 4, 5]

$$t=\frac{\hat{y}_{n+1}-\tilde{y}_{n+1}}{\hat{S}\sqrt{X_{n+1}^T(X^TX)^{-1}X_{n+1}+1}}, \tag{1.21}$$

имеющую такое же распределение, что и (1.19).

Пример 1[3]. Исследуется зависимость урожайности зерновых культур (y, ц/га) от ряда показателей, характеризующих различные факторы сельскохозяйственного производства :

X_1 - число тракторов на 100 га;

X_2 - число зерноуборочных комбайнов на 100 га;

X_3 - число орудий поверхностной обработки почвы на 100 га;

X_4 - количество удобрений, расходуемых на гектар (т/га);

X_5 - количество химических средств защиты растений, расходуемых на гектар (ц/га).

Исходные данные для 20 сельскохозяйственных районов области приведены в таблице А1 (Приложение А). В данном примере располагаем пространственной выборкой объема $n=20$; число объясняющих переменных $\kappa=5$.

[3] Пример взят из книги Айвазян С.А. Прикладная статистика. Основы эконометрики. – М.: ЮНИТИ - ДАНА, 2001. – С. 52.

Для изучения статистической взаимосвязи между исходными переменные рассмотрим линейную функцию регрессии:

$\tilde{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$.

Применение формулы (1.5) позволяет получить МНК-оценки для параметров β_j, $j = 1,2,3,4,5$: $b = (3{,}515;\ -0{,}006;\ 15{,}542;\ 0{,}110;\ 4{,}475;\ -2{,}932)^T$. Таким образом, оценка линейной функции регрессии имеет вид:

$$\hat{y} = \underset{(5,41)}{3{,}515} - \underset{(0,60)}{0{,}006} X_1 + \underset{(21,59)}{15{,}542} X_2 + \underset{(0,85)}{0{,}110} X_3 + \underset{(1,54)}{4{,}475} X_4 - \underset{(3,09)}{2{,}932} X_5, \hat{R}^2 = 0{,}52,\ F_{набл} = 3{,}01.$$

В скобках под значениями оцененных коэффициентов указаны их стандартные ошибки S_{b_j}, полученные по формуле (1.13).

Проверка гипотезы о незначимости построенного уравнения регрессии H_0: $\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=0$ (ни одна из объясняющих переменных не оказывает существенного влияния на результативный признак), H_1: $\exists j \in [\overline{1,5}]:\ \beta_j \neq 0$, осуществляется с помощью F-критерия Фишера. При уровне значимости $\alpha = 0{,}05$ и $\nu_1 = 5,\ \nu_2 = 20 - 5 - 1 = 14$ по таблице распределения Фишера определяется критическое значение $F_{крит} = 2{,}982$; $F_{набл} = 3{,}01$. Так как $F_{набл} > F_{крит}$, то нулевая гипотеза отвергается, то есть существует хотя бы одна объясняющая переменная, оказывающая существенное влияние на урожайность зерновых культур (результативный признак).

Проверка гипотез о незначимости отдельных коэффициентов (H_0: $\beta_j=0$, $\beta_j \neq 0$) проводилась с помощью t-критерия Стьюдента, определяемого формулой (1.16): $t_{0набл} = 0{,}653$, $t_{1набл} = -0{,}012$, $t_{2набл} = 0{,}722$, $t_{3набл} = 0{,}138$, $t_{4набл} = 2{,}899$, $t_{5набл} = -0{,}949$. По таблице распределения Стьюдента определяется критическое значение при уровне значимости $\alpha = 0{,}05$, $\nu = 20 - 5 - 1 = 14$: $t_{крит}(0{,}05;14) = 2{,}14$. Так как $/ t_{jнабл} / < t_{крит}, j = 1,2,3,5$, то нулевая гипотеза оказывается принятой для всех объясняющих переменных, кроме X_4. Таким образом, пока можно сделать вывод, что существенное влияние на урожайность зерновых культур оказывает только X_4 – количество удобрений, расходуемых на гектар земли, при изменении которой на единицу урожайность в среднем увеличится на 4,475 ц/га. Однако приведенные соображения позволяют сделать вывод о необходимости проведения дальнейшего исследование.

1.5 Мультиколлинеарность: понятие, признаки и методы устранения

Если существует функциональная линейная зависимость между объясняющими переменными (столбцы матрицы X линейно зависимы), то говорят, что существует **полная мультиколлинеарность**[1]. Наличие функциональной зависимости означает, что матрица объясняющих переменных имеет ранг меньше, чем *к+1(rang X<K+1)*, а это в свою очередь, приводит к вырожденности матрицы $X^T X$ и невозможности вычисления обратной матрицы

$(X^T X)^{-1}$ и, следовательно, оценок коэффициентов методом наименьших квадратов.

Полную мультиколлинеарность нетрудно избежать на предварительной стадии анализа и отбора множества объясняющих переменных путем исключения дублирующих признаков. Формально для выявления полной мультиколлинеарности определяется ранг X (например, методом элементарных преобразований) и попутно выявляются, какие столбцы линейно зависят от других. Выявив эти столбцы, из модели линейной регрессии исключаются соответствующие этим столбцам признаки, и строится регрессионную модель меньшей размерности по линейно независимым признакам, при этом матрица $X^T X$ будет невырожденная, и, следовательно, существует возможность построения регрессионной модели.

Наиболее чаще в эконометрической практике встречается **реальная (или частичная) мультиколлинеарность,** возникающая в случаях существования достаточно тесных корреляционных связей между объясняющими переменными [1]. В этом случая, матрица $X^T X$ становится плохо обусловленной – ее определитель близок к нулю. Тогда элементы матрицы $(X^T X)^{-1}$ вычисляются с большой погрешностью, следовательно, снижается точность МНК-оценок, то есть увеличивается дисперсия оценок коэффициентов в регрессионной модели, и, как следствие, менее надежными становятся, а то и вовсе искажаются результаты проверки статистических гипотез.

1.5.1 Признаки мультиколлинерности

Внешние признаки мультиколлинеарности:

1) неоправданно большие с экономической точки зрения коэффициенты уравнения регрессии;

2) небольшие изменения исходных статистических данных приводят к существенному изменению оценок коэффициентов модели, вплоть до изменения их знаков;

3) неправильные с экономической точки зрения знаки отдельных коэффициентов регрессии;

4) среди коэффициентов уравнения регрессии много (может быть все) незначимы, а модель значима;

5) стандартные отклонения велики настолько, что сравнимы или даже превосходят сами коэффициенты;

6) доверительные интервалы для коэффициентов уравнения регрессии содержат внутри себя точку нуль.

Следует обратить внимание на то, что это только необходимые, но не достаточные признаки мультиколлинеарности [1].

Формальные признаки мультиколлинеарности:

1) среди оценок коэффициентов парной или частной корреляции объясняющих переменных есть такие, которые по абсолютной величине превышают 0,6;

2) достаточно высокие значения множественных коэффициентов корреляции (детерминации) одной из объясняющей переменной (X_j) на другие $\hat{R}^2_{x_j / x_1,...,x_{j-1},x_{j+1},...,x_k} > 0.6$;

3) существование тесных линейных статистических связей между объясняющими переменными приводит к так называемой плохой (слабой) обусловленности системы;

Необходимым условием плохой обусловленности является малость определителя матрицы $X^T X$.

Достаточным условием плохой обусловленности (мультиколлинеарности) является большое значение числа обусловленности [1]

$$M = \frac{\max\limits_{i=1,n} |\lambda_i|}{\min\limits_{i=1,n} |\lambda_i|}, \tag{1.20}$$

где λ_i - собственное число матрицы $X^T X$.

1.5.2 Методы устранения мультиколлинеарности

Поскольку при мультиколлинеарности объясняющих переменных в модели регрессии сталкиваются с ситуацией дублирования информации, доставляемой сильно взаимозависимыми объясняющими переменными, то естественно осуществить что переход от исходного числа k анализируемых переменных к меньшему числу l наиболее информативных некоррелированных переменных. Поэтому в основе методов устранения мультиколлинеарности лежит идея уменьшения общего числа объясняющих переменных за счет отбора наиболее существенных с точки зрения их влияния на результативный признак.

Существует несколько подходов к решению задачи отбора наиболее существенных объясняющих переменных в модели регрессии: метод «ридж-регрессии», метод главных компонент, наиболее распространен– метод пошаговой регрессии [1, 4].

1.5.2.1 Метод пошаговой регрессии

Выделяют метод пошаговой регрессии с включением и исключением переменных. Рассмотрим метод пошаговой регрессии с включением переменных [1].

На первом шаге $(l = 1)$ определяется первая объясняющая переменная $x^{(i_1(1))}$, которую можно назвать наиболее информативной, при условии, что в регрессионную модель Y по X мы можем включить только одну из набора объясняющих переменных.

На втором шаге $(l = 2)$ реализация критерия максимальности коэффициента детерминации определит уже наиболее информативную пару объясняющих переменных $x^{(i_1(1))}, x^{(i_2(2))}$, при чем одна их них та, которую отобрали на предыдущем шаге. Эта пара объясняющих переменных должна будет иметь наиболее тесную статистическую связь с результативным признаком, то есть

$$\hat{R}^2(y/x^{(i_1(1))}, x^{(i_2(2))}) = \max_{1 \le i_1,..,i_2,...i_l \le k} \hat{R}^2(\text{y/x}^{(\text{i}_1)},..., x^{(i_l)}) \qquad (1.21)$$

На третьем шаге $(l = 2)$ будет отобрана наиболее информативная тройка объясняющих переменных и т.д.

На каждом шаге рассчитываются несмещенная оценка коэффициента детерминации

$$\hat{R}^{*2}(l) \cong 1 - \left(1 - \hat{R}^2(l)\right)\frac{n-1}{n-l-1} \qquad (1.22)$$

и величина нижней границы доверительного интервала $\hat{R}^2_{\min}(l)$

$$\hat{R}^2_{\min}(l) = \hat{R}^{*2}(l) - 2\sqrt{\frac{2l(n-l-1)}{(n-1)(n^2-1)}}\left(1 - \hat{R}^2(l)\right). \qquad (1.23)$$

Предполагается выбирать в качестве оптимального числа l_0 объясняющих переменных регрессионной модели значение l, при котором величина $R^2_{\min}(l)$ достигает своего максимума (рисунок 1.3).

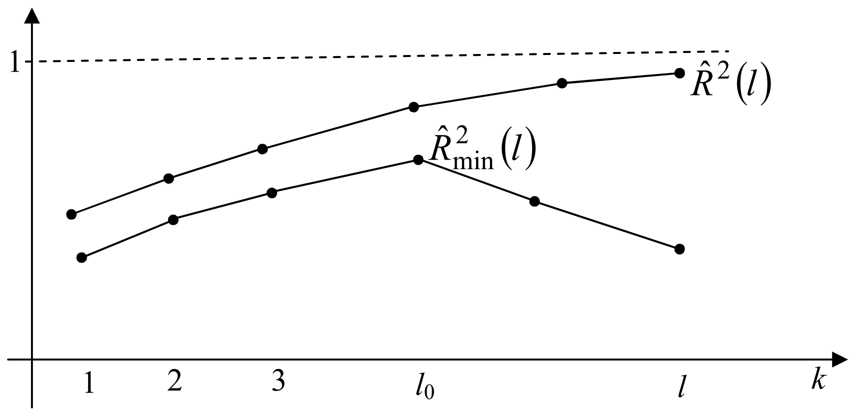


Рисунок 1.3 – Зависимость нижней границы доверительного интервала и числа объясняющих переменных

В случае метода пошаговой регрессии с исключением переменных, имеет место обратная процедура, то есть на первом шаге строится уравнение регрессии на все k факторных признаков и, если среди его коэффициентов есть незначимые, то на втором шаге строятся уравнения регрессии на $k-1$ признаков, среди которых выбирается то, которому соответствует наибольший выборочный коэффициент детерминации. Если и в этой модели есть незначимые коэффициенты, то процедура повторяется для $k-2$ переменных и т.д. [1].

При реализации версии «всех возможных регрессий» решается следующая задача: для заданного значения l ($l=\overline{1,k-1}$) путем полного перебора всех возможных комбинаций из l объясняющих переменных, отобранных из исходного набора $x_1, x_2, \ldots x_к$, определить такие переменные $x^{\left(i_1^0(l)\right)}, x^{\left(i_2^0(l)\right)}, \ldots, x^{\left(i_l^0(l)\right)}$, для которых коэффициент детерминации с результативным признаком был бы максимальным [1]. Использование описанного метода требует больших объемов вычислений, поэтому целесообразно на практике применять метод пошаговой регрессии.

Пример 2. На основе данных примера 1, проведем исследование модели регрессии на мультиколлинеарность. В результате реализации МНК была получена оценка уравнения регрессии:

$$\hat{y} = \underset{(5,41)}{3{,}515} - \underset{(0,60)}{0{,}006}\,X_1 + \underset{(21,59)}{15{,}542}\,X_2 + \underset{(0,85)}{0{,}110}\,X_3 + \underset{(1,54)}{4{,}475}\,X_4 - \underset{(3,09)}{2{,}932}\,X_5,\ \hat{R}^2 = 0{,}52,\ F_{набл} = 3{,}01.$$

Проанализируем внешние и формальные признаки мультиколлинеарности. Во-первых, вызывают сомнения знаки коэффициентов при показателях X_1, X_5, согласно которым урожайность в среднем уменьшится на 0,006 ц/га при увеличении числа тракторов на единицу (X_1), при увеличении количества химических средств защиты растений на единицу урожайность в среднем уменьшится на 2,932 ц/га. Во-вторых, несмотря на то, что регрессионная модель в целом значима, большинство коэффициентов незначимы стандартные ошибки оценок коэффициентов при переменных X_1, X_2, X_3, X_5 превосходят значения самих коэффициентов.

В матрице парных коэффициентов корреляции (таблица А2) имеются значимые коэффициенты корреляции по абсолютной величине превосходящие 0,75 – 0,80: $r(x_1,x_2)=0{,}85$; $r(x_1,x_3)=0{,}98$; $r(x_2,x_3)=0{,}88$. Более внимательное изучение этого вопроса достигается с помощью расчета значений коэффициентов детерминации каждой из объясняющих переменных по всем остальным переменным: $\hat{R}^2_{x1/x2,x3,x4,x5}=0{,}97$; $\hat{R}^2_{x2/x1,x3,x4,x5}=0{,}86$; $\hat{R}^2_{x3/x2,x1,x4,x5}=0{,}98$; $\hat{R}^2_{x4/x2,x3,x1,x5}=0{,}45$; $\hat{R}^2_{x5/x2,x3,x4,x1}=0{,}62$. Все это дает основание заподозрить наличие мультиколлинеарности, что подтверждается значением числа обусловленности: $M=\dfrac{402{,}318}{5{,}19\times10^{-3}}=7{,}752\times10^{4}$.

Для устранения мультиколлинеарности воспользуемся методом пошаговой регрессии с включением переменных.

На первом шаге оценивается линейная функция регрессии y по одной-объясняющей переменной, выбирается наиболее информативная объясняющая переменная в смысле критерия (1.21):

$$\hat{R}^2_{Y/X_1} = 0{,}185;\ \hat{R}^2_{Y/X_2} = 0{,}139;\ \hat{R}^2_{Y/X_3} = 0{,}163;\ \hat{R}^2_{Y/X_4} = 0{,}333;\ \hat{R}^2_{Y/X_5} = 0{,}110$$

$$\max_{j=1,5} \hat{R}^2_{Y/X_j} = \hat{R}^2_{Y/X_4}\, 0{,}333$$

Таким образом, наиболее информативной объясняющей переменной является X_4 - количество удобрений, расходуемых на гектар. Рассчитаны несмещенная оценка коэффициента детерминации $\hat{R}^{*2}(1) = 0{,}296$ и нижняя граница доверительного интервала $\hat{R}^2_{\min}(1) = 0{,}204$.

На втором шаге оценивается линейная функция регрессии y по двум объясняющим переменным, одна из которых отобрана на предыдущем шаге, то есть X_4:

$$\hat{R}^2_{Y/X_4,X_1} = 0{,}469;\ \hat{R}^2_{Y/X_4,X_2} = 0{,}462;\ \hat{R}^2_{Y/X_4,X_3} = 0{,}482;\ \hat{R}^2_{Y/X_4,X_5} = 0{,}33$$

$$\max_{j=1,2,3,5} \hat{R}^2_{Y/X_4X_j} = \hat{R}^2_{Y/X_4X_3} = 0{,}482.$$

Вторая информативная объясняющая переменная - X_3 (число орудий поверхностной обработки почвы на 100 га). Рассчитаны несмещенная оценка коэффициента детерминации $\hat{R}^{*2}(2) = 0{,}421$ и нижняя граница доверительного интервала $\hat{R}^2_{\min}(2) = 0{,}323$.

На третьем шаге оценивается линейная функция регрессии y по трем объясняющим переменным, две из которых отобраны на предыдущих шагах, то есть X_4, X_3:

$$\hat{R}^2_{Y/X_4,X_3,X_1} = 0{,}485;\ \hat{R}^2_{Y/X_4,X_3,X_2} = 0{,}484;\ \hat{R}^2_{Y/X_4,X_3,X_5} = 0{,}498;$$

$$\max_{j=1,2,5} \hat{R}^2_{Y/X_4,X_3,X_j} = \hat{R}^2_{Y/X_4,X_3,X_5} = 0{,}482.$$

Третья информативная объясняющая переменная - X_5 (количество химических средств защиты растений, расходуемых на 100 га). Рассчитаны несмещенная оценка коэффициента детерминации $\hat{R}^{*2}(3) = 0{,}404$ и нижняя граница доверительного интервала $\hat{R}^2_{\min}(3) = 0{,}291$.

Сравнение нижней границы доверительного интервала $\hat{R}^2_{\min}(3)$ и $\hat{R}^2_{\min}(2)$ свидетельствует о том, что третью объясняющую переменную X_5 в модель включать нецелесообразно, т.к. $\hat{R}^2_{\min}(3) < \hat{R}^2_{\min}(2)$ (рисунок 1.4).

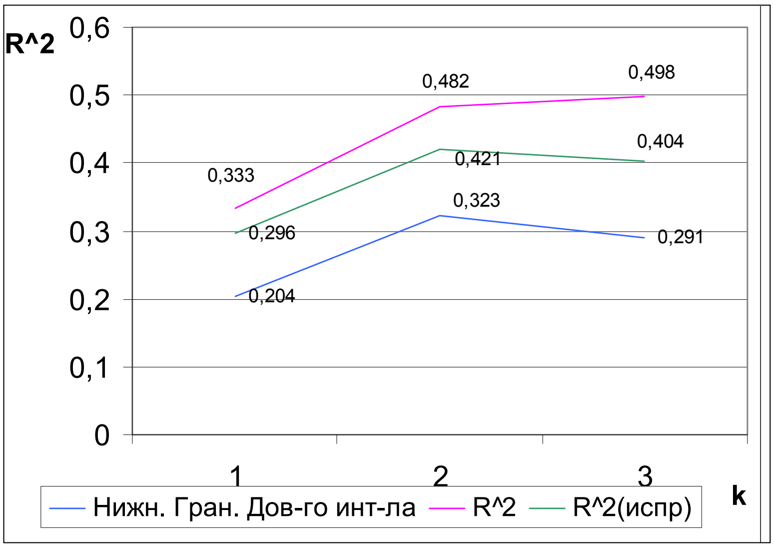


Рисунок 1.4 – Зависимость нижней границы доверительного интервала и числа объясняющих переменных

В результате реализации метода пошаговой регрессии, получили оценку уравнения регрессии:

$$\hat{y} = \underset{(0,66)}{7{,}29} + \underset{(0,18)}{0{,}28} X_3 + \underset{(1,07)}{3{,}47} X_4, \ \hat{R}^2 = 0{,}48; \quad F = 7{,}9.$$

Таким образом, полученная регрессионная модель значима, коэффициенты являются значимыми и интерпретируются следующим образом: увеличении числа орудий поверхностной обработки почвы на 1 урожайность зерновых культур возрастет в среднем на 0,28 ц/га и на 3,47 ц/га при увеличении количества химических средств защиты растений на единицу. Доля вариации результативного признака на 48,0 % объясняется вариацией факторных признаков, включенных в регрессионную модель.

1.5.2.2 Метод «ридж-регрессии»

Устранение мультиколлинеарности может осуществляться путем построения смещенных оценок **(**«ридж-регрессия» или «гребневая регрессия»**)**, за счет регуляризации матрицы нормальной системы [1]**:**

$$b_\tau = (X^T X + \tau \cdot E_{\kappa+1})^{-1} X^T Y. \tag{1.24}$$

В этом случае добавление к диагональным элементам матрицы ($X^T X$) «гребня» τ (τ - некоторое положительное число, $0{,}1 \le \tau \le 0{,}4$, $E_{\kappa+1}$ - единичная матрица $(\kappa+1)$ порядка) с одной стороны, делает получаемые при этом оценки смещенными, а с другой,- превращает матрицу $X^T X$ из «плохо обусловленной» в «хорошо обусловленную» [1] .

1.6 Тестовые задания для самоконтроля

1) Модель $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ называют классической линейной моделью множественной регрессии, если выполняются следующие условия:

а) $$\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \text{неслучайные переменные}; \\ rang\, X = k + 1 << n; \\ M\varepsilon_i \neq 0,\ i = 1,2,...,n; \\ M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 \text{ при } i = j \\ 0 \text{ при } i \neq j \end{cases} \quad i = \overline{1,n};\quad j = \overline{1,n} \end{cases}$$

б) $$\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \text{неслучайные переменные}; \\ rang\, X = k + 1 << n \\ M\varepsilon_i = 0,\ i = 1,2,...,n; \\ M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 \text{ при } i = j \\ 0 \text{ при } i \neq j \end{cases} \quad i = \overline{1,n};\quad j = \overline{1,n} \end{cases}$$

в) $$\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \text{неслучайные переменные}; \\ rang\, X = k + 1 << n \\ M\varepsilon_i = 0,\ i = 1,2,...,n; \\ M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma_i^2 \text{ при } i = j \\ 0 \text{ при } i \neq j \end{cases} \quad i = \overline{1,n};\quad j = \overline{1,n} \end{cases}$$

г) $$\begin{cases} (x^{(1)}, x^{(2)}, ..., x^{(k)}) - \text{неслучайные переменные}; \\ rang\, X = k + 1 << n \\ M\varepsilon_i = 0,\ i = 1,2,...,n; \\ \sum_\varepsilon = \sigma^2 \sum_0,\ \text{где} \sum_0 \neq E_n \end{cases}$$

2) Для оценки значимости отдельных коэффициентов КЛММР $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ рассчитывают:

а) оценки коэффициентов регрессии и стандартные ошибки;

б) стандартные коэффициенты регрессии и случайные отклонения уравнения регрессии;

в) оценки коэффициентов регрессии и случайные отклонения уравнения регрессии;

г) парные коэффициенты корреляции между объясняющими переменными.

3) Внешний признак мультиколлинерности указан в варианте:

а) парные коэффициенты корреляции между некоторыми из объясняющих переменных $|r_{ij}| \leq 0{,}7$;

б) достаточно малое значение множественного коэффициента детерминации одного из факторных признаков на другие $\hat{R}^2_{x_i / x_1, ..., x_{i-1}, x_{i+1}, ..., x_k}$;

в) некоторые коэффициенты уравнения регрессии имеют неправильные с точки зрения экономической теории знаки;

г) дисперсии регрессионных остатков постоянны;

4) Получена оценка уравнения регрессии $\hat{y} = b_0 + b_1 x$. Коэффициент регрессии b_1 показывает:

а) тесноту связи;

б) долю дисперсии *"у"*, зависимую от *"х"*;

в) на сколько единиц в среднем изменяется *"у"* при изменении *"х"* на одну единицу;

г) ошибку коэффициента корреляции.

5) В КЛММР $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1, n}$ долю вариации зависимой переменной, обусловленной вариацией объясняющих переменных характеризует:

а) коэффициент вариации;

б) коэффициент корреляции;

в) коэффициент детерминации;

г) коэффициент эластичности.

6) Формальный признак мультиколлинерности представлен в следующем варианте:

а) между объясняющими переменными и результативной переменной существует сильная корреляционная зависимость.

б) наличие только отрицательных знаков у коэффициентов уравнения регрессии;

в) парные коэффициенты корреляции между некоторыми из объясняющих переменных $|r_{ij}| \geq 0{,}7$;

г) парные коэффициенты корреляции между некоторыми из объясняющих переменных $|r_{ij}| \leq 0{,}7$;

7) По выборке объемом *n* единиц получена оценка уравнения регрессии: $\hat{y} = \underset{(0,2)}{5{,}2} + \underset{(0,01)}{0{,}2}\, x_1 - \underset{(0,05)}{0{,}4}\, x_2$ в которой $\hat{R}^2 = 0{,}95$. Следовательно, вариация результативного признака объясняется вариацией всех объясняющих переменных, вошедших в модель на…

а) 90 %; б) 81 %; в) 95 % г) 45 %

8) Полученную оценку уравнения регрессии $\hat{y} = \underset{(0,66)}{7{,}29} + \underset{(0,957)}{3{,}48}\, x$ можно интерпретировать следующим образом:

а) с увеличением объясняющей переменной на 3,48 единиц результативная переменная увеличится на 1 единицу;

б) с увеличением объясняющей переменной на 3,48 единиц результативная переменная увеличится на 7,29 единиц;

в) с увеличением объясняющей переменной на 1 единицу результативная переменная уменьшится на 3,48 единиц;

г) с увеличением объясняющей переменной на 1 единицу результативная переменная увеличится на 3,48 единиц;

9) Чем меньше величина остаточной дисперсии, тем:

а) меньше влияние прочих, не учитываемых в уравнении регрессии факторов;

б) больше влияние прочих, не учитываемых в уравнении регрессии факторов;

в) больше используется факторов в уравнении регрессии;

г) меньше оценка коэффициента детерминации;

10) Суть метода наименьших квадратов заключается в том, что оценка коэффициентов регрессии определяется из условия:

а) максимизации суммы квадратов отклонений наблюденных значений y_i от модельных значений функции регрессии $\tilde{y}_i$;

б) минимизации суммы квадратов отклонений выборочной средней от выборочной дисперсии;

в) минимизации суммы отклонений наблюденных значений y_i от модельных значений функции регрессии $\tilde{y}_i$;

г) минимизации суммы квадратов отклонений наблюденных значений y_i от модельных значений функции регрессии $\tilde{y}_i$;

11) Если по t-критерию большинство коэффициентов регрессии статистически незначимы, а модель в целом по F- критерию значима, то можно сделать предположение

а) о наличии мультиколлинеарности

б) об отсутствии мультиколлинерности

в) о наличии автокорреляции

г) о наличии гетероскедастичности

12) В КЛММР $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ ковариационная матрица вектора ошибок имеет вид:

а) $\Sigma_\varepsilon = \sigma^2 E_n$ в) $\Sigma_\varepsilon = \sigma^2 \Sigma_0$, где $\Sigma_0 \neq E_n$

б) $\Sigma_\varepsilon \neq \sigma^2 E_n$ г) $\Sigma_\varepsilon \neq \sigma_0{}^2 \Sigma_0$

13) В КЛММР $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ ковариационная матрица вектора оценок параметров имеет вид:

а) $\Sigma_b = \sigma_0^2 (X^T \Sigma_0^{-1} X)^{-1}$ в) $\Sigma_b = n(X^T X)^{-1}$

б) $\Sigma_b = \sigma^2 (X^T X)^{-1}$ г) $\Sigma_b = (X^T X)^{-1}$

14) Для изучения рынка жилья в городе по данным о 43 коттеджах, получена следующая оценка уравнения регрессии: $\hat{y} = \underset{(8,2)}{21.1} - \underset{(3,8)}{6.2}x_1 + \underset{(0,83)}{3.57}x_2 + \underset{(0,59)}{1.17}x_3$, $\hat{R}^2 = 0.7$

где y –цена объекта, тыс.долл.;

x_1 - расстояние до центра города, км;

x_2 - полезная площадь объекта, кв.м;

x_3 -–число комнат в квартире;

Модель оказалась значимой. Укажите факторы, оказывающие существенное влияние на цену объекта на заданном уровне значимости α=0,05.

а) расстояние до центра города,

б) полезная площадь объекта;

в) число комнат в квартире;

г) расстояние до центра города, и полезная площадь объекта;

P.S. $t_{кр}(0{,}05{,}39) = 2{,}022$; $t_{кр}(0{,}05{,}40) = 2{,}021$;

$t_{кр}(0{,}05{,}41) = 2{,}020$; $t_{кр}(0{,}05{,}47) = 2{,}012$.

15) По 16 наблюдениям получена оценка уравнения регрессии $\hat{y} = b_0 + b_1x_1$. При проверке значимости коэффициента регрессии b_1 вычислено $t_{набл.}$ = 2,01. Следовательно, коэффициент регрессии является значимым при уровне значимости:

а) α = 0,05; б) α = 0,01; в) α = 0,1; г) α = 0,02;

P.S. $t_{кр}(0{,}01;14) = 2{,}977$; $t_{кр}(0{,}1;14) = 1{,}761$;

$t_{кр}(0{,}05;14) = 2{,}145$; $t_{кр}(0{,}02;14) = 2{,}625$.

1.7 Индивидуальное задание №1

По данным, приведенным в Приложении А:

1) построить МНК-оценки коэффициентов линейной модели множественной регрессии и провести ее анализ;

2) провести анализ построенной модели на мультиколлинеарность;

3) устранить мультиколлинеарность методом пошаговой регрессии

1.8 Пример выполнения индивидуального задания №1[4]

Рассмотрим пример построения линейной регрессионной модели на основе информации об ожидаемой продолжительности жизни мужчин, число лет (y), рождаемости населения на 1000 человек (x_1), смертности населения на 1000 человек (x_2), числе браков на 1000 человек (x_3), числе разводов на 1000 человек (x_4), коэффициенте младенческой смертности (x_5), соотношении

[4] Индивидуальные задания и пример их выполнения разработаны авторами совместно с преподавателем кафедры математических методов и моделей в экономике Жемчужниковой Ю.А.

денежного дохода и прожиточного минимума, % (x_6), соотношении средней оплаты труда и прожиточного минимума трудоспособного населения, % (x_7), численности населения с денежными доходами ниже прожиточного минимума в % от численности населения (x_8), числа зарегистрированных преступлений на 100000 населения (x_9)

Запуск и подготовка данных. Запустить ППП Statistica. После запуска на экране откроется основное окно системы Statistica, представленное на рисунке 1.5.

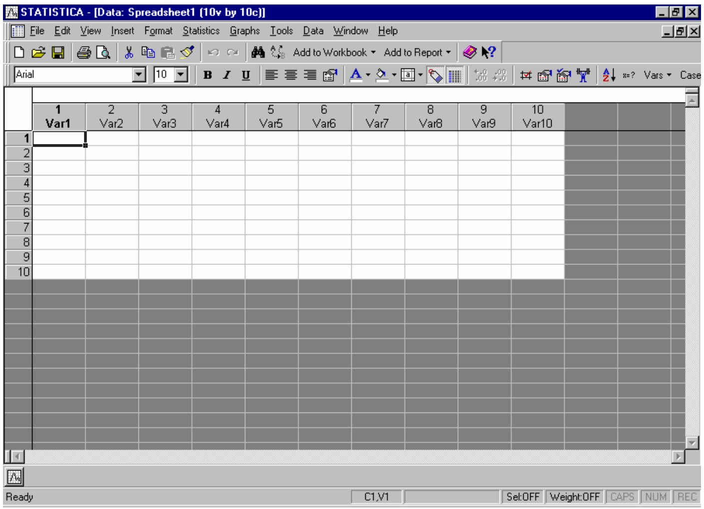


Рисунок 1.5 – Стартовое окно пакета Statistica

Стандартный вид исходной таблицы содержит 10 строк (10 cases) и 10 столбцов (10 variables). Так как исходная информация может быть представлена произвольного размера, то возникает необходимость в корректировке размерности таблицы. Если необходимо увеличить число столбцов, то в меню Insert, выбираем Add Variables, если необходимо изменить число строк, то – Add Cases. При этом откроется меню возможных операций со столбцами (строками).

Далее необходимо ввести данные для проведения регрессионного анализа. Если исходная информация уже имеется, то следует открыть нужный файл – для этого используется кнопка **Open Data – Открыть данные.** Окно с частью данных для анализа представлено на рисунке 1.6.

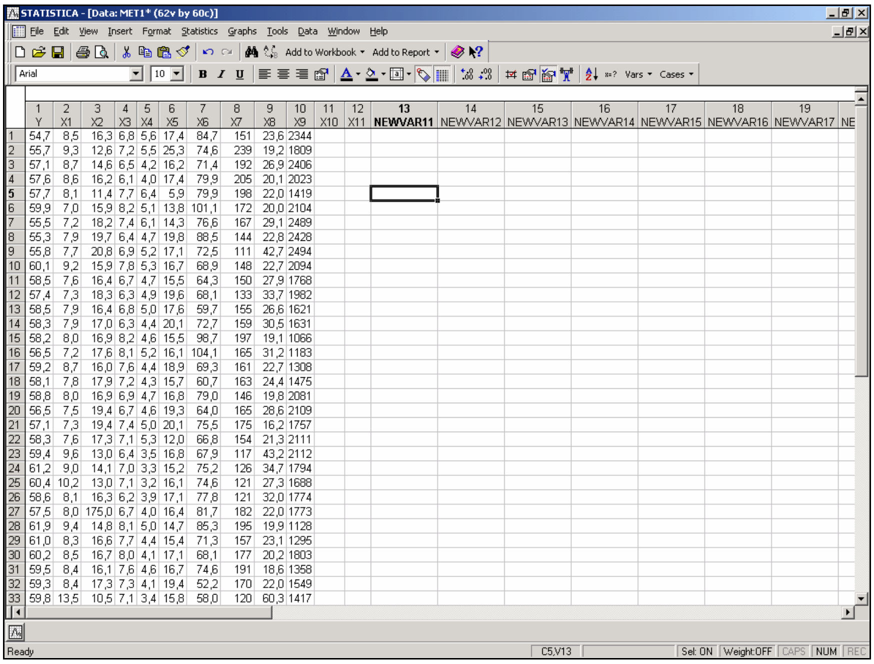


Рисунок 1.6 – Исходные данные

Для построения уравнения множественной регрессии в меню системы открыть **Statistics - Критерии** и выбрать в появившемся меню строку **Multiple Regression – Множественная регрессия**:

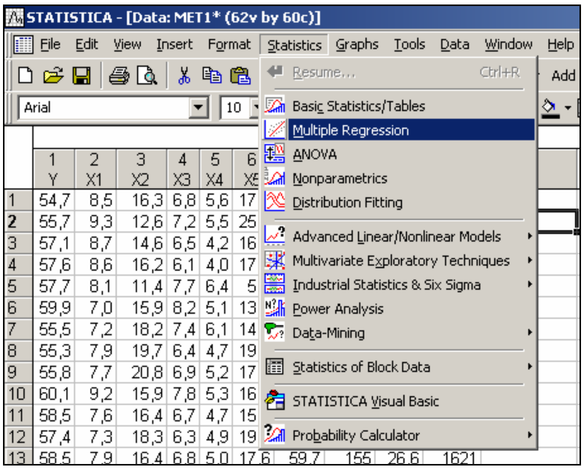


Рисунок 1.7 – Выбор пункта меню для проведения регрессионного анализа

На экране появится окно:

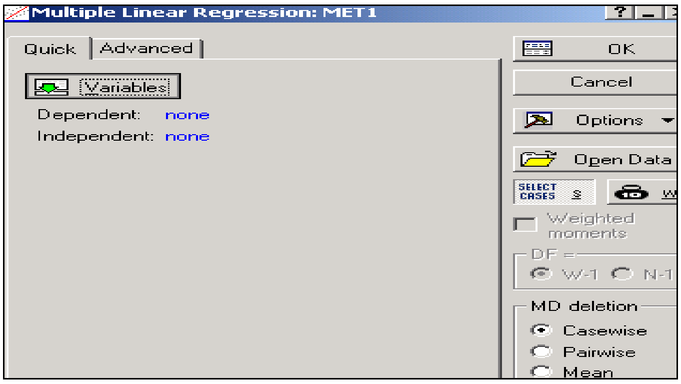


Рисунок 1.8 – Окно выбора переменных.

Далее необходимо выбрать зависимую (результирующую, объясняемую) и независимые (объясняющие) переменные для анализа.

Для задания переменных воспользуемся кнопкой **Variables** – **Переменные** из панели **Multiple Regression – Множественная регрессия** (рисунок 1.8).

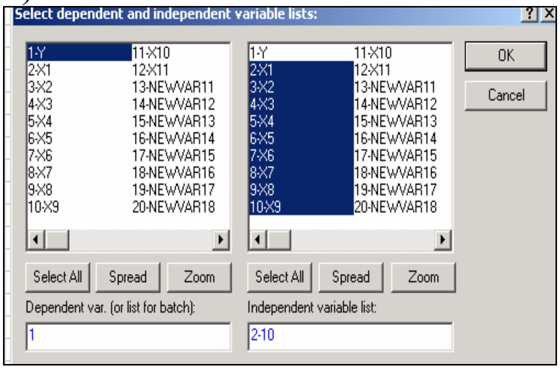


Рисунок 1.9 – Выбор зависимой и независимых переменных для проведения регрессионного анализа

В окне **Select dependent and independent variable list – Выбор зависимой переменной и списка независимых переменных**, выделяя имя переменной в левой части окна, производится выбор зависимой переменной **Dependent**. В правой части окна выбираем независимую переменную (**Independent)**. Выбор нескольких несмежных переменных производят при нажатой клавише **CTRL**. После выбора переменных необходимо щелкнуть на кнопке **OK**, вновь окажемся в панели модуля Множественная регрессия. Нажатие на кнопку **Advanced** позволяет перейти к окну функциональных возможностей модуля Множественная регрессия.

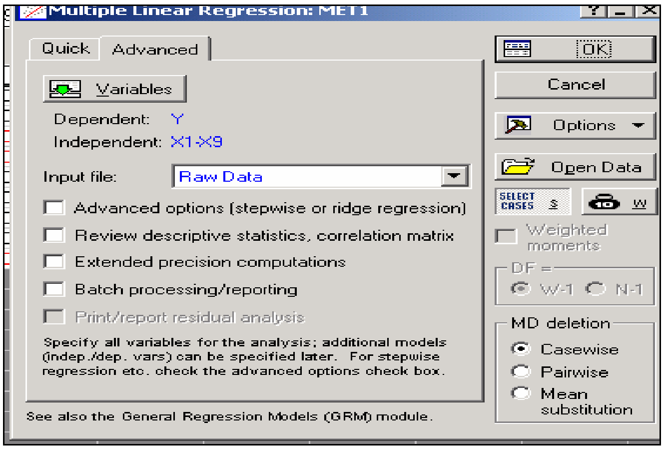


Рисунок 1.10 – Модуль множественная регрессия

Строка **Input file** определяет тип входной информации. Если входная информация представляет собой массив исходных данных, следует оставить **Raw Data** (необработанные данные). В поле окна **MD deletion** можно задать правило обработки пропущенных данных. Установка флажка в поле **Advanced options** позволит перейти к диалоговому окну **Model Defenition,** открывающему возможность выбора метода анализа, среди которых методы пошаговой регрессии и гребневой. Установка флажка в поле **Review descriptive statistics, correlations matrix** позволит провести предварительный анализ исходных переменных и построить корреляционную матрицу, анализ которой дает возможность сделать важные выводы о структуре связей между выбранными переменными. Установка флажка в поле **Extended precision computations** позволит выбрать метод расчета с расширенной точностью. После определения всей необходимой информации для построения модели, щелкните по кнопке **OK** в правом углу окна. Результаты расчетов приведены в виде отчета на рисунке 1.11.

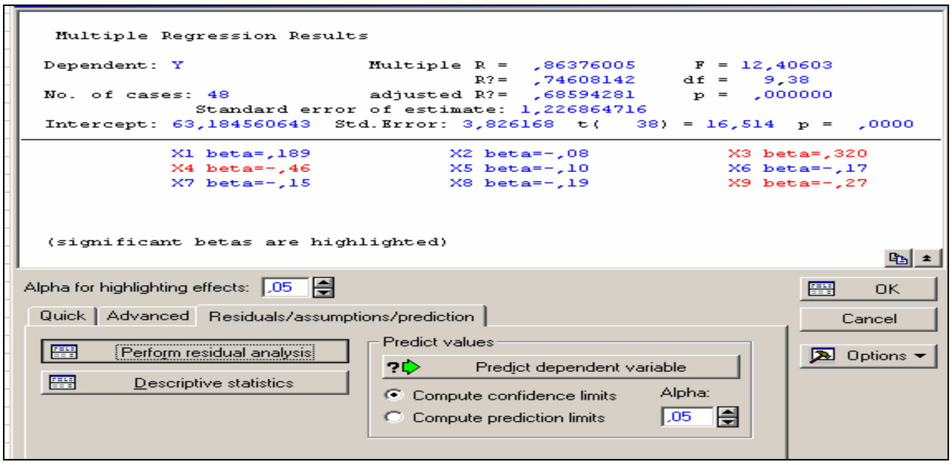


Рисунок 1.11 – Окно с результатами вычислений

В верхней информационной части окна результатов представлены основные характеристики построенной модели, а нижняя – содержит кнопки доступа к дополнительной информации, позволяющей провести исчерпывающий анализ модели, дать интерпретацию вычисленным параметрам и оценить адекватность модели исходным данным.

Рассмотрим содержание информационной части окна.

В левой части окна приводится имя зависимой переменной (**Dependent**) и число наблюдений, по которым построено уравнение регрессии (**No. Of Cases**).

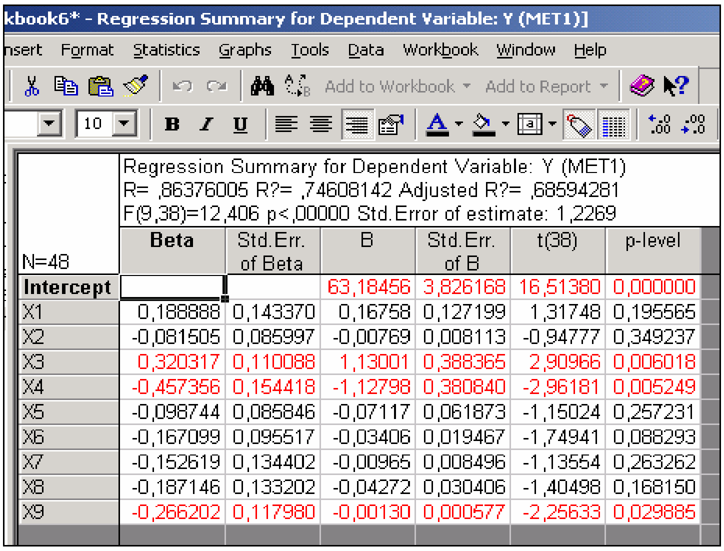
В правой части окна приводится оценка коэффициента множественной корреляции (**Multiple R**) и значение квадрата этого коэффициента (**R^2**) – коэффициента детерминации, несмещенная оценка R^2 (**Adjusted R^2**), значение **F–критерий**.

Также в верхней части окна результатов анализа приводится оценка свободного члена уравнения регрессии **(Intercept)**, стандартная ошибка (среднеквадратическое отклонение) этой оценки **(Std. Error)**, значение t-критерия и уровень значимости.

Standard Error of estimate является оценкой $\sqrt{S^2_{ост}}$, где $S^2_{ост}$- несмещенная оценка остаточной дисперсии.

Во второй части информационного окна подсвечены оценки значимых регрессионных коэффициентов (речь в данном случае идет о нормированных оценках: Beta**-** коэффициентах).

Более подробную информацию получим после нажатия на кнопку **Regression summary** (рисунок 1.12).



N=48	Beta	Std.Err. of Beta	B	Std.Err. of B	t(38)	p-level
Intercept			63,18456	3,826168	16,51380	0,000000
X1	0,188888	0,143370	0,16758	0,127199	1,31748	0,195565
X2	-0,081505	0,085997	-0,00769	0,008113	-0,94777	0,349237
X3	0,320317	0,110088	1,13001	0,388365	2,90966	0,006018
X4	-0,457356	0,154418	-1,12798	0,380840	-2,96181	0,005249
X5	-0,098744	0,085846	-0,07117	0,061873	-1,15024	0,257231
X6	-0,167099	0,095517	-0,03406	0,019467	-1,74941	0,088293
X7	-0,152619	0,134402	-0,00965	0,008496	-1,13554	0,263262
X8	-0,187146	0,133202	-0,04272	0,030406	-1,40498	0,168150
X9	-0,266202	0,117980	-0,00130	0,000577	-2,25633	0,029885

Рисунок 1.12 – Результаты оценивания параметров линейной модели множественной регрессии

В данном окне модуля представлены оценки параметров модели (**B**- обычные оценки и **Beta**- нормированные оценки), оценки их стандартных ошибок (**St. Error**) и уровни значимости **(p-level)** t-критерий Стьюдента.

Далее можно приступить к исследованию остатков регрессионной модели. Остатки исследуются в специальном окне **Residuals analysis – Анализ**

остатков. В нем приведен широкий набор статистических и визуальных методов исследования остатков модели. Для этого необходимо щелкнуть мышкой по кнопке **Residuals/assumptions/prediction – Остатки/распределение/предсказанные** в окне рисунка 1.11 (рисунок 1.13).

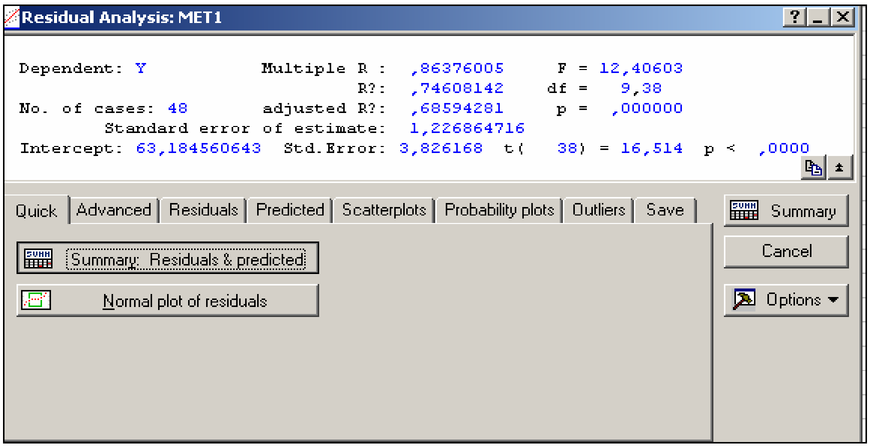
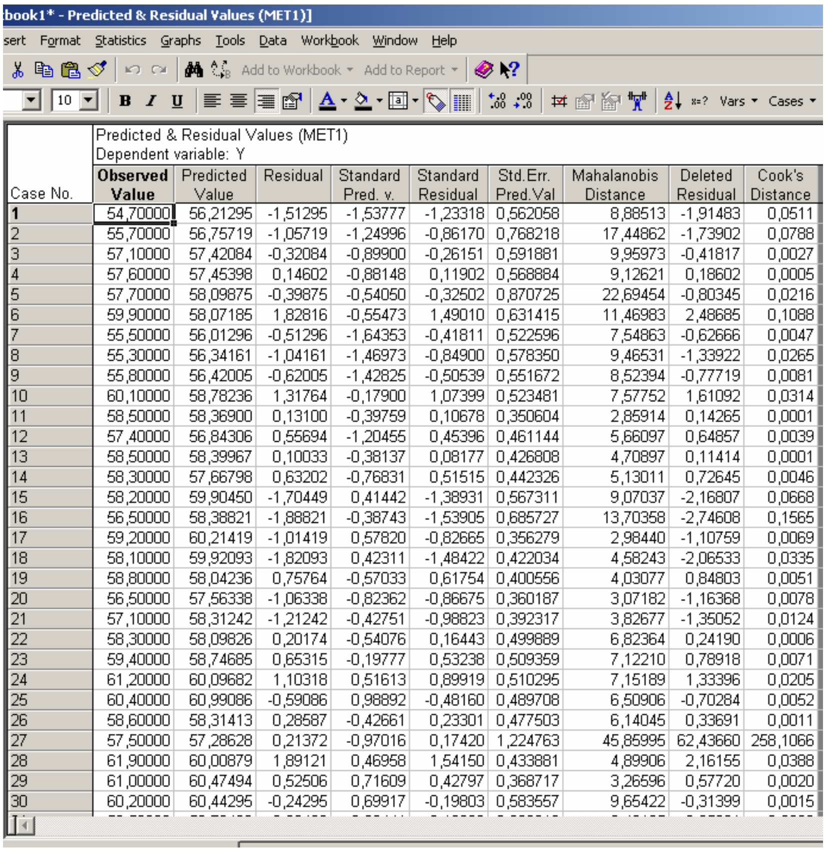


Рисунок 1.13 – Окно для анализа регрессионных остатков

Информация о значениях остатков может быть получена нажатием на кнопку **Summary: Residuals & predicted** (рисунок 1.14).



Predicted & Residual Values (MET1)
Dependent variable: Y

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
1	54,70000	56,21295	-1,51295	-1,53777	-1,23318	0,562058	8,88513	-1,91483	0,0511
2	55,70000	56,75719	-1,05719	-1,24996	-0,86170	0,768218	17,44862	-1,73902	0,0788
3	57,10000	57,42084	-0,32084	-0,89900	-0,26151	0,591881	9,95973	-0,41817	0,0027
4	57,60000	57,45398	0,14602	-0,88148	0,11902	0,568884	9,12621	0,18602	0,0005
5	57,70000	58,09875	-0,39875	-0,54050	-0,32502	0,870725	22,69454	-0,80345	0,0216
6	59,90000	58,07185	1,82816	-0,55473	1,49010	0,631415	11,46983	2,48685	0,1088
7	55,50000	56,01296	-0,51296	-1,64353	-0,41811	0,522596	7,54863	-0,62666	0,0047
8	55,30000	56,34161	-1,04161	-1,46973	-0,84900	0,578350	9,46531	-1,33922	0,0265
9	55,80000	56,42005	-0,62005	-1,42825	-0,50539	0,551672	8,52394	-0,77719	0,0081
10	60,10000	58,78236	1,31764	-0,17900	1,07399	0,523481	7,57752	1,61092	0,0314
11	58,50000	58,36900	0,13100	-0,39759	0,10678	0,350604	2,85914	0,14265	0,0001
12	57,40000	56,84306	0,55694	-1,20455	0,45396	0,461144	5,66097	0,64857	0,0039
13	58,50000	58,39967	0,10033	-0,38137	0,08177	0,426808	4,70897	0,11414	0,0001
14	58,30000	57,66798	0,63202	-0,76831	0,51515	0,442326	5,13011	0,72645	0,0046
15	58,20000	59,90450	-1,70449	0,41442	-1,38931	0,567311	9,07037	-2,16807	0,0668
16	56,50000	58,38821	-1,88821	-0,38743	-1,53905	0,685727	13,70358	-2,74608	0,1565
17	59,20000	60,21419	-1,01419	0,57820	-0,82665	0,356279	2,98440	-1,10759	0,0069
18	58,10000	59,92093	-1,82093	0,42311	-1,48422	0,422034	4,58243	-2,06533	0,0335
19	58,80000	58,04236	0,75764	-0,57033	0,61754	0,400556	4,03077	0,84803	0,0051
20	56,50000	57,56338	-1,06338	-0,82362	-0,86675	0,360187	3,07182	-1,16368	0,0078
21	57,10000	58,31242	-1,21242	-0,42751	-0,98823	0,392317	3,82677	-1,35052	0,0124
22	58,30000	58,09826	0,20174	-0,54076	0,16443	0,499889	6,82364	0,24190	0,0006
23	59,40000	58,74685	0,65315	-0,19777	0,53238	0,509359	7,12210	0,78918	0,0071
24	61,20000	60,09682	1,10318	0,51613	0,89919	0,510295	7,15189	1,33396	0,0205
25	60,40000	60,99086	-0,59086	0,98892	-0,48160	0,489708	6,50906	-0,70284	0,0052
26	58,60000	58,31413	0,28587	-0,42661	0,23301	0,477503	6,14045	0,33691	0,0011
27	57,50000	57,28628	0,21372	-0,97016	0,17420	1,224763	45,85995	62,43660	258,1066
28	61,90000	60,00879	1,89121	0,46958	1,54150	0,433881	4,89906	2,16155	0,0388
29	61,00000	60,47494	0,52506	0,71609	0,42797	0,368717	3,26596	0,57720	0,0020
30	60,20000	60,44295	-0,24295	0,69917	-0,19803	0,583557	9,65422	-0,31399	0,0015

Рисунок 1.14 – Наблюденные, модельные значения результативного признака, значение регрессионных остатков

Для проведения теста на нормальный характер распределения регрессионных остатков, скопируем столбец **Residual** в окно с исходными данными. Затем в меню системы Statistica выберем пункт **Distribution Fitting**. На экране появится окно:

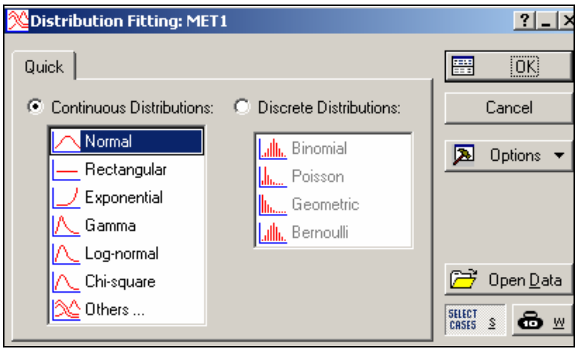


Рисунок 1.15 – Выбор вида распределения регрессионных остатков

В появившемся окне выберем распределение **Normal – Нормальное** и щелкнем по кнопке **OK.** После чего на экране появится окно:

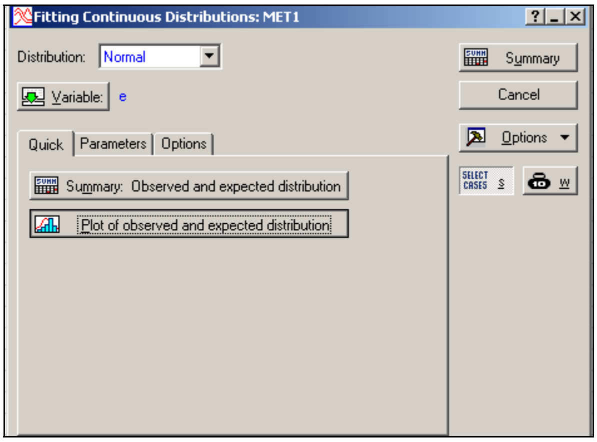


Рисунок 1.16 – Выбор пунктов для построения гистограммы регрессионных остатков

В данном окне сначала необходимо выбрать переменные, используя кнопку **Variable.** Кроме того, в данном модуле, используя кнопку **Parameters – Параметры,** можно изменить количество интервалов, верхнюю и нижнюю границы интервалов и т.д. Для получения графика нормального распределения, нажмем по кнопке **Plot of observed and expected distribution.**

На экране появится окно, содержащее гистограмму распределения, значение $\mathbf{X^2}$ – критерия, степени свободы, значимость нулевой гипотезы.

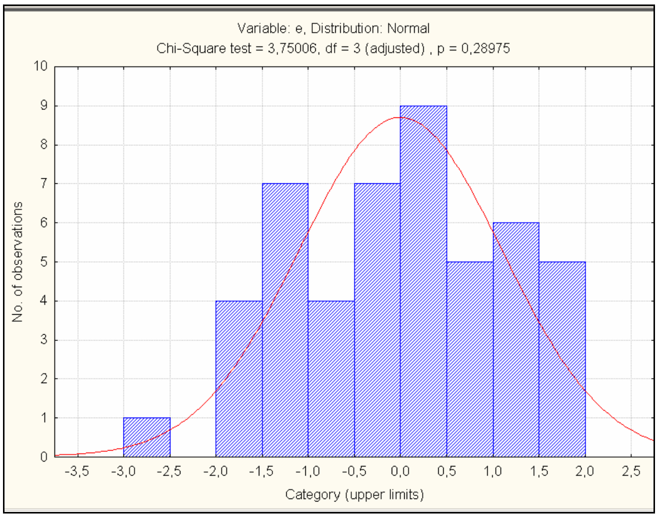
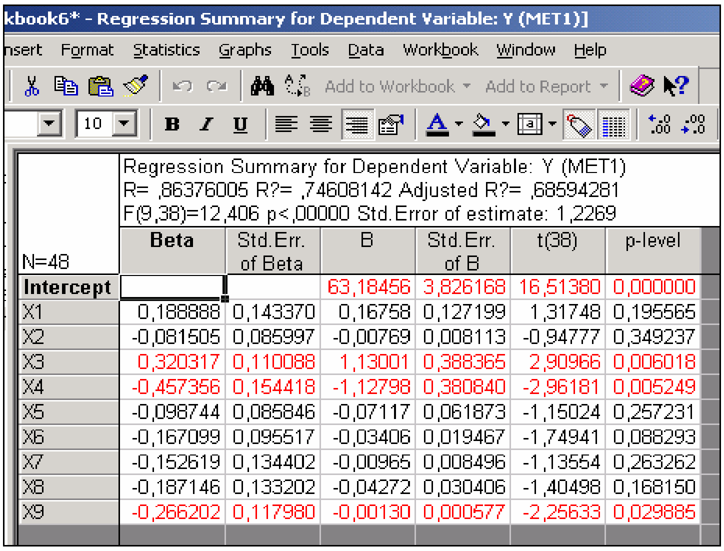


Рисунок 1.17 – График распределения регрессионных остатков

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что распределение регрессионных остатков не отличаются от нормального, так как значимость нулевой гипотезы (p=0,29).

Так как регрессионные остатки имеют нормальное распределение, то есть смысл проводить дальнейший анализ построенного уравнения множественной регрессии.

Итак, вернемся к окну **Multiple Regression Results** - Результаты множественной регрессии:



Regression Summary for Dependent Variable: Y (MET1)
R= ,86376005 R?= ,74608142 Adjusted R?= ,68594281
F(9,38)=12,406 p<,00000 Std.Error of estimate: 1,2269

N=48	Beta	Std.Err. of Beta	B	Std.Err. of B	t(38)	p-level
Intercept			63,18456	3,826168	16,51380	0,000000
X1	0,188888	0,143370	0,16758	0,127199	1,31748	0,195565
X2	-0,081505	0,085997	-0,00769	0,008113	-0,94777	0,349237
X3	0,320317	0,110088	1,13001	0,388365	2,90966	0,006018
X4	-0,457356	0,154418	-1,12798	0,380840	-2,96181	0,005249
X5	-0,098744	0,085846	-0,07117	0,061873	-1,15024	0,257231
X6	-0,167099	0,095517	-0,03406	0,019467	-1,74941	0,088293
X7	-0,152619	0,134402	-0,00965	0,008496	-1,13554	0,263262
X8	-0,187146	0,133202	-0,04272	0,030406	-1,40498	0,168150
X9	-0,266202	0,117980	-0,00130	0,000577	-2,25633	0,029885

Рисунок 1.18 - Результаты оценивания параметров линейной модели множественной регрессии

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = \underset{(3,83)}{63,18} + \underset{(0,13)}{0,17}\, X_1 - \underset{(0,008)}{0,008}\, X_2 + \underset{(0,039)}{1,13}\, X_3 - \underset{(0,38)}{1,128}\, X_4 - \underset{(0,06)}{0,07}\, X_5 - \underset{(0,02)}{0,003}\, X_6 -$$

$$- \underset{(0,008)}{0,0097}\, X_7 - \underset{(0,03)}{0,04}\, X_8 - \underset{(0,0005)}{0,001}\, X_9$$

Как видно из отчета, уравнение регрессии значимо, т.е. модель адекватна экспериментальным данным, значимыми оказались только коэффициенты при переменных X_3, X_4, X_9, среднеквадратические ошибки S_{b_j} оказались того же порядка, что и коэффициенты регрессии при переменных X_1, X_2, X_5, X_6, X_7, X_8. Согласно полученной модели при увеличении соотношения денежного дохода и прожиточного минимума на 1% ожидаемая продолжительность жизни мужчин уменьшится в среднем на 0,034 (коэффициент при переменной X_6 имеет отрицательный знак), что противоречит экономическому смыслу. Все эти внешние признаки позволяют нам заподозрить наличие мультиколлинеарности между объясняющими переменными. Итак, перейдем к рассмотрению формальных признаков по выявлению мультиколлинеарности

1. Для вычисления оценки матрицы парных коэффициентов корреляции в окне множественная регрессия (рисунок 1.10) установим флажок в поле **Review descriptive statistics, correlations matrix.** После нажатия на кнопку **ОК** на экране откроется окно.

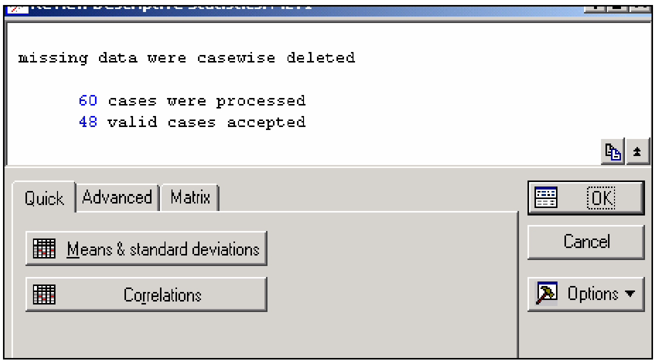
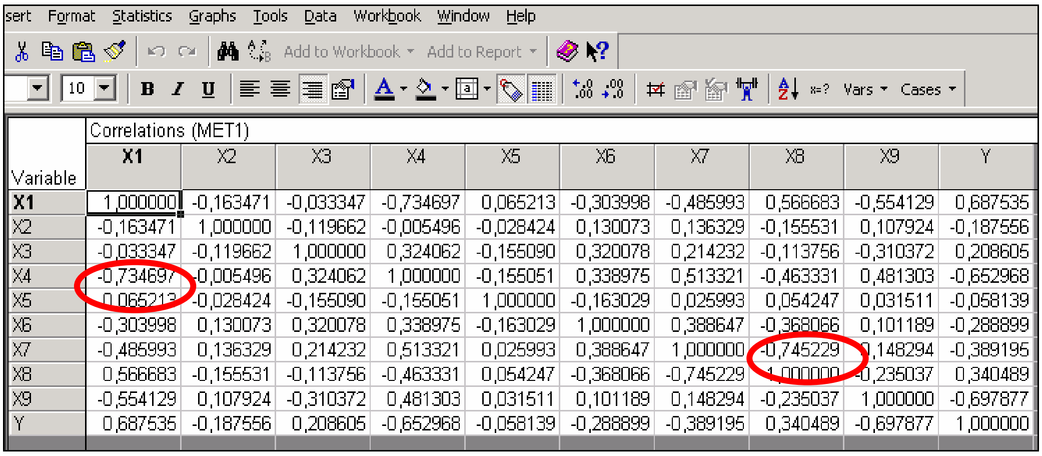


Рисунок 1.19 – Окно для вычисления оценки матрицы парных коэффициентов корреляции

В открывшемся окне нажимаем кнопку **Correlations** для вычисления оценки матрицы парных коэффициентов корреляции.



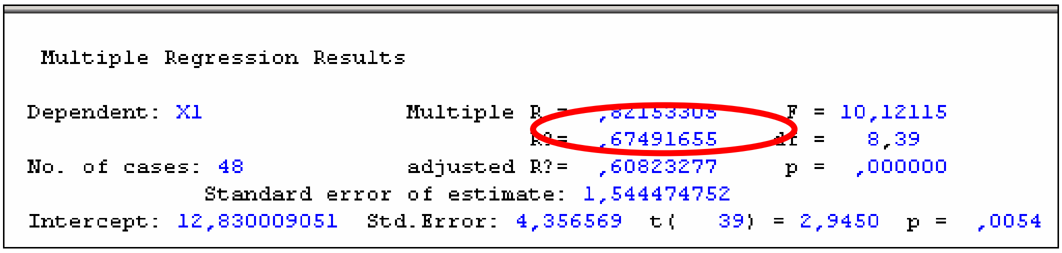
Correlations (MET1)

Variable	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
X1	1,000000	-0,163471	-0,033347	-0,734697	0,065213	-0,303998	-0,485993	0,566683	-0,554129	0,687535
X2	-0,163471	1,000000	-0,119662	-0,005496	-0,028424	0,130073	0,136329	-0,155531	0,107924	-0,187556
X3	-0,033347	-0,119662	1,000000	0,324062	-0,155090	0,320078	0,214232	-0,113756	-0,310372	0,208605
X4	-0,734697	-0,005496	0,324062	1,000000	-0,155051	0,338975	0,513321	-0,463331	0,481303	-0,652968
X5	0,065213	-0,028424	-0,155090	-0,155051	1,000000	-0,163029	0,025993	0,054247	0,031511	-0,058139
X6	-0,303998	0,130073	0,320078	0,338975	-0,163029	1,000000	0,388647	-0,368066	0,101189	-0,288899
X7	-0,485993	0,136329	0,214232	0,513321	0,025993	0,388647	1,000000	-0,745229	0,148294	-0,389195
X8	0,566683	-0,155531	-0,113756	-0,463331	0,054247	-0,368066	-0,745229	1,000000	-0,235037	0,340489
X9	-0,554129	0,107924	-0,310372	0,481303	0,031511	0,101189	0,148294	-0,235037	1,000000	-0,697877
Y	0,687535	-0,187556	0,208605	-0,652968	-0,058139	-0,288899	-0,389195	0,340489	-0,697877	1,000000

Рисунок 1.20 – Оценка матрицы парных коэффициентов корреляции

На основе вычисленной матрицы есть основания подозревать тесную связь между X_1 и X_4 ($r(x^{(1)}, x^{(4)}) = 0{,}73$) и X_7 и X_8 ($r(x^{(7)}, x^{(8)}) = -0{,}75$).

2. Для определения коэффициентов детерминации следует воспользоваться модулем множественная регрессия, где в качестве зависимой переменной выбрать $x^{(j)}$, все остальные объясняющие переменные в качестве независимых (рисунок 1.21).



```
Multiple Regression Results

Dependent: X1              Multiple R = ,82153305        F = 10,12115
                                  R?= ,67491655         df =    8,39
No. of cases: 48           adjusted R?= ,60823277        p =  ,000000
               Standard error of estimate: 1,544474752
Intercept: 12,830009051  Std.Error: 4,356569  t(   39) = 2,9450  p =  ,0054
```

Рисунок 1.21 – Оценка коэффициента детерминации переменной x_1

Все расчеты остальных коэффициенты детерминации производятся аналогичным образом. В результате получили:

$$\hat{R}^2_{x_1/x_2x_3x_4x_5x_6x_7x_8x_9} = 0{,}67498$$

$$\hat{R}^2_{x_2/x_1x_3x_4x_5x_6x_7x_8x_9} = 0{,}0965$$

$$\hat{R}^2_{x_3/x_1x_2x_4x_5x_6x_7x_8x_9} = 0{,}4486$$

$$\hat{R}^2_{x_4/x_1x_2x_3x_5x_6x_7x_8x_9} = 0{,}7198$$

$$\hat{R}^2_{x_5/x_1x_2x_3x_4x_6x_7x_8x_9} = 0{,}0933$$

$$\hat{R}^2_{x_6/x_1x_2x_3x_4x_5x_7x_8x_9} = 0{,}2676$$

$$\hat{R}^2_{x_7/x_1x_2x_3x_4x_5x_6x_8x_9} = 0{,}6301$$

$$\hat{R}^2_{x_8 / x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_9} = 0{,}6234$$

Рисунок 1.22 – Результаты вычислений оценок коэффициента детерминации

Анализ оценок коэффициентов детерминации показал наличие тесной линейной связи между объясняющей переменной X_4 и всеми остальными признаками, то же самое можно сказать о переменных X_7, X_8, X_1.

3. Для вычисления собственных чисел матрицы $X^T X$ воспользуемся функцией eiganvals из Mathcad (рисунок 1.23).

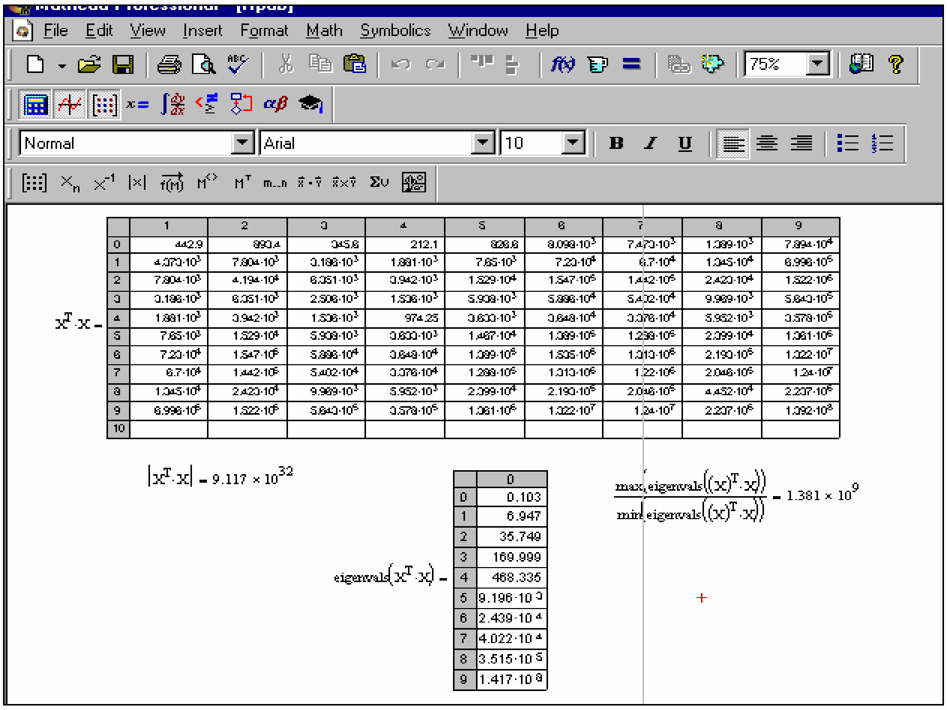


Рисунок 1.23 – Результаты вычислений в программе Mathcad

Таким образом, можно говорить о наличии мультиколлинеарности между объясняющими переменными X1,….,X17.

Все вышеприведенные методы устранения мультиколлинеарности реализуются в ППП Statistica. Рассмотрим метод пошаговой регрессии, используя модуль «множественная регрессия».

Установка флажка в поле **Advanced options** модуля множественная регрессия (рисунок 1.9) позволит перейти к диалоговому окну **Model Definition,** открывающему возможность выбора метода анализа, среди которых методы пошаговой регрессии и гребневой (метод ридж-регрессии). В прокручиваемом списке методов можно выбрать один из методов пошаговой регрессии. В модуле реализованы две процедуры отбора переменных, каждая из которых может давать различный конечный набор переменных: последовательное включение **(Forward stepwise)** и последовательное исключение **(Backward stepwise).**

В данном случае выбран пошаговый метод включения:

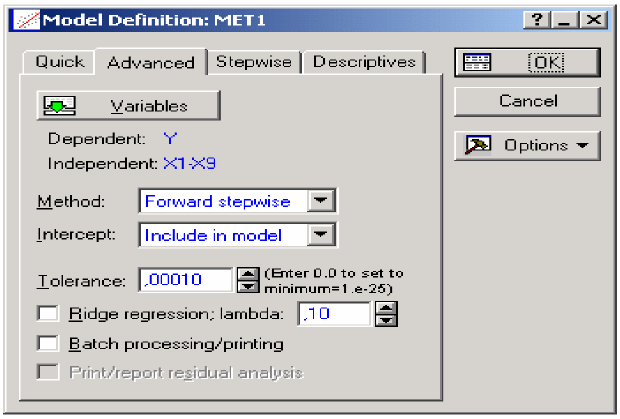
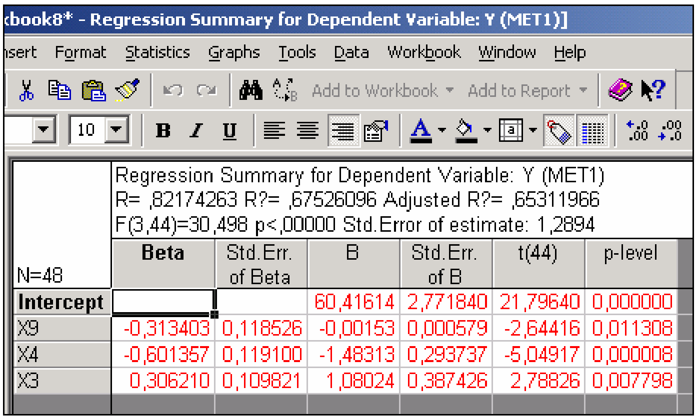


Рисунок 1.24 - Выбор метода оценивания параметров регрессионной модели

Результаты расчетов приведены в виде отчета на рисунке 1.24.



Regression Summary for Dependent Variable: Y (MET1)
R= ,82174263 R?= ,67526096 Adjusted R?= ,65311966
F(3,44)=30,498 p<,00000 Std.Error of estimate: 1,2894

N=48	Beta	Std.Err. of Beta	B	Std.Err. of B	t(44)	p-level
Intercept			60,41614	2,771840	21,79640	0,000000
X9	-0,313403	0,118526	-0,00153	0,000579	-2,64416	0,011308
X4	-0,601357	0,119100	-1,48313	0,293737	-5,04917	0,000008
X3	0,306210	0,109821	1,08024	0,387426	2,78826	0,007798

Рисунок 1.25 – Результаты оценивания параметров линейной модели множественной регрессии методом пошаговой регрессии

Были исследованы также регрессионные остатки, анализ которых показал нормальность их распределения (рисунок 1.26).

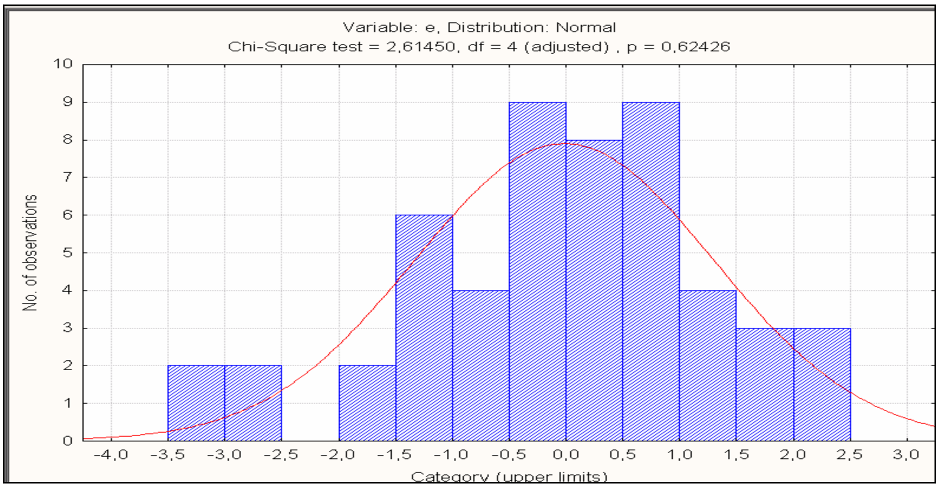


Рисунок 1.26 – Гистограмма распределения регрессионных остатков

В результате проведения пошаговой регрессии получили следующую оценку уравнения регрессии:

$$\hat{y} = \underset{(2,27)}{60{,}42} + \underset{(0,39)}{1{,}08}\,x_3 - \underset{(0,29)}{1{,}48}\,x_4 - \underset{(0,0006)}{0{,}0015}\,x_9$$

Оценка уравнения регрессии значима т.к. нулевая гипотеза отклонена; коэффициенты при переменных также значимы. Коэффициент детерминации составил 0,675, т.е. 67,5% доли вариации результирующей переменной объясняется переменными X_3, X_4 и X_9, а 32,5% доли вариации, вероятно, объясняется неучтенными в модели факторами.

Согласно полученной модели, можно сделать вывод о том, что увеличение количества браков приводит к росту ожидаемой продолжительности жизни мужчин в среднем на 1,08 лет, при росте количества разводов ожидаемая продолжительность жизни мужчин в среднем сокращается на 1,48 лет, при увеличении числа зарегистрированных преступлений ожидаемая продолжительность жизни мужчин в среднем также сокращается на 0,0015 лет.

1.9 Вопросы для самоконтроля

1 Назовите основные задачи эконометрики.

2 Приведите примеры эконометрических моделей.

3 Сформулируйте условия Гаусса-Маркова.

4 Как определяются стандартные ошибки оценок коэффициентов регрессии?

5 Как проверить значимость уравнения регрессии и его коэффициентов?

6 Как проверяется адекватность модели?

7 В чем суть коэффициента детерминации и в каких пределах он изменяется?

8 Раскройте понятие полной и частичной мультиколлинеарности.

9 Укажите причины и признаки мультиколлинеарности.

10 Укажите формальные признаки мультиколлинеарности.

11 К чему ведет отбрасывание незначимых коэффициентов в модели регрессии.

12 Поясните суть пошаговой регрессии.

13 Поясните суть «ридж-регрессии».

2 Обобщенная линейная модель множественной регрессии. Обобщенный метод наименьших квадратов

При моделировании реальных социально-экономических процессов можно столкнуться с ситуациями, в которых условия классической линейной модели множественной регрессии оказываются нарушенными. Если в качестве исходных статистических данных используются временные или пространственно-временные ряды, то чрезмерно нереалистичным становится, как правило, условие взаимной некоррелированности. При работе с пространственно-неоднородными данными излишне оптимистичным становится выполнение условия гомоскедастичности регрессионных остатков. К примеру, часто приходится анализировать регрессионные модели, в которых меняется разброс остатков около линии регрессии, например, возрастая пропорционально значениям функции регрессии. Такая ситуация наблюдается при исследовании зависимости расходов на потребление от уровня доходов семей. В этом случае можно ожидать, что в более обеспеченных семьях вариация расходов выше, чем в малообеспеченных, что приводит к росту дисперсии регрессионных остатков с увеличением расходов семьи.

2.1 Обобщенная линейная модель множественной регрессии

Линейная модель множественной регрессии

$$Y = X\beta + \varepsilon, \tag{2.1}$$

для которой нарушено 4 и/или 5 условие Гаусса-Маркова называется **обобщенной линейной моделью множественной регрессии** (ОЛММР), а именно [1]:

1) $x_1,\ldots,x_k$ – детерминированные переменные;

2) ранг матрицы X равен "$k+1$" – среди признаков нет линейно зависимых;

3) $M\varepsilon_i = 0$, $i = \overline{1,n}$ - нет систематических ошибок в измерении y;

4\`) $\Sigma_\varepsilon = M(\overline{\varepsilon}\overline{\varepsilon}^T) = \sigma_0{}^2\Sigma_0$, $\Sigma_0 \neq E_n$ - положительно-определенная, симметричная матрица.

В обобщенной модели множественной регрессии ковариации между регрессионными остатками могут быть отличными от нуля, а дисперсии могут различаться у различных объектов. В дальнейшем рассмотрим два типа наиболее распространенных эконометрических моделей: линейную модель множественной регрессии с гетероскедастичными остатками (регрессионные остатки которой не отвечают требованиям гомоскедастичности) и линейную модель множественной регрессии с автокоррелированными остатками (регрессионные остатки являются коррелированными между собой). Наша задача, как и прежде, заключается в оценивании вектора β, а также изучении свойств полученных оценок.

2.2 Свойства МНК-оценок и обобщенный метод наименьших квадратов (ОМНК)

МНК-оценки коэффициентов регрессионной модели, определенные соотношениями (1.5) остаются и в рамках ОЛММР состоятельными (при тех же требованиях к матрице наблюдений X) и несмещенными (доказательство несмещенности оценок $b_{МНК}$ в задаче оценивания параметров ОЛММР в точности повторяет доказательство этого факта в условиях КЛММР – (1.10) и (1.10а)). Однако полученные ранее формулы (1.11) и (1.13) для ковариационной матрицы $\Sigma_{b_{мнк}}$ оказываются неработоспособными, неприемлемыми в условиях ОЛММР [1]:

$$\begin{aligned}\sum\nolimits_{\bar b} &= M[((X^T X)^{-1} X^T \bar\varepsilon)((X^T X)^{-1} X^T \varepsilon)^T)] = M[(X^T X)^{-1} X^T \bar\varepsilon\bar\varepsilon^T (X^T X)^{-1}] = \\ &= (X^T X)^{-1} X^T M(\bar\varepsilon\bar\varepsilon^T) X (X^T X)^{-1} = (X^T X)^{-1} X^T \sigma_0{}^2 \Sigma_0 X (X^T X)^{-1}\end{aligned} \qquad (2.2)$$

в то время как для классической модели $\sum_{\bar b} = \sigma^2 (X^T X)^{-1}$.

Таким образом, оценка b, найденная обычным методом наименьших квадратов (1.5) хотя и будет несмещенной и состоятельной, но не будет эффективной. Для получения эффективной оценки преобразуем исходные данные с помощью любой невырожденной матрицы C: $\underbrace{C^{-1}Y}_{Y_{пр}} = \underbrace{C^{-1}X}_{X_{пр}}\beta + \underbrace{C^{-1}\varepsilon}_{\varepsilon_{пр}}$ ($Y_{пр} = X_{пр}\beta + \varepsilon_{пр}$). Преобразуемая модель является КЛММР, тогда оценки коэффициентов определяются следующим образом [1,2,7]:

$$b_{ОМНК} = (X_{пр}^T X_{пр})^{-1} X_{пр}^T Y_{пр} = (X^T (C^T)^{-1} X)^{-1} X^T (C^T)^{-1} C^{-1} Y = (X^T \Sigma_0^{-1} X)^{-1} (X^T \Sigma_0^{-1} Y). \qquad (2.3)$$

Оценки, определенные соотношением (2.3) называются оценки обобщенного метода наименьших квадратов.

При этом несмещенная оценка остаточной дисперсии, ковариационной матрицы $\Sigma_{b_{ОМНК}}$ имеет вид:

$$\hat S^2_{ост(ОМНК)} = \frac{1}{n-k-1} (Y - Xb_{ОМНК})^T \Sigma_0^{-1} (Y - Xb_{ОМНК}), \qquad (2.4)$$

$$\hat\Sigma_{b_{ОМНК}} = \hat S^2_{ост(ОМНК)} (X^T \Sigma_0^{-1} X)^{-1}. \qquad (2.5)$$

Выборочный коэффициент детерминации определяется по формуле (2.6):

$$\hat{R}^2_{ОМНК} = 1 - \frac{\overline{e}^T \Sigma_0^{-1} \overline{e}}{(\overline{Y} - \overline{Y}_{cp})^T \Sigma_0^{-1} (\overline{Y} - \overline{Y}_{cp})} . \qquad (2.6)$$

Однако здесь в $\hat{R}^2_{ОМНК}$ вообще говоря может выходить за пределы [0;1]. Поэтому выражение (2.6) используется лишь как приближенная характеристика качества модели.

В заключение отметим, что для применения обобщенного метода наименьших квадратов необходимо уметь оценивать ковариационную матрицу Σ_0 или подбирать матрицу C, что удается только в некоторых случаях, рассмотренных ниже.

2.3 Обобщенная линейная модель множественной регрессии с гетероскедастичными остатками

Линейная модель множественной регрессии

$$Y = X\beta + \varepsilon , \qquad (2.7)$$

для которой нарушено 4 условие Гаусса-Маркова называется **обобщенной линейной моделью множественной регрессии с гетероскедастичными остатками**, а именно [1]:

1) $x_1,...,x_к$ – детерминированные переменные;

2) ранг матрицы X равен "$к+1$" – среди признаков нет линейно зависимых;

3) $M\varepsilon_i = 0$, $i = \overline{1,n}$ - нет систематических ошибок в измерении y;

4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma_i^{\ 2}$, $i = \overline{1,n}$

5) $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0$, $i \neq j$, $i = \overline{1,n}$ $j = \overline{1,n}$

4') $\Sigma_\varepsilon = M\overline{\varepsilon}\overline{\varepsilon}^T = \sigma_0^2\Sigma_0$ (где Σ_0 - диагональная матрица, но $\Sigma_0 \neq E_n$; σ_0^2 - некоторый общий множитель диагональных элементов Σ_ε).

Рассмотрим пример моделирования зависимости y (ден. ед.) – среднедушевых сбережений от x (ден. ед.) – среднедушевого дохода за один месяц в семье, извлеченных из совокупности семей, однородных по своему потребительскому поведению (таблица 2.1). Естественно ожидать, что с ростом дохода разброс денежных сбережений относительно условной средней величины (функции регрессии) увеличивается (рисунок 2.1). Таким образом, наблюдается рост дисперсии регрессионных остатков с увеличением дохода семьи.

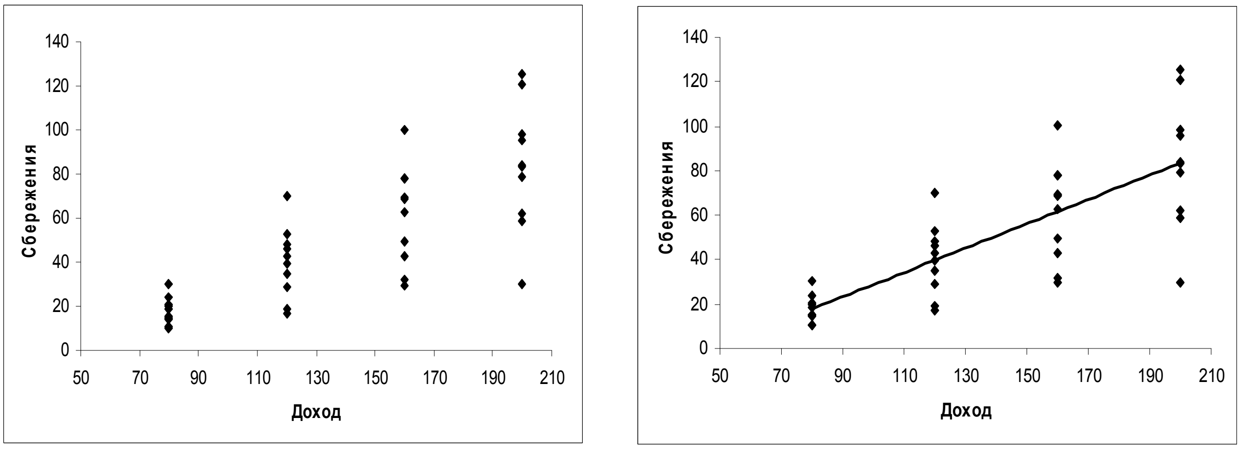


Рисунок 2.1 – Графическое изображение зависимости сбережений от дохода

Влияние гетероскедастичности на свойства МНК-оценок коэффициентов регрессионной модели с гетероскедастичными остатками имеет двойственный характер. Во-первых, оценки, полученные с помощью МНК, в этом случае уже не будут обладать наименьшей дисперсией среди всех линейных несмещенных оценок. Во-вторых, стандартные ошибки таких оценок будут смещены, это обстоятельство приводит к тому, что при проверке гипотез о несущественности параметров регрессионной модели будет снижаться достоверность статистических выводов.

Таблица 2.1 - Данные о среднедушевых сбережений и доходах за месяц

Средне-душевой доход (ден.ед.)	Средне-душевые сбережения (ден.ед.)	Средние сбережения для семей данной группы (ден. ед.)	Оценка среднеквадратического отклонения $\widehat{S}$ и коэффициента вариации $\widehat{V}$ сбережений для семей данной доходной группы
1	2	3	4
$x_1 = x_2 = \ldots = = x_{10} = x_1^0 = 80$	$y_1 = 12.5$ $y_2 = 10.7$ $y_3 = 18.5$ $y_4 = 14.9$ $y_5 = 24.1$ $y_6 = 10.3$ $y_7 = 14.2$ $y_8 = 31.0$ $y_9 = 20.4$ $y_{10} = 20.0$	$\overline{y}\left(x_1^0\right) = = \frac{1}{10}\sum_{i=1}^{10} y_i = 17.9$	$\widehat{S}(x_1^0) = \left[\frac{1}{9}\sum_{i=1}^{10}\left(y_i - \overline{y}\left(x_1^{(0)}\right)\right)^2\right]^{\frac{1}{2}} = 6.4$ $\widehat{V}\left(x_1^{(0)}\right) = 36\%$
$x_{11} = x_{12} = \ldots = = x_{20} = x_2^0 = 120$	$y_{11} = 70.1$ $y_{12} = 35.0$ $y_{13} = 43.0$ $y_{14} = 29.0$ $y_{15} = 17.0$ $y_{16} = 48.2$ $y_{17} = 18.9$ $y_{18} = 53.0$ $y_{19} = 39.4$ $y_{20} = 46.2$	$\overline{y}\left(x_2^0\right) = = \frac{1}{10}\sum_{i=11}^{20} y_i = 40$	$\widehat{S}(x_2^0) = \left[\frac{1}{9}\sum_{i=11}^{20}\left(y_i - \overline{y}\left(x_2^{(0)}\right)\right)^2\right]^{\frac{1}{2}} = = 16.6$ $\widehat{V}\left(x_2^{(0)}\right) = 40\%$
$x_{21} = x_{22} = \ldots = = x_{30} = x_3^0 = 160$	$y_{21} = 49.6$ $y_{22} = 69.4$ $y_{23} = 77.8$ $y_{24} = 43.0$ $y_{25} = 31.8$ $y_{26} = 62.6$ $y_{27} = 100.2$ $y_{28} = 68.8$ $y_{29} = 78.0$ $y_{30} = 29.8$	$\overline{y}\left(x_3^0\right) = = \frac{1}{10}\sum_{i=21}^{30} y_i = 61.1$	$\widehat{S}(x_3^0) = \left[\frac{1}{9}\sum_{i=21}^{30}\left(y_i - \overline{y}\left(x_3^{(0)}\right)\right)^2\right]^{\frac{1}{2}} = = 26.6$ $\widehat{V}\left(x_3^{(0)}\right) = 37\%$
$x_{31} = x_{32} = \ldots = = x_{40} = x_4^0 = 200$	$y_{31} = 125.5$ $y_{32} = 88.3$ $y_{33} = 62.0$ $y_{34} = 58.8$ $y_{35} = 84.0$ $y_{36} = 79.0$ $y_{37} = 95.5$ $y_{38} = 120.8$ $y_{39} = 98.7$ $y_{40} = 29.7$	$\overline{y}\left(x_4^0\right) = = \frac{1}{10}\sum_{i=31}^{40} y_i = 84.2$	$\widehat{S}(x_4^0) = \left[\frac{1}{9}\sum_{i=31}^{40}\left(y_i - \overline{y}\left(x_4^{(0)}\right)\right)^2\right]^{\frac{1}{2}} = = 28.4$ $\widehat{V}\left(x_4^{(0)}\right) = 34\%$

2.3.1 Выявление и тестирование гетероскедастичности

Исследование модели регрессии начинается в предположении справедливости условий Гаусса-Маркова. После оценки ее коэффициентов МНК мы должны подтвердить справедливость предположений 4 и 5, для чего оцениваются регрессионные остатки $e_i = y_i - \hat{y}_i$. Для проверки на гетероскедастичность, в случае множественной функции регрессии строится график зависимости $/e_i/$, упорядоченных в порядке возрастания значений объясняющей переменной x_l, относительно которой модель исследуется на геретоскедастичность. Если по мере возрастания переменной x_l наблюдается, например, тенденция к возрастанию (или убыванию) $|e_i|$ и т.п., то можно предположить наличие гетероскедастичности. Высказанное предположение о наличие гетероскедастичности следует проверить с помощью приведенных ниже тестов.

Тест ранговой корреляции Спирмена

Тест Спирмена используется, если есть основание предполагать, что дисперсия регрессионных остатков изменяется прямо или обратно пропорциональна значению объясняющей переменной (x_l).

Вычисляется коэффициент ранговой корреляции Спирмена [2]:

$$r_{x/e/} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n^3 - n}, \tag{2.8}$$

где $D_i = rang(x_l) - rang(e)$.

Гипотеза об отсутствии (H_0) или наличии (H_1) явления гетероскедастичности формируется следующим образом:

$H_0 : \rho_{xe} = 0$ (нет гетероскедастичности)

$H_1 : \rho_{xe} \neq 0$ (есть гетероскедастичность)

Для проверки гипотезы используется статистика $t = r_{x/e/}\sqrt{n-1}$, которая в случае справедливости нулевой гипотезы имеет нормальный закон распределения с параметрами 0 и 1.

Если нулевая гипотеза не принимается (то есть существует гетероскедастичность), то для оценки коэффициентов линейной модели множественной регрессии используется ОМНК. Структура матрицы Σ_0 для реализации ОМНК имеет следующий вид:

$$\Sigma_0 = \begin{pmatrix} x_{1l}^2 & 0 & \cdots & 0 \\ 0 & x_{2l}^2 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & x_{nl}^2 \end{pmatrix}, \textit{если имеется тенденция к росту модуля регрессионных остатков}$$

$$\textit{с ростом } x_l, \textit{то есть } r_{xe} > 0$$

$$\text{или } \Sigma_0 = \begin{pmatrix} \frac{1}{x_{1l}^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_{2l}^2} & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_{nl}^2} \end{pmatrix}, \textit{в противном случае}.$$

Тест Голдфелда-Квандта

Тест Голдфелда-Квандта используется, если есть основание предполагать, что дисперсия регрессионных остатков пропорциональна значению объясняющей переменной (x_l), вариацией которых порождается гетероскедастичность.

Выдвигается гипотеза [2, 3]:

$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$ (нет гетероскедастичности)

$H_1 : \exists i, j : \sigma_i^2 \neq \sigma_j^2$ (есть гетероскедастичность)

Шаги теста:

1 – Упорядочить в порядке возрастания значения объясняющей переменной (x_l).

2 – Упорядочить наблюдаемые значения результативного признака и объясняющих переменных в порядке возрастания (x_l).

3 – Сформировать (по результатам упорядочивания) выборки, состоящие из n′ первых и n″ последних объектов наблюдения, где $n' = n'' = \dfrac{n - 0.25n}{2}$.

4 – Оцениваются уравнения регрессии Y на X, сформированных по выборкам объемами n′ и n″:

5 – Вычисляются регрессионные остатки $\bar{e}' = Y' - \hat{Y}'$ и $\bar{e}'' = Y'' - \hat{Y}''$ и их суммы квадратов отклонений: $Q' = (\bar{e}')^T \cdot \bar{e}'$ и $Q'' = (\bar{e}'')^T \cdot \bar{e}''$.

6 – Для проверки нулевой гипотезы строится статистика $F = \dfrac{\max\{Q'; Q''\}/(n' - k - 1)}{\min\{Q'; Q''\}/(n' - k - 1)}$, которая в случае справедливости нулевой гипотезы имеет закон распределения Фишера-Снедекора с числом степеней свободы $\nu_1 = n' - k - 1$, $\nu_2 = n' - k - 1$.

Если гипотеза об отсутствии гетероскедастичности отвергнута, при этом $Q^{/} < Q^{//}$, то в качестве оценки Σ_0 можно взять $\Sigma_0 = \begin{pmatrix} x_{1l}^2 & 0 & \cdots & 0 \\ 0 & x_{2l}^2 & \cdots & 0 \\ \vdots & \vdots & ... & \vdots \\ 0 & 0 & \cdots & x_{nl}^2 \end{pmatrix}$, а в случае $Q^{/} > Q^{//}$, оценка матрица $\Sigma_0 = \begin{pmatrix} \frac{1}{x_{1l}^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_{2l}^2} & \cdots & 0 \\ \vdots & \vdots & ... & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_{nl}^2} \end{pmatrix}$.

Критерий Глейзера

Критерий Глейзера позволяет более точно аппроксимировать зависимость между регрессионными остатками и объясняющей переменной. В качестве такой аппроксимации берется линейная регрессия абсолютных значений регрессионных остатков $/e_i/$ на $/x_l/^{\gamma}$ [1, 2, 3]:

$$|e_i| = \alpha + \beta |x_{il}|^{\gamma} + \delta_i, \tag{2.9}$$

где

$\gamma -$ *параметр, который подбирается из промежутка* $(-3;3)$ [2]

$\delta_i -$ *случайная компонента, удовлетворяющая свойствам*:

$$M(\delta_i, \delta_j) = \begin{cases} 0, \text{ если } i = j \\ \sigma_{\delta}^2, \text{ если } i \neq j \end{cases}$$

Если среди моделей вида (2.9), хотя бы для одного значения γ найдется значимая, то есть гипотеза $H_0 : \beta = 0; H_1 : \beta \neq 0$ будет отвергнута, то наличие гетероскедастичности будет доказано. Выбрав среди значимых моделей $|\hat{e}_i| = \hat{\alpha} + \hat{\beta}|x_{il}|^{\hat{\gamma}}$ с наибольшим коэффициентом детерминации, построим оценку $\hat{\Sigma}_0 = \begin{pmatrix} (\hat{\alpha} + \hat{\beta}|x_{1l}|^{\hat{\gamma}})^2 & 0 & 0 & ... & 0 \\ 0 & (\hat{\alpha} + \hat{\beta}|x_{2l}|^{\hat{\gamma}})^2 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & (\hat{\alpha} + \hat{\beta}|x_{nl}|^{\hat{\gamma}})^2 \end{pmatrix}$.

Точечный прогноз для ОЛММР с гетероскедастичными остатками, определяется значением функции регрессии в точке $\overline{X}_{n+1} = (1, x_{n+1,1}, x_{n+2,1}, ..., x_{n+1,k})^T$:

$$\hat{y}_{n+1} = \overline{X}_{n+1}{}^T b_{омнк}. \tag{2.10}$$

Замечание. Зачастую не удается выявить гетероскедастичность с помощью тестов, позволяющих приближенно оценить Σ_0, тогда используют МНК-оценки коэффициентов регрессии, а для статистических исследований используют оценку $\hat{\Sigma}_b$, уточненную по формулам Уайта и Невье-Веста [3].

Проиллюстрируем вышеизложенное на примере.

Пример 3. По данным 20 семей (таблицы А3, Приложение А) построена МНК - оценка уравнения регрессии, описывающая зависимость между расходами на питание ($y, ден.ед.$) и доходами ($x, ден.ед.$):

$$\hat{y} = \underset{(4,5)}{20{,}84} + \underset{(0{,}0022)}{0{,}44}\;x,\ \hat{R}^2 = 0{,}916$$

Проведем исследование регрессионных остатков на гетероскедастичность.

На рисунке 2.2 представлен график зависимости упорядоченных в порядке возрастания значений доходов (x) и абсолютных значений регрессионных остатков, значения которых получены на основе построенной регрессионной модели (таблицы А3, Приложение А). Как видно, прослеживается тенденция к росту абсолютных значений регрессионных остатков с ростом объясняющей переменной, что позволяет предположить наличие гетероскедастичности. Используя критерии ранговой корреляции Спирмена, Голдфелда-Квандта, проверим выдвинутое предположение.

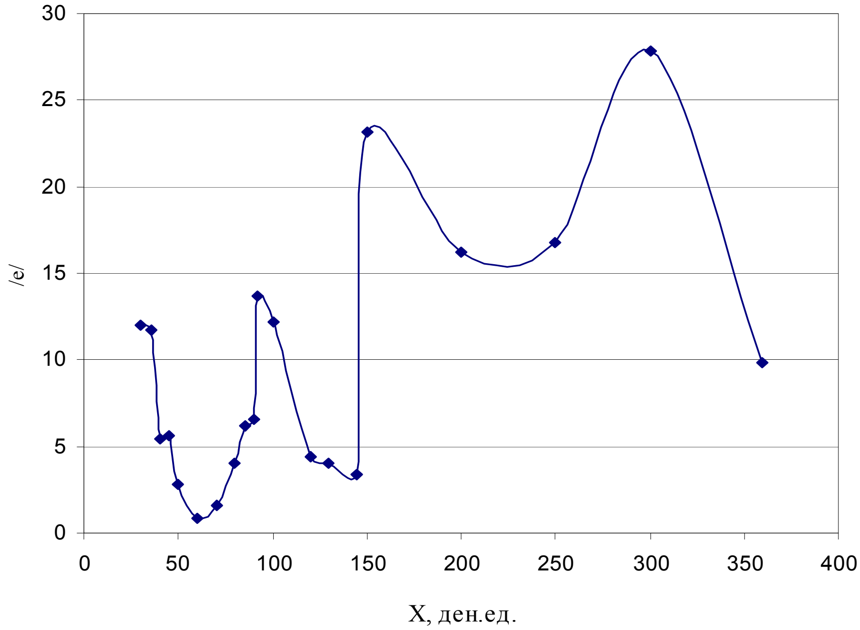


Рисунок 2.2 - Зависимость между упорядоченными в порядке возрастания значениями доходами (x) и абсолютными значеними регрессионных остатков

Рассчитаем коэффициент ранговой корреляции Спирмена

$r_{x/e/} = 1 - \dfrac{6\sum\limits_{i=1}^{n} D_i^2}{n^3 - n} = 1 - \dfrac{6 \cdot 739{,}5}{20^3 - 20} = 0{,}444$ (расчет D_i представлен в таблице А2).

$H_0 : \rho_{xe} = 0$ (нет гетероскедастичности)

$H_1 : \rho_{xe} \neq 0$ (есть гетероскедастичность)

Для проверки гипотезы используется статистика. По таблице нормального распределения определяем критическое значение на уровне значимости $\alpha = 0{,}05$ $t_{крит} = 1{,}96$, $t_{набл} = 1{,}92$. Так как $/t_{набл}/ < t_{крит}$, то нулевая гипотеза об отсутствии гетероскедастичности принимается.

Для применения теста Голдфелда-Квандта необходимо сформировать выборки, состоящие из n′ первых и n″ последних объектов наблюдения, где $n' = n'' = \dfrac{n - 0.25n}{2} = 8$.

Оценим уравнения регрессии Y на X, сформированных по выборкам объемами n′ и n″:

$\hat{y}' = 5{,}785 + 0{,}629X$, $\hat{R}^2 = 0{,}924$ и $\hat{y}'' = 38{,}566 + 0{,}364X$, $\hat{R}^2 = 0{,}814$, по которым оцениваются регрессионные остатки и определяются суммы их квадратов (таблица 2.2).

Таблица 2.2 – Расчет критерия Голдфелда-Квандта

№ п/п	x	$\hat{y}$	$\bar{e}$	$\bar{e}^2$
1	30	24,7	-2,7	7,29
2	36	28,4	-3,4	11,56
3	40	30,9	2,1	4,41
4	45	34,1	0,9	0,81
5	50	37,2	2,8	7,84
6	60	43,5	4,5	20,25
7	70	49,8	0,2	0,04
8	80	56,1	-4,1	16,81
Сумма				**69,01**
13	120	82,2	-4,1	16,81
14	130	85,9	-3,9	15,21
15	145	91,3	-3,3	10,89
16	150	93,2	16,8	282,24
17	200	111,4	13,6	184,96
18	250	129,6	-15,6	243,36
19	300	147,8	-22,8	519,84
20	360	169,8	19,4	376,36
Сумма				**1649,7**

Для проверки нулевой гипотезы $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$ (нет гетероскедастичности); $H_1 : \exists i, j : \sigma_i^2 \neq \sigma_j^2$ (есть гетероскедастичность) строится статистика $F=\dfrac{\max\{Q^{/};Q^{//}\}/(n^{'}-k-1)}{\min\{Q^{/};Q^{//}\}/(n^{'}-k-1)}=\dfrac{1649{,}7}{69{,}01}=23{,}9$. По таблице распределения Фишера определяется $F_{крит}(0{,}05;6;6)=4{,}28$. Так как $F>F_{крит}$, то нулевая гипотеза об отсутствии гетероскедастичности отвергается.

Для устранения гетероскедастичности воспользуемся ОМНК (2.3), где матрица $\hat{\Sigma}_0=\begin{pmatrix}30^2 & 0 & ... & 0\\ 0 & 36^2 & ... & 0\\ ... & ... & .... & ...\\ 0 & ... & 0 & 360^2\end{pmatrix}$. Оценка уравнения регрессии в этом случае имеет вид: $\hat{y}=\underset{(2,1)}{19{,}46}+\underset{(0,0001)}{0{,}32}\,x$. Таким образом, учет гетероскедастичности позволил существенно скорректировать зависимость между доходами и расходами, и, в частности, с ростом доходов на 1, расходы на питание возрастут в среднем на 0,32 ден.ед., а не на 0,44 ден. ед.

2.4 Обобщенная линейная модель множественной регрессии с автокоррелированными остатками

2.4.1 Автокорреляционная зависимость первого порядка

Линейная модель множественной регрессии

$$Y=X\beta+\varepsilon, \tag{2.11}$$

для которой нарушено 5 условие Гаусса-Маркова называется **обобщенной линейной моделью множественной регрессии с автокоррелированными остатками** [1].

Рассмотрим частный случай автокорреляции регрессионных остатков – автокорреляционную зависимость первого порядка, которая описывается соотношением [1]:

$$\varepsilon_i=\rho\varepsilon_{i-1}+\delta_i, \tag{2.12}$$

где ρ - коэффициент корреляции между ε_i и ε_{i-1} $(/\rho/\le 1)$.

δ_i - случайные величины, удовлетворяющие условиям $M[\delta_i]=0,\ M[\delta_i,\delta_j]=\begin{cases}\sigma_\delta^2,\ при\ \ i=j;\\ 0,\ \ при\ i\ne j\end{cases}$.

Анализируя выражение (2.12) при $0<\rho<1$, можно утверждать, что величина ε_i будет сохранять определенный знак на некоторых промежутках изменения i (рисунок 2.3). Подобную ситуацию будем характеризовать как **положительную автокорреляцию регрессионных остатков**.

При $-1<\rho<0$ величина ε_i будет почти регулярно менять знак (рисунок 2.4). Такую ситуацию будем называть **отрицательной автокорреляцией регрессионных остатков**.

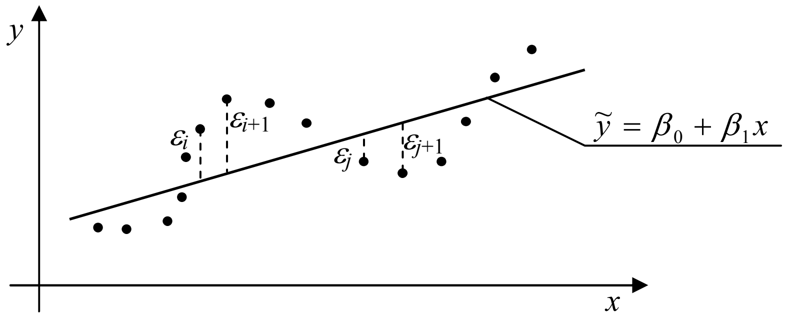


Рисунок 2.3 – Иллюстрация положительной автокорреляции

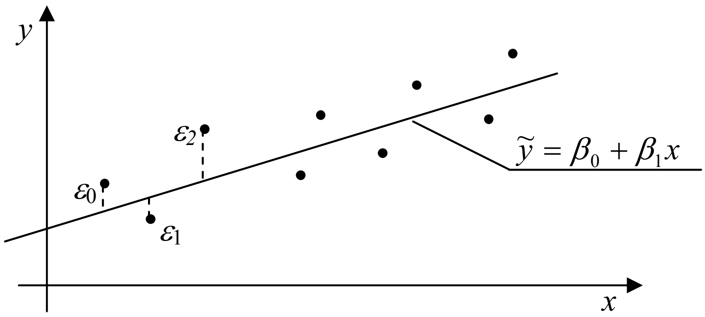


Рисунок 2.4 - Иллюстрация отрицательной автокорреляции

Приведенное геометрическое свойство ЛММР с автокоррелированными остатками будем использовать для выявления подозреваемых на такое свойство моделей.

Пример 4. На основе (таблица А.4, Приложение А) данных о ежегодным потреблением бананов (y) и годовым семейным доходом (x) построена оценка уравнения регрессии $\hat{y} = \underset{(1,23)}{5{,}089} + \underset{(0,198)}{0{,}734}x$, $\hat{R}^2 = 0{,}637$. Проведем исследование регрессионных остатков на наличие автокорреляции, для этого построим график разброса наблюдаемых значений результативного признака относительно функции регрессии (рисунок 2.5). По графику можно предположить наличие положительной автокорреляции.

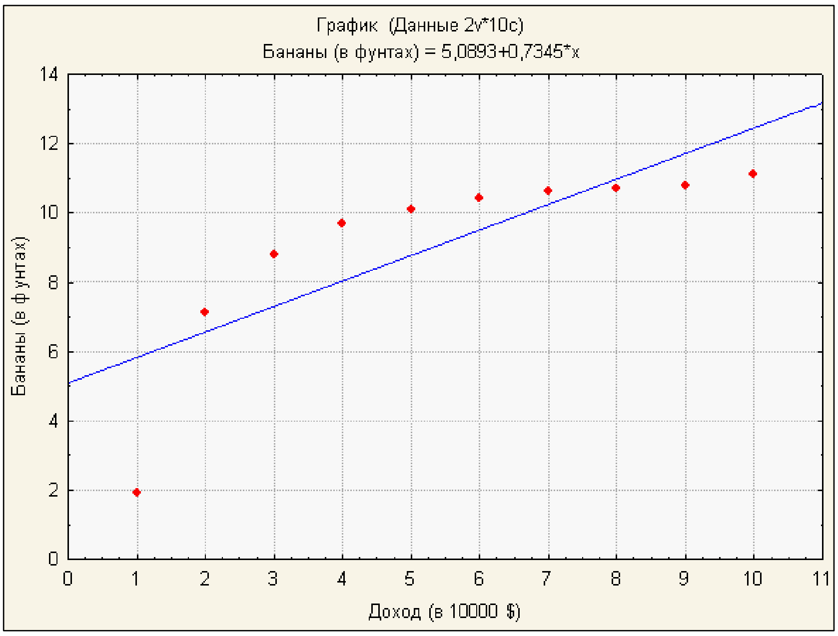


Рисунок 2.5 - График разброса наблюдаемых значений результативного признака относительно функции регрессии.

В случае множественной функции регрессии строится график зависимости e_i, упорядоченных в порядке возрастания значений объясняющей переменной x_l, по которой исследуется автокорреляция. Если по мере возрастания упорядоченной переменной оценки регрессионных остатков сохраняют определенный знак на некоторых промежутках, то можно выдвинуть предположение о положительной автокорреляции. Если e_i, в основном, регулярно меняют знак, то выдвигается предположение об отрицательной автокорреляций. Ниже рассмотрим критерий, позволяющий устанавливать наличие/отсуствие автокорреляции.

2.4.2 Тестирование автокоррелированности регрессионных остатков

Допустим, что регрессионные остатки ЛММР удовлетворяют условию (2.12). С учетом соотношения (2.12) рассмотрим величину

$$D = \frac{M(\varepsilon_i - \varepsilon_{i-1})^2}{M\varepsilon_i^2} = \frac{M\varepsilon_i^2 + M\varepsilon_{i-1}^2 - 2M(\varepsilon_i \varepsilon_{i-1})}{M\varepsilon_i^2} = \frac{\sigma^2(2-2\rho)}{\sigma^2} = 2(1-\rho)\,. \qquad (2.13)$$

Из (2.13), следует, что в случае положительной автокорреляции ($\rho > 0$) будем иметь $0 \le D \le 2$ и, в частности, при ρ близких к единице величина $D \to 0$. В случае отрицательной автокорреляции ($\rho < 0$) величина D может находиться в интервале $2 \le D \le 4$ и, в случае, ρ близких к «-1» величина D близка к «-4».

Проверить подтверждается или нет предположение (гипотеза) о наличии автокорреляции можно попытаться с помощью критерия Дарбина-Уотсона, приведенного ниже:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \approx 2(1-\hat{\rho}), \tag{2.14}$$

где $e_i = y_i - y_i$, $i = 1,n$ - оценки регрессионных остатков, пронумерованных в порядке возрастания проверяемой объясняющей переменной x_l.

Статистика Дарбина-Уотсона обладает свойствами аналогичными свойствам D из (2.13). Действительно, при положительной автокорреляции можно почти всегда считать $e_i \approx e_{i-1}$, а это означает, что DW близко к нулю; при отрицательной автокорреляции почти всегда $e_i \approx -e_{i-1}$, из чего следует - DW близко к 4; в сучае отсутствия оснований подозревать автокорреляцию, в половине случаев считаем, что $e_i \approx e_{i-1}$, а в другой – $e_i \approx -e_{i-1}$, что влечет близкие к 2 значения DW.

Для проверки гипотезы H_0: $\rho = 0$ (нет явления автокорреляции), H_1: $\rho \neq 0$ (есть явление автокорреляции) используется статистика Дарбина-Уотсона (2.14).

По таблице Дарбина-Уотсона находятся (при данном уровне значимости α, числе наблюдений n и объясняющих переменных k) два критических значения: нижнее $d_н$ и верхнее $d_в$.

Если фактически наблюдаемое значение DW (рисунок 2.6):

1) $d_в < DW < 4 - d_в$, то гипотеза об отсутствии автокорреляции принимается.

2) $d_н < DW < d_в$ или $4 - d_в < DW < 4 - d_н$, область неопределенности критерия (вопрос об отвержении или принятии гипотезы остается открытым).

3) $0 < DW < d_н$, то принимается альтернативная гипотеза о положительной автокорреляции.

4) $4 - d_н < DW < 4$, то принимается альтернативная гипотеза об отрицательной автокорреляции.

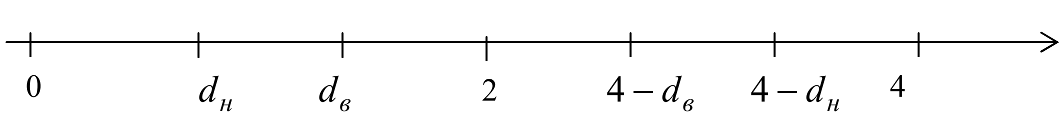


Рисунок 2.6 – Критическая область и область принятия нулевой гипотезы об отсутствии автокорреляции первого порядка

Пример 5 (продолжение примера 4). По графику разброса наблюдаемых значений результативного признака относительно функции регрессии, было сделано предположение о наличии положительной автокорреляции. Для проверки нулевой гипотезы H_0: $\rho = 0$ (нет явления автокорреляции), H_1: $\rho \neq 0$ (есть явление автокорреляции) рассчитаем статистика Дарбина-Уотсона (2.14): $DW = 0{,}866$. По таблице Дарбина – Уотсона для к=1 и n=10 определены два критических значения: $d_н = 1{,}08,\ d_в = 1{,}36$. Так как $0 < DW < d_н$, то принимается альтернативная гипотеза о положительной автокорреляции.

Наличие автокорреляции регрессионных остатков в некоторых практических задачах обусловлено выбором неверной спецификации модели.

Так, в примере 4, наличие автокорреляции является следствием нелинейного характера связи между исследуемыми признаками. Поэтому целесообразным представляется в качестве модели рассмотреть гиперболическую функцию, оценка которой имеет вид:
$\hat{y} = 12{,}08 + 10{,}08\frac{1}{x},\ \hat{R}^2 = 0{,}998;\ DW = 2{,}02$.

В том случае, если изменением спецификации модели не удалось устранить автокорреляцию, то для оценки коэффициентов ЛММР рекомендуется использовать ОМНК, реализация которого сводится к оценке параметра ρ.

2.4.3 Практические рекомендации по оцениванию коэффициентов обобщенной линейной модели множественной регрессии с автокоррелированными остатками

Можно показать [1], что для ОЛММР с автокорреляционной зависимостью первого ковариационная матрица регрессионных остатков имеет вид:

$$\Sigma_\varepsilon = \frac{\sigma_\delta^2}{1-\rho^2}\underbrace{\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}}_{\Sigma_0}.$$

Таким образом, реализация ОМНК для оценки коэффициентов обобщенной линейной модели множественной регрессии с автокоррелированными остатками первого порядка сводится к нахождению параметра ρ.

В ситуациях, когда значение коэффициента корреляции ρ между соседними значениями регрессионных остатков известно, то исследователь не должен испытывать затруднений в практической реализации основных формул ОМНК (2.3) – (2.6).

Поэтому остановимся на ситуации, когда значение параметра ρ априорно не известно. Практически все процедуры, предложенные для реализации

ОМНК в модели регрессии с автокоррелированными остатками при неизвестном значении ρ, имеют итерационных характер [1]. Рассмотрим описание одной из наиболее распространенных процедур подобного типа, известной в литературе под названием **процедуры Кохрейна-Оркатта** [1].

1) МНК оцениваются коэффициенты $\overline{b}^{(1)}$ регрессионной модели $Y = X\beta + \varepsilon$;

2) рассчитываются регрессионные остатки первой итерации: $e_i^{(1)} = y_i - \hat{y}_i^{(1)}$, где $\hat{y}_i^{(1)} = b_0^{(1)} + b_1^{(1)} x_{i1} + ... + b_k^{(1)} x_{ik}$;

3) первое приближение $r^{(1)}$ оценки неизвестного параметра ρ определяется с помощью МНК-оценки коэффициента регрессии ρ в модели $e_i^{(1)} = \rho e_{i-1}^{(1)} + \delta_i^{(1)}$;

4) вычисляются ОМНК-оценки $\overline{b}_{ОМНК}(r^{(1)}) = \overline{b}^{(2)}$ по формуле (2.2), с матрицей $\hat{\Sigma}_0(r^{(1)})$, определенной соотношением (2.12), в котором вместо ρ подставлены $r^{(1)}$;

5) рассчитываются регрессионные остатки второй итерации: $e_i^{(2)} = y_i - \hat{y}_i^{(2)}$, где $\hat{y}_i^{(2)} = b_0^{(2)} + b_1^{(2)} x_{i1} + ... + b_k^{(2)} x_{ik}$ и т.д.

Процесс продолжается до тех пор, пока не будет достигнута требуемая точность (пока r не стабилизируются), а именно пока разность между предыдущей и последующей оценками параметра ρ не станет любого наперед заданного числа.

Точечный прогноз для ОЛММР с автокоррелированными остатками, определяется: $\hat{y}_{n+1} = \overline{X}_{n+1}{}^T b_{омнк} + \hat{\rho} e_n$.

Замечание. При работе с пространственной статистической информацией, наличие автокоррелированных регрессионных остатков, как правило, обусловлено неправильной спецификацией модели. Поэтому в некоторых практических задачах методом устранения автокорреляции является изменение спецификации (вида функции) регрессионной модели.

2.5 Тестовые задания для самоконтроля

1) Оценка вектора коэффициентов обобщенной линейной модели множественной регрессии, найденная методом наименьших квадратов будет обладать свойствами:

а) несмещенности и неэффективности;

б) смещенности и неэффективности;

в) смещенности и эффективности;

г) несмещенности и эффективности.

2) Укажите тип линейной модели множественной регрессии, для которой 4 и 5 условия Гаусса-Маркова имеют вид:

$$M\varepsilon_i\varepsilon_j = \begin{cases} 0, & i \neq j, \quad i = \overline{1,n}, \quad j = \overline{1,n} \\ \sigma_i^2, & i = j, \end{cases}$$

а) обобщенная линейная модель с гетероскедастичными остатками;

б) обобщенная линейная модель с автокоррелированными остатками;

в) классическая линейная модель;

г) линейная модель с гомоскедастичными остатками.

3) К условиям гетероскедастичности в линейной модели множественной регрессии $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ относятся:

а) $M\varepsilon_i\varepsilon_j = \begin{cases} 0, & i = j, \quad i = \overline{1,n}, \quad j = \overline{1,n} \\ \sigma^2, & i \neq j, \end{cases}$;

б) $M(\varepsilon_i\varepsilon_j) = \begin{cases} 1, & i \neq j \\ \sigma^2, & i = j \end{cases} \qquad i = \overline{1,n}, j = \overline{1,n}$;

в) $\Sigma_\varepsilon = \sigma^2 E_n$;

г) $\Sigma_\varepsilon = \sigma_0^2 \Sigma_0, \quad где \quad \Sigma_0 \neq E_n, \quad \Sigma_0 - диагональная\ матрица$.

4) Условие автокорреляции первого порядка регрессионных остатков в линейной модели множественной регрессии $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ указано в ответе:

а) $\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i$, где δ_i - белый шум;

б) $\Sigma_\varepsilon = \sigma^2 E_n$;

в) $M\bar{\varepsilon}\bar{\varepsilon}^T = \sigma^2 E_n$;

г) $cov(\varepsilon_i\varepsilon_j) = M(\varepsilon_i\varepsilon_j) = 0, \quad i \neq j$, $i = \overline{1n}$, $j = \overline{1n}$.

5) Линейная модель множественной регрессии $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$, $i = \overline{1,n}$ с автокоррелированными остатками удовлетворяет условию:

а) $M\varepsilon_i\varepsilon_j = \begin{cases} 0, & i = j, \quad i = \overline{1,n}, \quad j = \overline{1,n} \\ \sigma^2, & i \neq j, \end{cases}$;

б) $cov(\varepsilon_i\varepsilon_j) = M(\varepsilon_i\varepsilon_j) \neq 0$, $i \neq j$, $i = 1,2,...,n$, $j = 1,2,...,n.$;

в) $\Sigma_\varepsilon = \sigma^2 E_n$;

г) $M(\varepsilon_i\varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma_i^2, & i = j \end{cases} \qquad i = \overline{1,n}, \ j = \overline{1,n}$.

6) Укажите тип линейной модели множественной регрессии, для которой ковариационная матрица регрессионных остатков имеет вид:

$$\Sigma_\varepsilon = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & ... & \rho^{n-1} \\ \rho & 1 & \rho & ... & \rho^{n-2} \\ \rho^2 & \rho & 1 & ... & \rho^{n-3} \\ ... & ... & ... & ... & ... \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & ... & 1 \end{pmatrix}$$

а) обобщенная линейная модель множественной регрессии с гетероскедастичными остатками;

б) обобщенная линейная модель множественной регрессии с автокоррелированными, но гомоскедастичными остатками;

в) линейная модель множественной регрессии с гомоскедастичными остатками;

г) классическая линейная модель множественной регрессии.

7) Гипотеза о наличие гетероскедастичности по переменной x_j подтвердилась с помощью теста Голдфелда-Квандта, используемая статистика $F = \dfrac{\max(Q_1;Q_2)/(n' - k - 1)}{\min(Q_1;Q_2)/(n' - k - 1)}$ (где Q_1, Q_2-это сумма квадратов оценок регрессионных остатков для первых и последних n' наблюдений), при этом оказалось, что $Q_2 > Q_1$. Укажите вид матрицы $\hat{\Sigma}_0$ для оценки коэффициентов ОМНК.

а) $\hat{\Sigma}_0 = \begin{pmatrix} \frac{1}{x_{1j}^2} & 0 & 0 & ... & 0 \\ 0 & \frac{1}{x_{2j}^2} & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \frac{1}{x_{nj}^2} \end{pmatrix}$;

б) $\hat{\Sigma}_0 = \begin{pmatrix} x_{1j}^2 & 0 & 0 & ... & 0 \\ 0 & x_{2j}^2 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & x_{nj}^2 \end{pmatrix}$;

в) $\hat{\Sigma}_0 = \begin{pmatrix} (\hat{\alpha} + \hat{\beta}\left|x_{1j}\right|^{\gamma})^2 & 0 & 0 & ... & 0 \\ 0 & (\hat{\alpha} + \hat{\beta}\left|x_{2j}\right|^{\gamma})^2 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & (\hat{\alpha} + \hat{\beta}\left|x_{nj}\right|^{\gamma})^2 \end{pmatrix}$;

г) $\hat{\Sigma}_0 = \begin{pmatrix} x_{1j}^2 & 1 & 1 & ... & 1 \\ 1 & x_{2j}^2 & 1 & ... & 1 \\ ... & ... & ... & ... & ... \\ 1 & 1 & 1 & ... & x_{nj}^2 \end{pmatrix}$.

8) Если наблюдаемое значение статистики Дарбина-Уотсона (DW) находится в пределах $4 - d_l \le DW \le 4$, то:

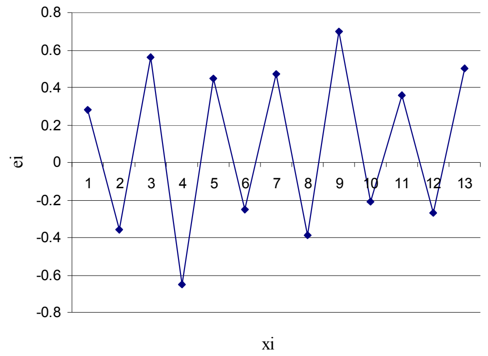
а) принимается гипотеза об отсутствии автокорреляции;

б) принимается гипотеза о наличии положительной автокорреляции;

в) принимается гипотеза о наличии отрицательной автокорреляции;
г) вывод о наличии автокорреляции не определен.

9) Укажите, какой можно сделать вывод, если статистика Дарбина-Уотсона (DW) находится в пределах:
а) принимается гипотеза об отсутствии автокорреляции;
б) принимается гипотеза о наличии положительной автокорреляции;
в) принимается гипотеза о наличии отрицательной автокорреляции;
г) вывод о наличии автокорреляции не определен.

10) На рисунке представлена зависимость оценок регрессионных остатков от значений объясняющей переменной. Укажите, какую гипотезу можно выдвинуть.



а) о наличие положительной автокорреляции;
б) об отсутствии автокорреляции;
в) о наличии отрицательной автокорреляции;
с) по рисунку ничего определенного сказать нельзя.

11) Статистики $t = r_{x|e|}\sqrt{n-1}$, где $r_{x|e|}$ - коэффициент ранговой корреляции Спирмена, применяемая для проверки остатков нормальной модели регрессии на наличие гетероскедастичности, имеет распределение:
а) Стьюдента;
б) Пирсона;
в) нормальный закон распределения;
г) Фишера-Снедекора.

12) Наличие гетероскедастичности регрессионных остатков по переменной x_j подтверждается с помощью теста Голдфелда-Квандта на уровне значимости α=0,05, при k=3 и n'=21, в варианте:
Табличный материал: $F_{кр}(0{,}05;17;17) = 2{,}272$; $F_{кр}(0{,}05;21;21) = 2{,}084$;
$F_{кр}(0{,}05;3;21) = 3{,}073$; Статистика Фишера-Снедекора: $F = \dfrac{\max(Q_1;Q_2)\big/(n'-k-1)}{\min(Q_1;Q_2)\big/(n'-k-1)}$

а) $\begin{aligned} Q_1 &= \bar{e}_1^T \cdot \bar{e}_1 = 3{,}05 \\ Q_2 &= \bar{e}_2^T \cdot \bar{e}_2 = 1{,}69 \end{aligned}$

б) $\begin{aligned} Q_1 &= \bar{e}_1^T \cdot \bar{e}_1 = 1{,}36 \\ Q_2 &= \bar{e}_2^T \cdot \bar{e}_2 = 6{,}04 \end{aligned}$;

в) $\begin{aligned} Q_1 &= \bar{e}_1^T \cdot \bar{e}_1 = 4{,}15 \\ Q_2 &= \bar{e}_2^T \cdot \bar{e}_2 = 2{,}03 \end{aligned}$;

г) $\begin{aligned} Q_1 &= \bar{e}_1^T \cdot \bar{e}_1 = 2{,}10 \\ Q_2 &= \bar{e}_2^T \cdot \bar{e}_2 = 4{,}528 \end{aligned}$.

13) Наблюдаемое значение статистики Дарбина-Уотсона DW=3,2. При уровне значимости "α =0,05", объеме выборки "n=17",числе объясняющих переменных "k=2" справедливо утверждение:

Табличный материал: при $\alpha = 0{,}05$ n=15 k=2 $d_н = 0{,}95$ $d_в = 1{,}54$

при $\alpha = 0{,}05$ n=17 k=1 $d_н = 1{,}13$ $d_в = 1{,}38$

при $\alpha = 0{,}05$ n=17 k=2 $d_н = 1{,}02$ $d_в = 1{,}54$

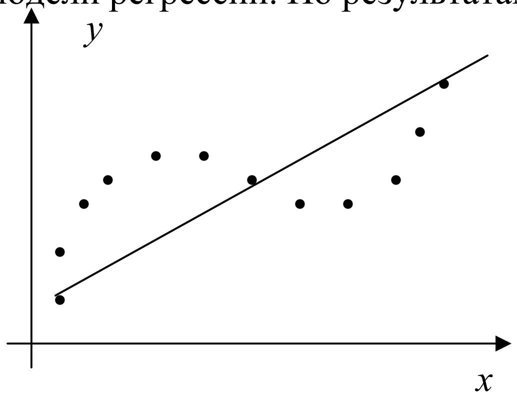
а) принимается гипотеза об отсутствии автокорреляции первого порядка

б) принимается гипотеза о наличии положительной автокорреляции первого порядка

в) принимается гипотеза о наличии отрицательной автокорреляции первого порядка

г) вывод о наличии автокорреляции не определен.

14) На рисунке представлено корреляционное поле и график оцененной модели регрессии. По результатам можно выдвинуть гипотезу



а) о наличии положительной автокорреляции;
б) об отсутствии автокорреляции;
в) о наличии отрицательной автокорреляции;
г) по рисунку ничего определенного сказать нельзя.

15) Прибыль предприятия является случайной величиной, зависящей от размера основных фондов. Естественно ожидать, что для больших предприятий вариация прибыли будет выше, чем для малых. В этом случае при построении классической линейной модели множественной регрессии следует ожидать:
а) отсутствие гетероскедастичности;
б) наличие положительной автокорреляции;
в) наличие отрицательной автокорреляции;
г) наличие гетероскедастичности.

2.6 Индивидуальное задание №2

По регрессионной модели, полученной в результате выполнения индивидуального задания №1:

1) исследовать регрессионные остатки на гетероскедастичность, используя один из рассмотренных тестов (Спирмена, Голдфелда-Квандта, Глейзера);

2) в случае ее наличии устранить, используя практически реализуемый ОМНК.

2.7 Пример выполнения индивидуального задания №2[5]

Пусть по результатам оценивания получена оценка уравнения регрессии, описывающая зависимость между расходами на образование (y) и валовом внутреннем продуктом (x_1) (рисунок 2.7): $\hat{y} = \underset{(0,91)}{-2,32} + \underset{(0,002)}{0,067}\,x$.

[5] Индивидуальные задания и пример их выполнения разработаны авторами совместно с преподавателем кафедры математических методов и моделей в экономике Жемчужниковой Ю.А.

Regression Summary for Dependent Variable: y (Spreadsheet
R= ,98966700 R?= ,97944078 Adjusted R?= ,97879830
F(1,32)=1524,5 p<0,0000 Std.Error of estimate: 4,7108

N=34	Beta	Std.Err. of Beta	B	Std.Err. of B	t(32)	p-level
Intercept			-2,31990	0,913881	-2,53851	0,016197
x1	0,989667	0,025347	0,06689	0,001713	39,04458	0,000000

Рисунок 2.7 - Результаты оценивания параметров регрессионной модели

Как видно из отчета, коэффициент при переменной X_1 значимо отличается от нуля, т.к. нулевая гипотеза (H_0: $\beta 1=0$ при альтернативной H_1: $\beta 1 \neq 0$) отклонена. Регрессионная модель адекватна экспериментальным данным.

Поскольку построили значимую регрессионную модель, то следующим этапом является исследование регрессионных остатков на наличие/отсутствие гетероскедастичности.

Наличие гетероскедастичности можно предположить по графику зависимости остатков $/e_i/$ от упорядоченных по возрастанию значений той объясняющей переменной, вариацией которой возможно порождается гетероскедастичность. Для построения графика можно воспользоваться MS Excel. Для этого из ППП «Statistica» в MS Excel скопируем упорядоченные столбцы регрессионных остатков и значений объясняющей переменной. После чего выделяем указанные столбцы и в строке меню выбираем пункт «Мастер диаграмм». После выполнения запросов данного пункта получаем следующий график (рисунок 2.8).

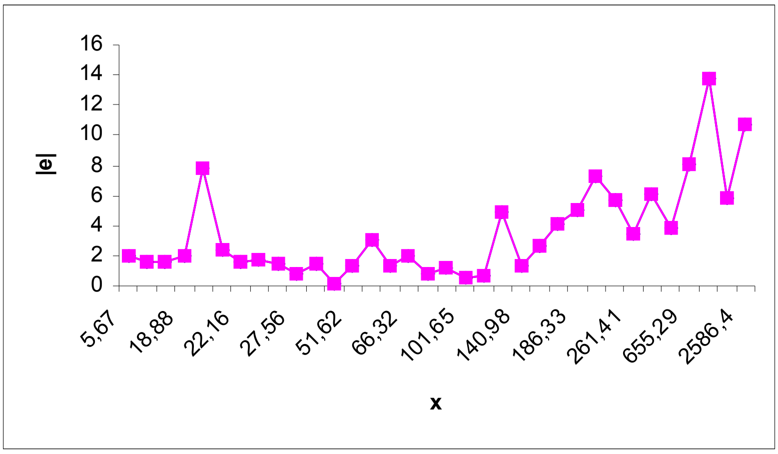


Рисунок 2.8 – График зависимости модуля значений регрессионных остатков и значений объясняющей переменной

На графике видно, что модули регрессионных остатков имеют тенденцию к росту при увеличении значений объясняющей переменной. Следовательно, можно заподозрить гетероскедастичность по переменной X_1.

Тест ранговой корреляции Спирмена

Для определения коэффициента ранговой корреляции Спирмена в меню системы открыть **Statistics - Критерии** и выбрать в появившемся меню строку **Nonparametrics – Непараметрические критерии.** На экране откроется окно.

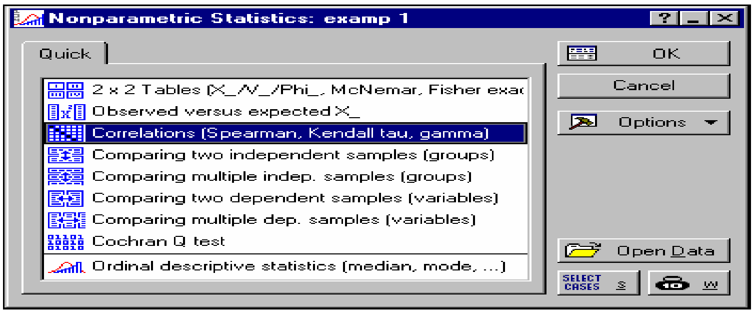


Рисунок 2.9 – Выбор пунктов меню для расчета коэффициента ранговой корреляции Спирмена

Выбираем пункт **Correlaitions (Spearmam, …) – Корреляциия (Спирмен, …),** далее в открывшемся окне выбираем переменные между которыми необходимо рассчитать данный коэффициент (в нашем случае между регрессионными остатками и значениями объясняющей переменной).

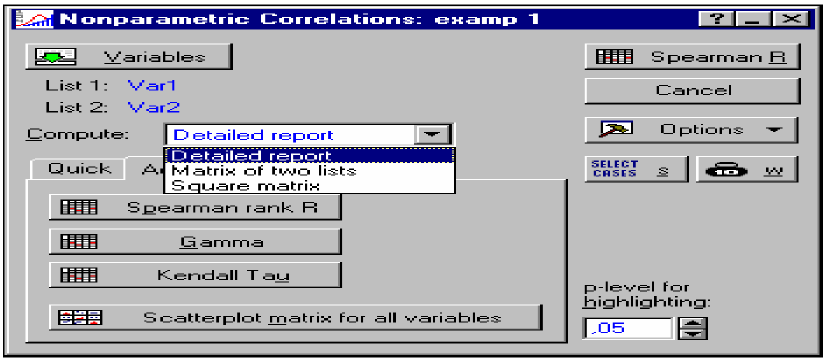
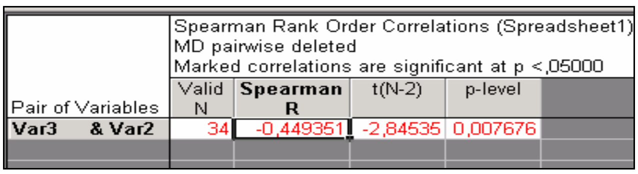


Рисунок 2.10 – Выбор пунктов для расчета коэффициента ранговой корреляции Спирмена

После нажатия на кнопку **Spearman rank R** программа произведет расчеты (рисунок 2.11).



Spearman Rank Order Correlations (Spreadsheet1)
MD pairwise deleted
Marked correlations are significant at p <,05000

Pair of Variables	Valid N	Spearman R	t(N-2)	p-level
Var3 & Var2	34	-0,449351	-2,84535	0,007676

Рисунок 2.11 – Результаты оценивания теста ранговой корреляции Спирмена

Во втором столбце данного окна определяется оценка коэффициента ранговой корреляции Спирмена, в третьем – значении статистики с помощью которой проверяется нулевая гипотеза $H_0: \rho_{xe} = 0,$ (нет явления гетероскедастичности), $H_1: \rho_{xe} \neq 0,$ (есть явление гетероскедастичности). В данном случае нулевая гипотеза отвергается, то есть можно сделать вывод о наличии гетероскедастичности.

Тест Голдфелда-Квандта.

Для реализации проверки на гетероскедастичность с помощью теста Голдфелда-Квандта сначала необходимо упорядочить данные по возрастанию независимой переменной (ВНП); величину, определяющую число исключенных средних наблюдений ($\mathrm{d} \approx \frac{1}{4}\mathrm{n}$), возьмем равной 10.

Тогда сначала необходимо построить оценки уравнений регрессии для первых 12 стран, а затем для 12 последних.

Для этого в окне Multiple Regression сначала укажем в качестве независимой переменной – X, а в качестве зависимой – Y. При задании параметров в данном окне следует осуществить отбор того подмножества наблюдений, которое будет участвовать в расчетах, используя для этого кнопку **Select Cases**. После нажатия этой кнопки откроется диалоговое окно **Cases Selection Conditions,** в котором следует задавать условия отбора наблюдений.

При построении оценки уравнения регрессии по первым 12 наблюдениям в строке этого окна **include if** (включать если) укажем неравенство v0<=12, т.е. с 1 по 12 наблюдение (рисунок 2.12).

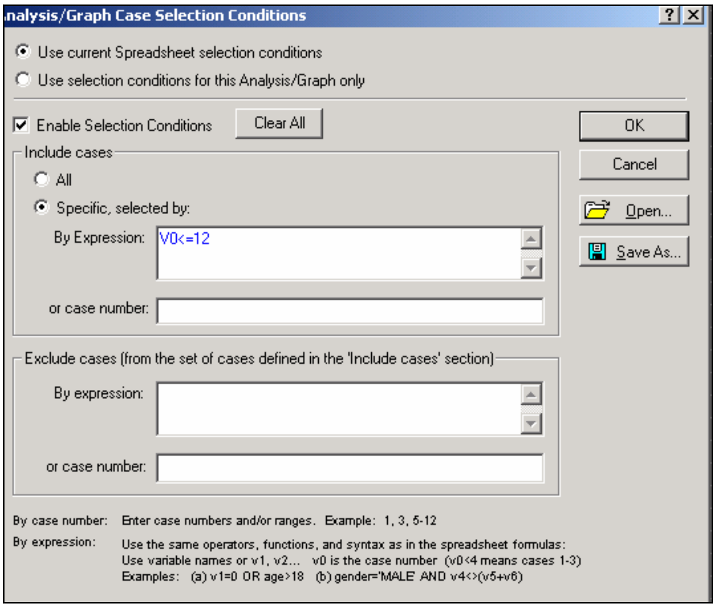


Рисунок 2.12 – Выбор подмножества наблюдений для оценки уравнения регрессии

При оценивании коэффициентов модели по 12 последним (с 13 по 24 наблюдение) укажем неравенство v0>=13. Символы v0 в логических операциях определяют номер строки. После задания всех необходимых параметров, произведем вычисления.

В нижней части окна результатов регрессионного анализа нажмем на кнопку **ANOVA (**рисунок 2.13).

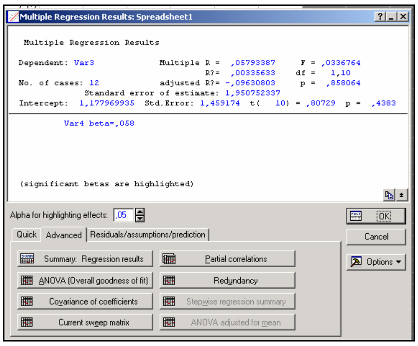


Рисунок 2.13 – Результаты оценивания для подмножества, состоящего из 12 первых наблюдений

На рисунке 2.14 представлена таблица с результатами дисперсионного анализа. Требуемое для теста Голдфелда-Квандта значение суммы квадратов остатков находится на пересечении строки **Residual** (остаточная) и столбца **Sums of Squares** (сумма квадратов).

	Analysis of Variance; DV: Var3 (Spreadsheet1)				
Effect	**Sums of Squares**	df	Mean Squares	F	p-level
Regress.	0,12815	1	0,128153	0,033676	0,858064
Residual	38,05435	10	3,805435		
Total	38,18250				

Рисунок 2.14 – Результаты дисперсионного анализа для 12 первых наблюдений

Значение этого показателя для модели, построенной по последним 12 наблюдениям равно Q_2=378,78.

$F_{набл} = \dfrac{Q_2}{Q_1} = \dfrac{373,78}{38,054} = 9,82$. Критическое значение определим в окне Excel, используя **Мастер функций**, категорию **Статистические**, функцию **FРАСПОБР** (рисунок 2.15).

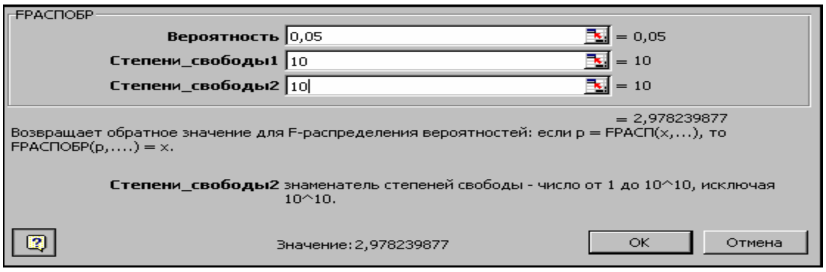


Рисунок 2.15 – Расчет критического значения для F- статистики

Так как $F_{набл.} > F_{крит}$, следовательно нулевая гипотеза об отсутствии гетероскедастичности отклоняется.

В данном случае, подтвердилась прямая зависимость дисперсий регрессионных остатков от значений переменной X_1, и оценка матрицы Σ_0:

$$\hat{\Sigma}_0 = \begin{pmatrix} x_{11}^2 & 0 & 0... & 0 \\ 0 & x_{21}^2 & 0.... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & x_{n1}^2 \end{pmatrix} \qquad (2.13)$$

Тест Глейзера

В тех случаях, когда хотим установить более точный характер поведения σ_i, целесообразно использовать тест Глейзера. Оценив регрессионные остатки исходной модели, будем строить модель (2.8):

$|e_i| = \alpha + \beta |x_{i1}|^{\gamma} + \delta_i$, где $/ e_i / = / y_i - \hat{y}_i /$.

Перебирая γ в промежутке от $-\infty$ до ∞ (при чем если регрессионные остатки имеют тенденцию к росту, то γ целесооброзно выбирать из промежутка $(0,+\infty)$) оценивают регрессионную модель вида (2.8). Отбираются только значимые модели, поскольку в случае отклонения нулевой гипотезы (H_0: β=0 при альтернативной H_1: $\beta \neq 0$), гипотеза об отсутствии гетероскедастичности не принимается.

В нашем случае, подбирая γ в промежутке от –3 до 3 были оценены уравнения с использованием модуля множественная регрессия. Результаты представлены в обобщенном виде на рисунке 2.16.

γ	b_0	S_{b_0}	b_1	S_{b_1}	R^2	F
-1	4,25	0,648	-28,9	14,602	0,11	3,92
-0.5	5,77	0,864	-16,87	5,285	0,24	10,194
0.5	0,684	0,574	0,231	0,037	0,55	39,822
1	2,396	0,459	0,004	0,0009	0,45	26,075
1.5	2,9	0,47	0,00008	0,00002	0,33	15,513

Рисунок 2.16 – Результаты оценивания регрессионной модели вида (2.8)

Статистически значимые оценки были получены для последних четырех значений γ. Наилучший результат соответствует значению γ=0,5 (рисунок 2.17).

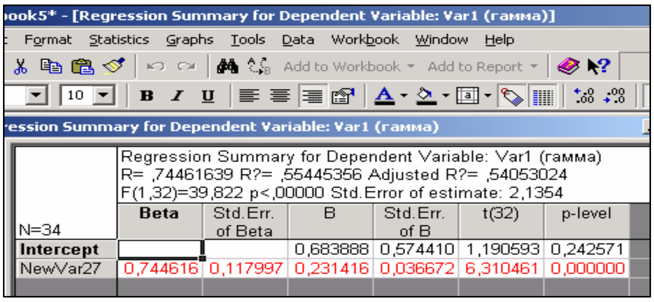


Рисунок 2.17 – Результаты оценки регрессионной модели, соответствующей значению γ=0,5

Таким образом, наилучшая аппроксимация $|\hat{e}_i| = 0.6839 + 0.231\sqrt{x_i}$. Оценка матрицы Σ_0 вида:

$$\hat{\Sigma}_0 = \begin{pmatrix} \left(\hat{\alpha} + \hat{\beta}|x_{11}|^{0,5}\right)^2 & 0 & 0... & 0 \\ 0 & \left(\hat{\alpha} + \hat{\beta}|x_{21}|^{0,5}\right)^2 & 0.... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \left(\hat{\alpha} + \hat{\beta}|x_{n1}|^{0.5}\right)^2 \end{pmatrix}, \qquad (2.14)$$

Построение обобщенной линейной модели множественной регрессии

ОМНК-оценки коэффициентов уравнения регрессии: $b = \left(X^T \hat{\Sigma}_0^{-1} X\right)^{-1} X^T \hat{\Sigma}_0^{-1} Y$.

Для определения вектора-оценок коэффициентов уравнения регрессии воспользуемся функциональными возможностями Mathcad. Сначала матрицы X, Y, $\hat{\Sigma}^{-1}{}_0$ формируем в Excel, сохраняем в текстовом формате, затем открываем Mathcad, в меню **Insert** выбираем пункт **Components** (рисунок 2.18).

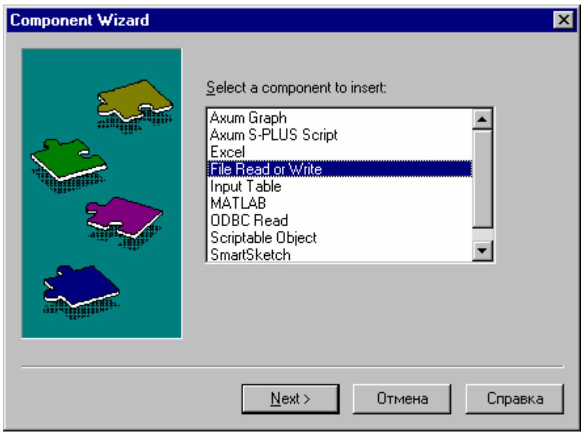


Рисунок 2.18 – Выбор пунктов меню для импортирования данных из MS Excel

В появившемся окне находим пункт **File Read or Word**. В окне **File Read or Word** нажимаем на кнопку **Browse - Обзор** и открываем текстовый файл, в котором сохранили матрицы X, Y, $\hat{\Sigma}^{-1}{}_0$. Выбрав нужный файл, нажимаем на кнопку Готово. В появившемся окне полученной матрице присваиваем имя, например X, и соответственно, получаем матрицы X, Y и $\hat{\Sigma}_0^{-1}$.

Используя функциональные возможности Mathcad вычисляем вектор-оценок коэффициентов уравнения регрессии, S_{b_j}, критерии, с помощью которых проверяется значимость модели в целом и отдельных коэффициентов.

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = \underset{(0,91)}{-0,816} + \underset{(4,507\cdot 10^{-6})}{0,064}\, x$$

Согласно полученной модели, при изменении валового внутреннего продукта на 1 млн. руб. расходы на образование возрастут в среднем на 0,064 мл.руб.

Проверим на значимость уравнение регрессии с помощью статистики Фишера – Снедекора:

$$\mathrm{F} = \frac{\mathrm{Q_{фак}}/\mathrm{k}}{\mathrm{Q_{ост}}/(\mathrm{n-k-1})}, \qquad (2.15)$$

где $\mathrm{Q_{фак}} = (\mathrm{X}\bar{b} - \mathrm{y_{ср}})^T\, \hat{\Sigma}_0^{-1}\, (\mathrm{X}\bar{b} - \mathrm{y_{ср}}) = 3931$,

$\mathrm{Q_{ост}} = (\bar{Y} - \mathrm{X}\bar{b})^T\, \hat{\Sigma}_0^{-1} (\bar{Y} - \mathrm{X}\bar{b}) = 120,194$

F= 1046,5

Так как $F_{набл.} > F_{крит}$, следовательно регрессионная модель адекватна экспериментальным данным.

2.8 Индивидуальное задание №3[6]

По регрессионной модели, полученной в результате выполнения индивидуального задания №1:

1) исследовать регрессионные остатки на наличие автокорреляции,

2) в случае ее наличии, используя процедуру Кохрейна-Оркатта, построить ОМНК-оценки параметров ОЛММР.

2.9 Пример выполнения индивидуального задания №3

Пусть по результатам оценивания получена оценка уравнения регрессии, описывающая зависимость между чистым доходом, млрд. долл (y) и оборотом капитала, млрд. долл. (x_1) (рисунок 2.19): $\hat{y} = \underset{(0,086)}{2,945} + \underset{(0,005)}{0,053}\, x$

[6] Индивидуальные задания и пример их выполнения разработаны авторами совместно с преподавателями кафедры математических методов и моделей в экономике Васяниной В.И., Жемчужниковой Ю.А.

N=30	Regression Summary for Dependent Variable: Y (danna1) R= ,88229307 R_= ,77844106 Adjusted R_= ,77052824 F(1,28)=98,377 p<,00000 Std.Error of estimate: ,42341					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(28)	p-level
Intercept			2,944958	0,086370	34,09713	0,000000
x1	0,882293	0,088954	0,053217	0,005365	9,91853	0,000000

Рисунок 2.19 - Результаты оценивания параметров регрессионной модели

Как видно из отчета (рисунок 2.19), регрессионная модель адекватна экспериментальным данным. Исследуем регрессионные остатки на наличие/отсутствие автокорреляции. Для визуального анализа регрессионных остатков построим график с использованием MS Excel (рисунок 2.20).

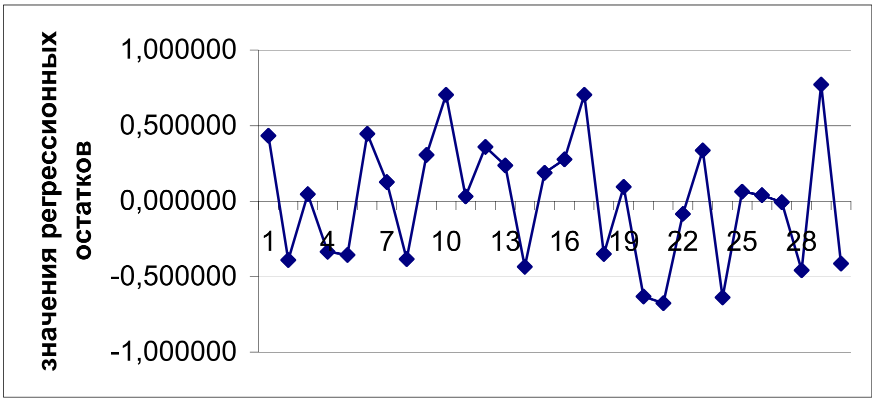


Рисунок 2.20 – График регрессионных остатков

По графику регрессионных остатков можно предположить наличие в регрессионных остатках положительной автокорреляции.

Кроме визуального анализа, существует критерий Дарбина-Уотсона, с помощью которого выявляется автокорреляции первого порядка.

Для вычисления значения данного критерия используется соответствующая процедура ППП Statistica. В окне **Residuals analysis – Анализ остатков** нажмем кнопку **Durbin-Watson statistic – Критерий Дарбина-Уотсона.** На экране появится окно, содержащее значение данного критерия.

- Durbin-Watson d (danna1)

	Durbin-Watson d (danna1) and serial correlation of residuals	
	Durbin-Watson d	Serial Corr.
Estimate	1,232193	0,157172

Рисунок 2.21 – Значение критерия Дарбина-Уотсона и оценка коэффициента корреляции регрессионных остатков

Так как $DW < 2$, то наше предположение о возможном наличии положительной автокорреляции допустимо. Для расчета критического значения воспользуемся таблицей значений статистики Дарбина-Уотсона. В нашем случае для n=30, k=1 получаем d_l=1,35 и d_u=1,49. Так как $DW \le d_1$, то нулевую гипотезу об отсутствии автокорреляции первого порядка ($H_0 : \rho = 0$) отвергаем, т.е. делаем вывод о наличии положительной автокорреляция.

Поскольку подтвердилось предположение о наличии положительной автокорреляции, то для оценивания коэффициентов используем ОМНК.

При наличии автокорреляции первого порядка матрица Σ_0^{-1} будет иметь вид (2.11). Таким образом, задача сводится к оцениванию параметра ρ. Для решения этой задачи воспользуемся процедурой Кохрейна –Оркатта.

На первом этапе МНК находим оценки коэффициентов уравнения регрессии, вычисляем регрессионные остатки $e_i^{(1)}$. На рисунке 2.18 представлены оценки коэффициентов уравнения регрессии. Информация о значениях остатков может быть получена нажатием на кнопку **Summary: Residuals & predicted**. Вектор регрессионных остатков представлен на рисунке 2.22.

8 e
0,432488
-0,390375
0,046367
-0,337069
-0,356990
0,445142
0,127023
-0,384014
0,307473
0,704347
0,030868
0,358953
0,237443
-0,435503
0,188592
0,277805
0,705071
-0,350825
0,096170
-0,632711
-0,675523

Рисунок 2.22 – Значения регрессионных остатков

Оценивая параметр ρ модели регрессии $e_i^{(1)} = \rho e_{i-1}^{(1)} + \delta_i^{(1)}$, получили $\hat{\rho}^{(1)}$=0,157. В качестве оценки матрицы Σ_0^{-1} берем матрицу $\hat{\Sigma}_0^{-1} = \dfrac{1}{1-\hat{\rho}^{(1)2}} \begin{bmatrix} 1 & -\hat{\rho}^{(1)} & \dots & 0 \\ -\hat{\rho}^{(1)} & 1+\hat{\rho}^{(1)2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$. Имея эту матрицу, построим ОМНК-оценки параметров уравнения регрессии.

Для определения вектора-оценок коэффициентов уравнения регрессии воспользуемся функциональными возможностями Mathcad. Сначала матрицы

X, Y, $\hat{\Sigma}^{-1}{}_0$ формируем в Excel, сохраняем в текстовом формате, затем открываем Mathcad, в меню **Insert** выбираем пункт **Components** (рисунок 2.23).

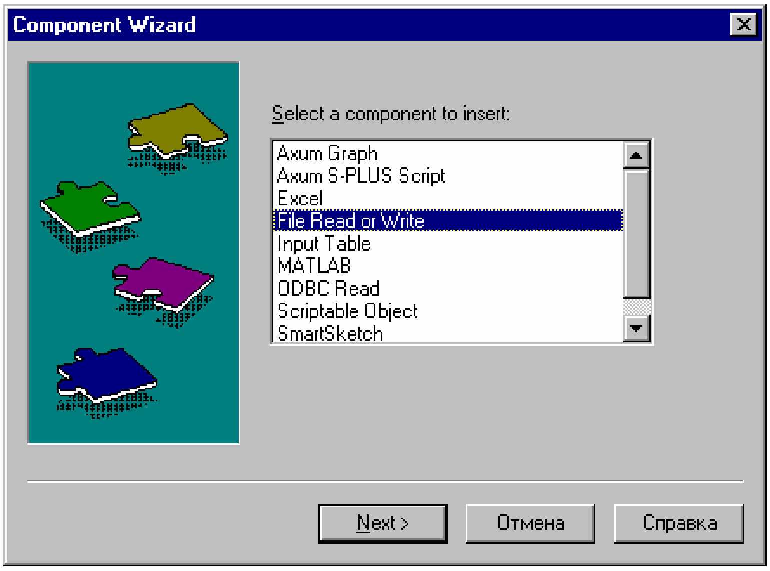


Рисунок 2.23 – Выбор пунктов меню для импортирования данных из MS Excel

В появившемся окне находим пункт **File Read or Word**. В окне **File Read or Word** нажимаем на кнопку **Browse - Обзор** и открываем текстовый файл, в котором сохранили матрицы X, Y, $\hat{\Sigma}^{-1}{}_0$. Выбрав нужный файл, нажимаем на кнопку Готово. В появившемся окне полученной матрице присваиваем имя, например X, и соответственно, получаем матрицы X, Y и $\hat{\Sigma}_0^{-1}$ (рисунок 2.23).

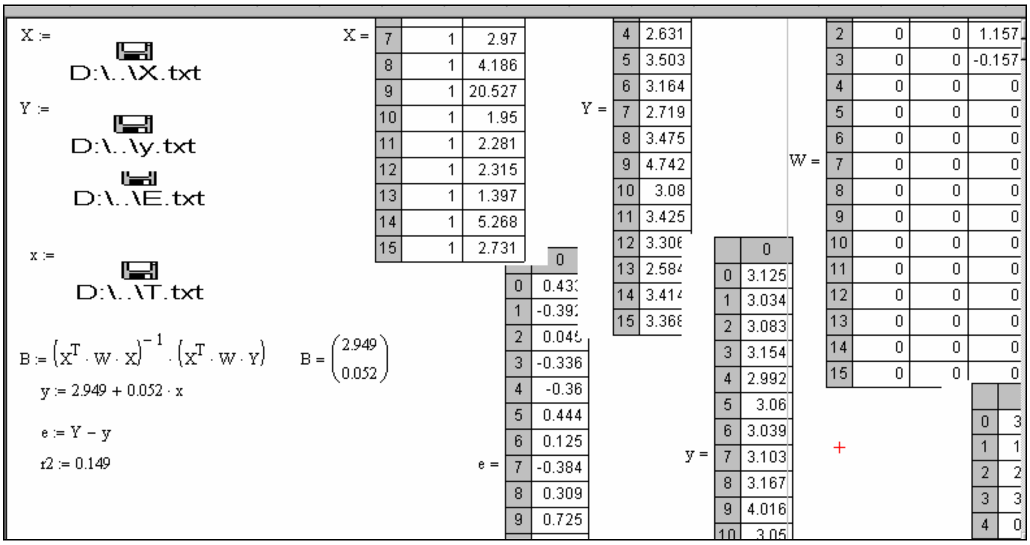


Рисунок 2.24 – Результаты расчетов в Mathcad

Повторим процедуру до тех пор пока соседние $\hat{\rho}^{(k)}$ и $\hat{\rho}^{(k-1)}$ не окажутся между собой приблизительно равны.

На 5 шаге итерации $\hat{\rho}^{(4)} \approx \hat{\rho}^{(5)} = 0{,}120$. На рисунке 2.25 представлены результаты вычислений на последней итерации.

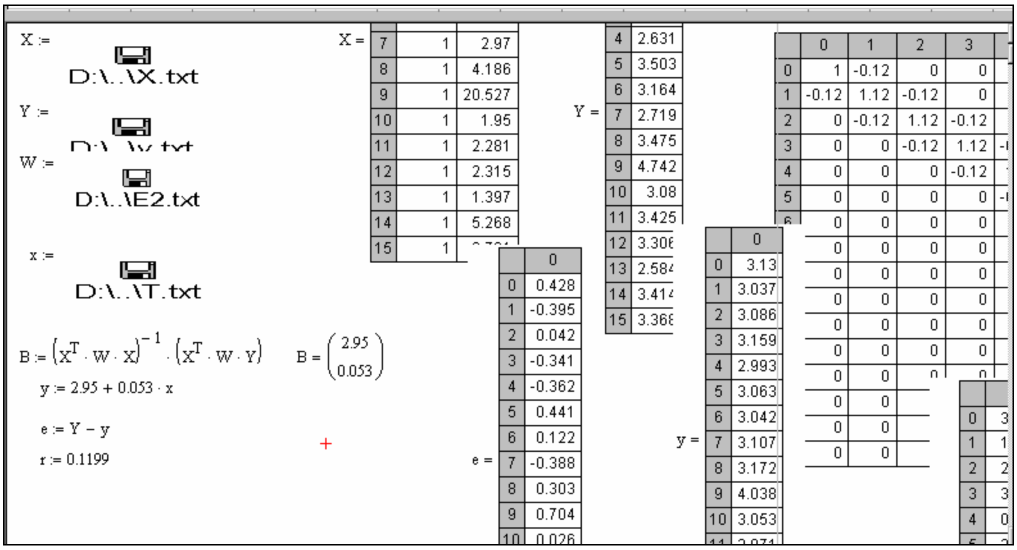


Рисунок 2.25 – Результаты расчетов на последнем шаге

На рисунках 2.24 и 2.25 под матрицей W подразумевается $\hat{\Sigma}_0^{-1}$.

Оценка уравнения регрессии выглядит следующим образом $\hat{y} = 2{,}95 + 0{,}053\,x$.

2.10 Вопросы для самоконтроля

1 Дайте определение ОЛММР с гетероскедастичными остатками.

2 Укажите свойства МНК-оценок ОЛММР с гетероскедастичными остатками.

3 Укажите возможные причины, порождающие гетероскедастичность.

4. Опишите алгоритм анализа модели на наличие гетераскедастичности.

5 Тест Спирмена и ОМНК.

6 Тест Голдфелда-Квандта и ОМНК.

7 Тест Глейзера и ОМНК.

8 Несмещенные оценки ковариационной матрицы вектора коэффициентов в форме Невье-Веста.

9 Несмещенные оценки ковариационной матрицы вектора коэффициентов в форме Уайта.

10 Дайте определение ОЛММР с автокоррелированными остатками.

11 Укажите свойства МНК-оценок ОЛММР с автокоррелированными остатками.

12 Ковариационная матрица регрессионных остатков в ОЛММР с автокоррелированными остатками первого порядка.

13 Укажите причины и признаки автокорреляции.

14 Тест Дарбина-Уотсона для проверки наличия/отсутствия автокорреляции.

15 Процедура Кохрейна-Оркатта.

16 Исследование статистических свойств ОМНК-оценок параметров модели с автокоррелированными остатками.

3 Линейные регрессионные модели с переменной структурой (построение линейной модели по неоднородным данным)

3.1 Проблема неоднородных данных

О линейных регрессионных моделях с переменной структурой будем говорить в ситуациях, когда в ходе сбора исходных статистических данных имеет место косвенное воздействие некоторых качественных факторов, в результате которых происходят скачкообразные сдвиги в структуре анализируемых линейных связей, то есть в значениях коэффициентов регрессии [1, 4]. Введем понятие однородных и неоднородных в регрессионном смысле статистических данных.

Статистические данные будем считать **однородными** (в регрессионном смысле), если все они наблюдаемы при одних и тех же условиях, то есть при одних и тех же значениях качественных переменных [1].

Статистические данные будем считать **неоднородными** (в регрессионном смысле), если они наблюдаемы при различных значениях качественных переменных [1].

На рисунке 3.1 изображена зависимость заработной платы (у) от производительности труда (х) работников, занятых физическим трудом. В этом случае, уровень заработной платы существенно различен для работников мужского пола (подвыборка В) и женского пола (подвыборка А), поэтому можно ожидать, что регрессионная модель, построенная по всей совокупности наблюдений, будет обладать плохим качеством.

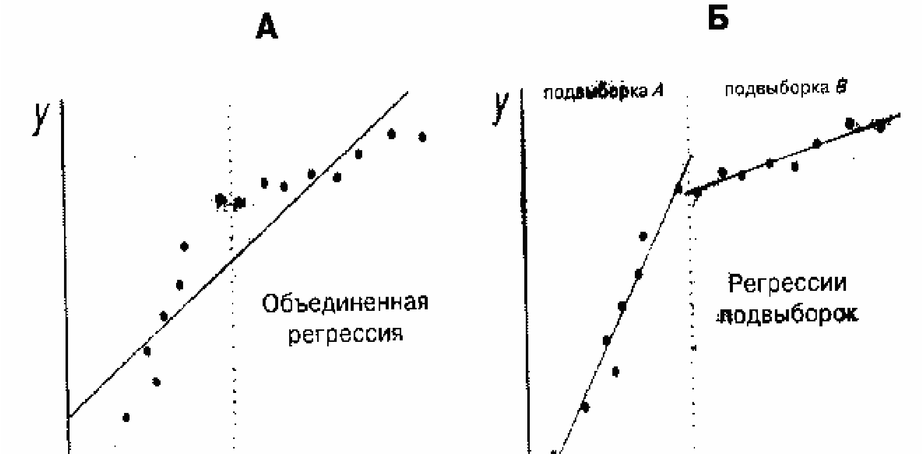


Рисунок 3.1 – Зависимость заработной платы работников (у) от производительности труда (х) с учетом пола работника

Если игнорировать различия по качественным (сопутствующим) признакам, то они или уйдут в остаточную вариацию, ухудшив тем самым модель, либо в неизвестной пропорции станут смешиваться с влиянием тех или иных количественных признаков, искажая степень их влияние на результативный признак. Поэтому возникает проблема отражения в модели регрессии влияния качественных переменных, которые могут быть учтены при моделировании двумя способами (рисунок 3.2).

Первый подход, позволяющий учесть влияние качественных признаков, заключается в том, что оцениваются уравнения регрессии для каждой однородной группы (подвыборки), то есть такой, внутри каждой из которых значения качественных переменных не меняется при переходе от одного наблюдения к другому [1]. Например, в предыдущем примере могут быть оценены отдельно уравнения регрессии для работников мужского пола: $y_i^{(м)} = \beta_0^{(м)} + \beta_1^{(м)} x_i^{(м)} + \varepsilon_i$ и женского пола: $y_i^{(ж)} = \beta_0^{(ж)} + \beta_1^{(ж)} x_i^{(ж)} + \delta_i$. Различия в уровне заработной платы проявятся в различии коэффициентов $b_0^{(м)}$ *и* $b_0^{(ж)}$, при этом сила влияния x на y может быть одинаковой, т.е. $b_1^{(м)} \approx b_1^{(ж)}$.

Если качественный признак ненаблюдаем, либо его значения не были своевременно зарегистрированы при сборе исходных статистических данных, то прямое разбиение исследуемой совокупности на регрессионно однородные группы невозможно и с этой целью приходится использовать методы кластерного анализа в пространстве ($y, x_1, ..., x_к$) [1].

Рисунок 3.2 – Схема учета качественных признаков при моделировании социально-экономических процессов

Второй подход применятся, как правило, когда разбиение на однородные группы, приводит к тому, что одна из групп имеет небольшой объем совокупности, что сказывается на точности оценок коэффициентов; возможны ситуации, когда не представляется возможным проводить регрессионный анализ, т.е. n<k, МНК-оценки построить можно, но нельзя определить их характеристики. В этом случае в регрессионную модель вводятся фиктивные переменные (манекены), то есть строится регрессионная модель с переменной структурой, отражающая неоднородность данных.

3.2 Введение фиктивных переменных в регрессионную модель

Прием введения в анализируемую модель регрессии фиктивных переменных, отражающих влияние на исследуемых результативный признак качественных переменных, оказывается удобным и выгодным в двух отношениях. Во-первых, статистическая надежность (точность) получаемых оценок коэффициентов регрессии будет выше той, которая имелась бы, оценивая коэффициенты отдельно по каждой однородной группе. Во-вторых, в ходе построения регрессионной модели с фиктивными переменными имеется возможность одновременно проверять гипотезы о наличии или отсутствии статистического значимого влияния качественных переменных на структуру анализируемой модели.

Учет влияния качественных переменных на структуру модели осуществляется, как правило, с помощью введения в правую часть регрессионного уравнения определенного числа дихотомических (бинарных) переменных, т.е. таких переменных, которые могут принимать одно из двух возможных значений («нуль» или «единица»):

$$d_j^{(l)} = \begin{cases} 1, \text{ если } i-й \text{ объект обладает } l-м \text{ } \textit{качественным} \text{ свойством;} \\ 0, \textit{ в противном случае} \end{cases}, \qquad (3.1)$$

$$j = 1,2,...,p-1$$

При этом, если качественная переменная имеет p градаций, то для отражения ее влияния на структуру регрессионной модели необходимо вводить $p-1$ фиктивных переменных (введение p фиктивных переменных приведет к вырожденной матрице исходных данных, а следовательно, к нарушению 2-ого условия Гаусса-Маркова, т.е. невозможности получения МНК-оценок) [1, 4]. Конкретная форма, в которой эти переменные будут представлены в анализируемой модели, будет зависеть от допущений о характере влияния качественной переменной на коэффициенты.

Проанализируем зависимость цены двухкомнатной квартиры $(n = 200)$ от ее жилой площади, типа дома («хрущевка», панельный, кирпичный) и наличия телефона. Для учета качественных переменных в соответствии с (3.1) рассмотрим фиктивные переменные:

$$d_1^{(1)} = \begin{cases} 1, \textit{ если дом панельный;} \\ 0, \textit{ в противном случае} \end{cases}; \qquad d_1^{(2)} = \begin{cases} 1, \textit{ если в квартире есть телефон;} \\ 0, \textit{ в противном случае} \end{cases};$$

$$d_2^{(1)} = \begin{cases} 1, \textit{ если дом кирпичный;} \\ 0, \textit{ в противном случае} \end{cases};$$

$$d_3^{(1)} = \begin{cases} 1, \textit{ если дом типа "хрущевка";} \\ 0, \textit{ в противном случае} \end{cases}; \quad i = 1,...,200$$

При использовании трех категорий домов в регрессионную модель необходимо вести только две фиктивные переменные, например, $d_{i1}^{(1)}$ и $d_{i2}^{(1)}$, за базу сравнения взяты дома типа «хрущевки». Оценка уравнения регрессии составила:

$$\hat{y} = \underset{(99,3)}{3020} + \underset{(173.,3)}{1050}\, x + \underset{(550,2)}{1110}\, d_1^{(1)} + \underset{(480,2)}{1060}\, d_2^{(1)} + \underset{(1116,5)}{990}\, d_1^{(2)}$$

Отсюда непосредственно следует, что коэффициент при фиктивной переменной $d_1^{(2)}$ статистически незначим, то есть на цену двухкомнатной квартиры при данном уровне значимости $\alpha = 0{,}05$ не оказывает существенного влияния наличие/отсутствие телефона в квартире. Параметр при $d_1^{(1)}$ означает, что при одной и той же жилой площади квартиры цена ее в панельных домах в среднем на 1110 долл. США выше, чем в «хрущевках»; в кирпичных цена выше в среднем на 1060 долл. США.

3.3 Проверка регрессионной однородности двух групп наблюдений (критерий Г.Чоу)

Следует отметить, что с введением фиктивных переменных увеличивается число переменных, что приводит к снижению точности получаемых оценок коэффициентов. Поэтому необходимо обоснованно вводить фиктивные переменные в регрессионную модель. Для этого рекомендуется проверить гипотезу о регрессионной неоднородности статистической совокупности.

Допустим, что наша выборка состоит из двух однородных групп, одна из которых объёмом n_1, а другая - n_2. Если n_1, n_2 значительно больше, чем $k+1$, то для каждой из таких подвыборок можно построить регрессионную модель. Если окажется, что оценки коэффициентов для одной однородной группы входят в доверительные интервалы для другой группы, то делается вывод о регрессионной однородности выборочной совокупности и переходят к построению оценок на основе объединённой выборки объёмом $n_1 + n_2$.

Рассмотрим другую ситуацию, когда объём одной из подвыборок, например, второй, меньше или равен $k+1$. В этом случае вторая подвыборка не позволяет построить уравнение регрессии для однородной группы и для проверки гипотезы об однородности выборочной совокупности используется критерий Чоу [1].

$$\begin{aligned} H_0 &: \beta^{(1)} = \beta^{(2)}; \sigma_\varepsilon^2(1) = \sigma_\varepsilon^2(2) \\ H_1 &: \beta^{(1)} \neq \beta^{(2)}; \sigma_\varepsilon^2(1) \neq \sigma_\varepsilon^2(2) \end{aligned} \tag{3.2}$$

Для проверки гипотезы используется статистика Чоу, которая в условиях справедливости нулевой гипотезы распределена по закону Фишера-Снедекора $(\nu_1 = n_2, \nu_2 = n_1 - k - 1)$:

$$\gamma_{n,n_1} = \frac{\left(e^T e - e^{(1)T} e^{(1)}\right)/ n_2}{e^{(1)T} e^{(1)} / (n_1 - k - 1)}, \tag{3.3}$$

где e - вектор регрессионных остатков, оцененных по всей выборке;

$e^{(1)}$ - вектор регрессионных остатков, оцененных по первой подвыборке.

Если n_2 достаточно велико, то более предпочтительно использовать статистику [1]:

$$\gamma_{n_1,n_2} = \frac{\left(e^T e - e^{(1)T} e^{(1)} - e^{(2)T} e^{(2)}\right) / (k+1)}{\left(e^{(1)T} e^{(1)} + e^{(2)T} e^{(2)}\right) / \left(n_1 + n_2 - 2k - 2\right)}, \tag{3.4}$$

где e - вектор регрессионных остатков, оцененных по всей выборке;

$e^{(1)}$- вектор регрессионных остатков, оцененных по первой подвыборке;

$e^{(2)}$ - вектор регрессионных остатков, оцененных по второй подвыборке.

3.4 Тестовые задания для самоконтроля

1) Пусть величина расходов “y” на оплату электроэнергии зависит от периода взимания платы – отопительный сезон или нет. Тогда линейная модель множественной регрессии:

а) $y_i = \beta_0 + \beta_1 d_{1i} + \beta_2 d_{2i} + \varepsilon_i, \quad i = \overline{1,n}$;

б) $y_i = \beta_0 + \beta_1 d_{1i} + \beta_2 d_{2i}, \quad i = \overline{1,n}$;

в) $y_i = \beta_0 + \varepsilon_i, \quad i = \overline{1,n}$;

г) $y_i = \beta_0 + \beta_1 d_{1i} + \varepsilon_i, \quad i = \overline{1,n}$;

где:

$$d_{1i} = \begin{cases} 1, & \textit{наблюдение сделано в отопительный сезон} \\ 0, & \textit{в остальных случаях} \end{cases}$$

$$d_{2i} = \begin{cases} 1, & \textit{наблюдение сделано в неотопительный сезон} \\ 0, & \textit{в остальных случаях} \end{cases}$$

ε_i - регрессионные остатки.

2) Время “Т” доставки груза от базы до торговой точки определяется в первую очередь расстоянием “S” и, предположительно, временем доставки (час пик или нет). Тогда линейная модель множественной регрессии:

а) $T_i = \beta_0 + \beta_1 S_i + \beta_2 d_{1i} + \varepsilon_i, \quad i = \overline{1,n}$;

б) $T_i = \beta_0 + \beta_1 S_i + \beta_2 d_{1i} + \beta_3 d_{2i} + \varepsilon_i, \quad i = \overline{1,n}$;

в) $T_i = \beta_0 + \beta_1 S_i + \varepsilon_i, \quad i = \overline{1,n}$;

г) $T_i = \beta_0 + \beta_1 S_i + \beta_2 d_{1i}, \quad i = \overline{1,n}$;

где: $d_{1i} = \begin{cases} 1, & \textit{наблюдение проведено в час пик} \\ 0, & \textit{в остальных случаях} \end{cases}$

$$d_{2i} = \begin{cases} 1, & \textit{наблюдение проведено не в час пик} \\ 0, & \textit{в остальных случаях} \end{cases}$$

ε_i - регрессионные остатки.

3) Если в ходе сбора исходных статистических данных имеет место косвенное воздействие (во времени и/или пространстве) качественных факторов, то речь идет о:

а) нелинейных моделях;

б) линейных регрессионных моделях с постоянной структурой;

в) линейных регрессионных моделях с переменной структурой;

г) системе одновременных регрессионных уравнений.

4) Объем “y” (в денежном исчислении) реализации продуктов питания торговой точкой зависит от периода реализации (будние дни, выходные или праздничные). Тогда линейная модель множественной регрессии:

а) $y_i = \beta_0 + \beta_1 d_{1i} + \beta_2 d_{2i} + \varepsilon_i, \quad i = \overline{1,n}$;

б) $y_i = \beta_0 + \varepsilon_i, \quad i = \overline{1,n}$;

в) $y_i = \beta_0 + \beta_1 d_{1i} + \varepsilon_i, \quad i = \overline{1,n}$;

г) $y_i = \beta_0 + \beta_1 d_{1i} + \beta_2 d_{2i}, \quad i = \overline{1,n}$;

где:

$$d_{1i} = \begin{cases} 1, & \textit{наблюдение проведено в будний день} \\ 0, & \textit{наблюдение проведено в выходной или праздничный день} \end{cases}$$

$$d_{1i} = \begin{cases} 1, & \textit{наблюдение проведено в выходной или праздничный день} \\ 0, & \textit{наблюдение проведено в будний день} \end{cases}$$

ε_i - регрессионные остатки.

5) Статистические данные называются неоднородными в регрессионном смысле, если они:

а) зарегистрированы при одних и тех же условиях;

б) зарегистрированы не своевременно;

в) не наблюдаемы;

г) зарегистрированы при различных условиях.

6) Изучается зависимость средней заработной платы работника в зависимости от стажа (x_1), производительности труда (x_2) и уровня образования (начальное (d_1), среднее (d_2), высшее).

Оценка уравнения регрессии имеет следующий вид:

$$\hat{y} = 2{,}2 + 0{,}5x_1 + 0{,}4x_2 - 0{,}1d_1 - 0{,}02d_2$$

где:

$$d_1 = \begin{cases} 1, & \textit{работник с начальным уровнем образования} \\ 0, & \textit{иначе} \end{cases}$$

$$d_1 = \begin{cases} 1, & \textit{работник со средним уровнем образования} \\ 0, & \textit{иначе} \end{cases}$$

Указать величину различия в оплате труда для работников с начальным и высшим уровнем образования при одних и тех же значениях производительности труда и стажа:

а) 0,5;

б) 0,4;

в) 0,1;

г) 0,02.

7) Если качественная переменная имеет p градаций, то для отражения ее влияния на структуру регрессионной связи необходимо ввести:

а) 1 фиктивную переменную;

б) $p-1$ фиктивных переменных;

в) p фиктивных переменных;

г) $p+1$ фиктивных переменных.

8) Изучается зависимость средней заработной платы работника в зависимости от стажа (x_1), производительности труда (x_2) и уровня образования (начальный (d_1), средний (d_2), высший).

Оценка уравнения регрессии имеет следующий вид:

$$\hat{y} = 2{,}2 + 0{,}5x_1 + 0{,}4x_2 - 0{,}1d_1 - 0{,}02d_2$$

где:

$$d_1 = \begin{cases} 1, & \textit{работник с начальным уровнем образования} \\ 0, & \textit{иначе} \end{cases}$$

$$d_1 = \begin{cases} 1, & \textit{работник со средним уровнем образования} \\ 0, & \textit{иначе} \end{cases}$$

Указать величину различия в оплате труда для работников со средним и высшим уровнем образования при одних и тех же значениях производительности труда и стажа:

а) 0,5;

б) 0,4;

в) 0,1;

г) 0,02.

9) Рассматривается регрессионная модель:

$$\hat{y} = 56{,}6 - 21{,}6d_1 - 10{,}1d_2,$$

где:

y - процент рабочих ручного труда в общей численности рабочих;

d - уровень автоматизации производства;

$$d_1 = \begin{cases} 1, & \text{для предприятий с высоким уровнем автоматизации производства} \\ 0, & \text{для остальных предприятий} \end{cases}$$

$$d_1 = \begin{cases} 1, & \text{для предприятий со средним уровнем автоматизации производства} \\ 0, & \text{для остальных предприятий} \end{cases}$$

Чему равен средний процент рабочих ручного труда на предприятиях с низким уровнем автоматизации производства:

а) 0;
б) 10,1;
в) 56,6;
г) 21,6.

10) При исследовании зависимости цены однокомнатной квартиры от ее полезной площади и типа дома («хрущевка», панельный, кирпичный), была получена следующая модель: $\hat{y} = 320 + 500x + 2200d_1 + 1600d_2$,

где:

$$d_1 = \begin{cases} 1, & \text{для панельного дома} \\ 0, & \text{для остальных типов домов} \end{cases}$$

$$d_2 = \begin{cases} 1, & \text{для кирпичного дома} \\ 0, & \text{для остальных типов домов} \end{cases}$$

На сколько единиц в среднем цена квартиры в кирпичных домах выше, чем в «хрущевках» при неизменной величине полезной площади:

а) 320;
б) 500;
в) 2200;
г)1600.

11) Для учета влияния качественного признака на структуру модели фиктивные переменные вводится:

а) аддитивно-линейно;
б) мультипликативно-линейно;
в) в зависимости от характера связи между переменными;
г) нелинейно.

12) Пусть объем потребления овощей “y” определяется доходом “x” и, предположительно, может зависеть от времени года. Тогда линейная модель множественной регрессии имеет вид:

а) $y_i = \beta_0 + \beta_1 x_i + \beta_2 d_{1i} + \beta_3 d_{2i} + \beta_4 d_{3i} + \beta_5 d_{4i} + \varepsilon_i, \quad i = \overline{1, n}$;

б) $y_i = \beta_0 + \beta_1 x_i + \beta_2 d_{1i} + \beta_3 d_{2i} + \beta_4 d_{3i} + \varepsilon_i, \quad i = \overline{1, n}$;

в) $y_i = \beta_0 + \beta_1 x_i + \beta_2 D_{1i} + \beta_3 D_{2i} + \varepsilon_i, \quad i = \overline{1, n}$;

г) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = \overline{1, n}$, где:

$$d_{1i} = \begin{cases} 1, & \text{наблюдение сделано зимой} \\ 0, & \text{в остальных случаях} \end{cases}; \quad d_{2i} = \begin{cases} 1, & \text{наблюдение сделано весной} \\ 0, & \text{в остальных случаях} \end{cases}$$

$$d_{3i} = \begin{cases} 1, & \text{наблюдение сделано летом} \\ 0, & \text{в остальных случаях} \end{cases} \quad d_{4i} = \begin{cases} 1, & \text{наблюдение сделано осенью} \\ 0, & \text{в остальных случаях} \end{cases}$$

$$D_{1i} = \begin{cases} 1, & \text{наблюдение сделано летом или осенью} \\ 0, & \text{в остальных случаях} \end{cases}$$

$$D_{2i} = \begin{cases} 1, & \text{наблюдение сделано зимой или весной} \\ 0, & \text{в остальных случаях} \end{cases}$$

ε_i - регрессионные остатки.

3.5 Индивидуальное задание №4[7]

По данным Приложения Б:

- выдвинуть и обосновать предположение о сопутствующих качественных факторах, числе уровней каждого, указать число фиктивных переменных и охарактеризовать каждую из них;

- записать линейную модель регрессии с переменной структурой и её матрицу “объект - свойства”;

- исследовать имеющиеся статистические данные на неоднородность с помощью критерия Чоу.

- оценить параметры регрессионной модели с переменной структурой и провести её анализ.

3.6 Порядок выполнения индивидуального задания № 4

Пусть по результатам оценивания, получена оценка уравнения регрессии, описывающая зависимость между ценой квартиры, тыс. долл. (у) и общей площадью квартиры, м^2 (x_2): $\hat{y} = \underset{(0,0098)}{0{,}491} x_2$. На основании отчета (рисунок 3.3) можно сделать вывод, что модель значима; существенное влияние на результативный признак – цена квартиры, оказывает объясняющая переменная – общая площадь квартиры.

Regression Summary for Dependent Variable: y (фиктив)
R= ,99095724 R?= ,98199626 Adjusted R?= ,98160487
F(1,46)=2509,0 p<0,0000 Std.Error of estimate: 4,7448

N=47	Beta	Std.Err. of Beta	B	Std.Err. of B	t(46)	p-level
x2	0,990957	0,019783	0,491029	0,009803	50,09016	0,000000

Рисунок 3.3 – Результаты множественной регрессии

[7] Индивидуальные задания и пример их выполнения разработаны авторами совместно с преподавателем кафедры математических методов и моделей в экономике Жемчужниковой Ю.А.

Очевидно, что на цену квартиры может оказывать влияние и тип дома (панельный, кирпичный). Проверим эту гипотезу. Так как качественный признак имеет две градации, то введем одну фиктивную переменную:

$$d_1 = \begin{cases} 1, \textit{кирпичный} \\ 0, \textit{иначе} \end{cases}$$

Таким образом, модель регрессии будем искать в виде:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 d_{i1} + \delta_i\,. \tag{3.10}$$

Тогда матрица “объект-свойства” будет иметь вид, изображенный на рисунке 3.4, где качественные свойства не формализованы. Окно с частью данных изображено на рисунке 3.4.

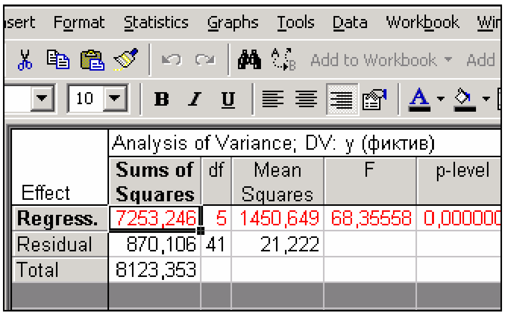
	1 y	2 x1	3 x2	4 x3	5 x4	6 x5	7 d1
1	15,9	1	39	20	8,2	0	Панельный
2	27	3	68,4	40,5	10,7	0	Панельный
3	13,5	1	34,8	16	10,7	12	Панельный
4	15,1	1	39	20	8,5	12	Панельный
5	21,1	2	54,7	28	10,7	12	Панельный
6	28,7	3	74,7	46,3	10,7	12	Панельный
7	27,2	3	71,7	45,9	10,7	0	Панельный
8	28,3	3	74,5	47,5	10,4	0	Панельный
9	52,3	4	137,7	87,2	14,6	0	Панельный
10	22	1	40	17,7	11	8	Кирпичный
11	28	2	53	31,1	10	8	Кирпичный
12	45	3	86	48,7	14	8	Кирпичный
13	51	4	98	65,8	13	8	Кирпичный
14	34,4	2	62,6	21,4	11	0	Кирпичный
15	24,7	1	45,3	20,6	10,4	8	Кирпичный
16	30,8	2	56,4	29,7	9,4	8	Кирпичный
17	15,9	1	37	17,8	8,3	0	Панельный
18	29	3	67,5	43,5	8,3	0	Панельный

Рисунок 3.4– Таблица данных

Но прежде чем вводить фиктивные переменные необходимо проверить выборочную совокупность на регрессионную однородность, применяя критерий Чоу.

Разделим всю совокупность на две подвыборки. В результате получим, что в выборку, где дома панельные вошло 15 объектов, а где кирпичные – 32 объекта. Так как объем подвыборок достаточно велик, то проверка гипотезы об однородности вида (3.2) осуществляется с помощью критерия (3.4)

Построив уравнение по объединенной выборки, получили следующие результаты. Остатки исследуются в специальном окне **Advanced** (рисунок 1.10). В нижней части окна результатов регрессионного анализа нажмем кнопку **ANOVA,** в результате появится окно:



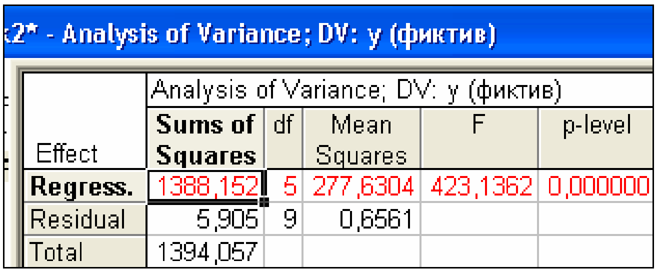
Effect	Analysis of Variance; DV: y (фиктив) Sums of Squares	df	Mean Squares	F	p-level
Regress.	7253,246	5	1450,649	68,35558	0,000000
Residual	870,106	41	21,222		
Total	8123,353				

Рисунок 3.5– Результаты дисперсионного анализа

Значение суммы квадратов остатков находится на пересечении строки **Residual** (остаточная) и столбца **Sums of Squares** (сумма квадратов).

Таким образом, $e^T \cdot e$=870,106

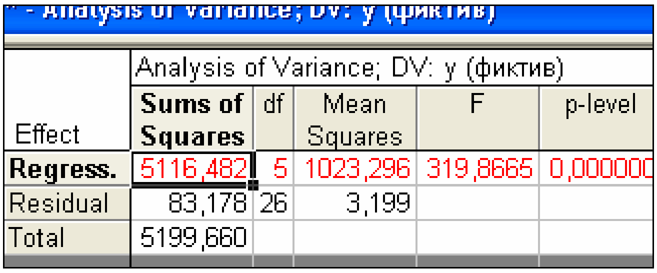
Аналогично оцениваем регрессионные остатки по первой подвыборки (Панельные дома) $n_1 = 15$, в результате получим $e^{(1)^T} \cdot e^{(1)} = 5{,}906$ (рисунок 3.6):



Effect	Analysis of Variance; DV: y (фиктив) Sums of Squares	df	Mean Squares	F	p-level
Regress.	1388,152	5	277,6304	423,1362	0,000000
Residual	5,905	9	0,6561		
Total	1394,057				

Рисунок 3.6 – Результаты дисперсионного анализа для 15 объектов

Для кирпичных домов $n_2 = 32$, получим следующие результаты $e^{(2)^T} \cdot e^{(2)} = 83{,}178$ (рисунок 3.7)



Effect	Analysis of Variance; DV: y (фиктив) Sums of Squares	df	Mean Squares	F	p-level
Regress.	5116,482	5	1023,296	319,8665	0,000000
Residual	83,178	26	3,199		
Total	5199,660				

Рисунок 3.7 – Результаты дисперсионного анализа для 32объектов

Подставим полученные результаты в формулу (3.4):

$$\gamma_{n_1 n_2} = \frac{(e^T \cdot e - e^{(1)^T} \cdot e^{(1)} - e^{(2)^T} \cdot e^{(2)}) / \kappa + 1}{\dfrac{e^{(1)^T} \cdot e^{(1)} + e^{(2)^T} \cdot e^{(2)}}{n_1 + n_2 - 2k - 2}} = \frac{(870{,}106 - 5{,}906 - 83{,}178)/5 + 1}{\dfrac{5{,}906 + 83{,}178}{15 + 32 - 2*5 - 2}} = 51{,}14$$

На уровне значимости 0,05 и числу степеней свободы $\nu_1 = 6$ и $\nu_2 = 35$, найдем $\gamma_{крит}$ =2,39 по таблице Фишера-Снедекора. Так как $\gamma_{крит} < \gamma_{наб}$, то гипотеза H_0 отвергается, следовательно, подвыборки неоднородны. Переходим к построению модели с фиктивными переменными.

В системе Statistica для удобной работы с переменными, принимающими текстовые значения, реализован так называемый механизм "двойной записи". Согласно этому соглашению, каждому текстовому значению переменной ставится в соответствие некоторое число. Таким образом, устанавливается соответствие вида *Число=Текстовое значение.* При работе с данными всегда можно переключиться между текстовой и числовой формой просмотра исходных данных.

Можно легко переключаться между различными способами отображения данных в электронной таблице. Для этого нужно только выбрать в меню **View/Display Text Labels:**

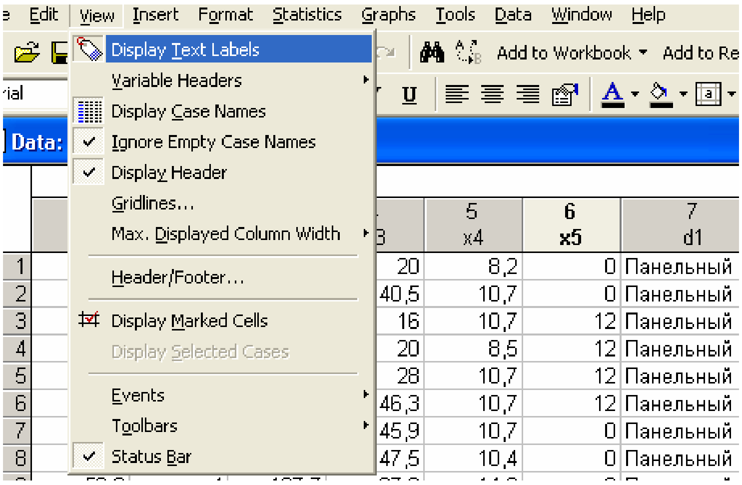
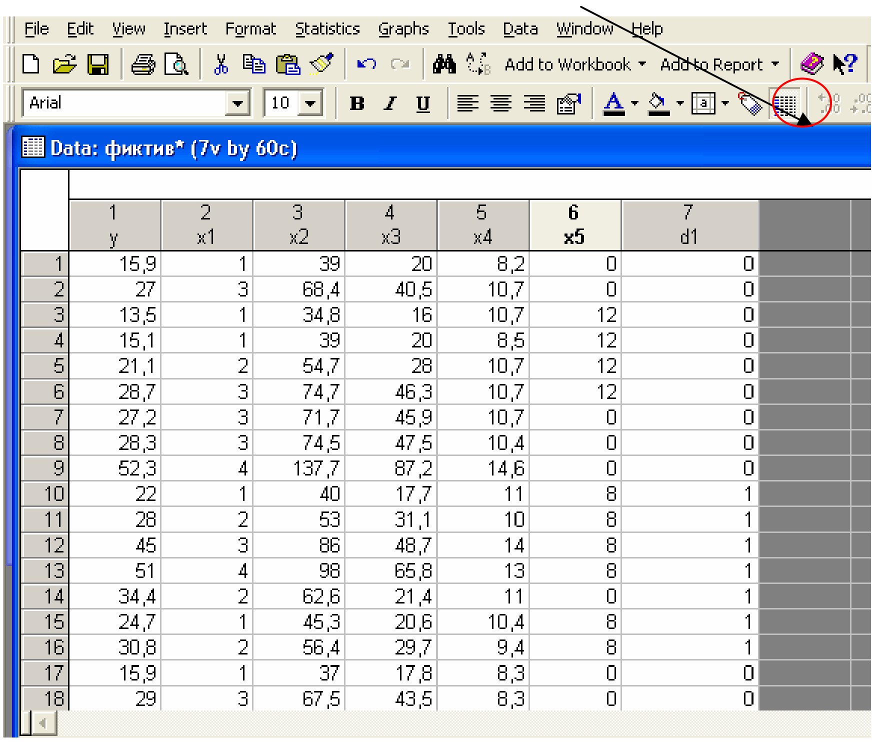


Рисунок 3.8

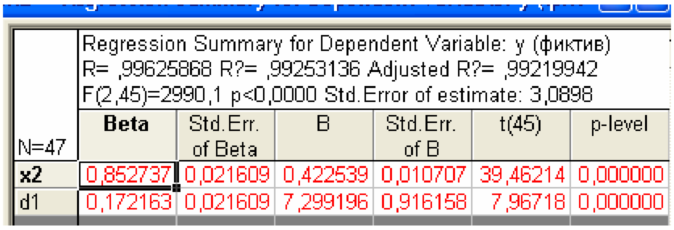
Или на панели можно перейти от качественных к бинарным переменным с помощью кнопки (рисунок 3.9).



	1 y	2 x1	3 x2	4 x3	5 x4	6 x5	7 d1
1	15,9	1	39	20	8,2	0	0
2	27	3	68,4	40,5	10,7	0	0
3	13,5	1	34,8	16	10,7	12	0
4	15,1	1	39	20	8,5	12	0
5	21,1	2	54,7	28	10,7	12	0
6	28,7	3	74,7	46,3	10,7	12	0
7	27,2	3	71,7	45,9	10,7	0	0
8	28,3	3	74,5	47,5	10,4	0	0
9	52,3	4	137,7	87,2	14,6	0	0
10	22	1	40	17,7	11	8	1
11	28	2	53	31,1	10	8	1
12	45	3	86	48,7	14	8	1
13	51	4	98	65,8	13	8	1
14	34,4	2	62,6	21,4	11	0	1
15	24,7	1	45,3	20,6	10,4	8	1
16	30,8	2	56,4	29,7	9,4	8	1
17	15,9	1	37	17,8	8,3	0	0
18	29	3	67,5	43,5	8,3	0	0

Рисунок 3.9

Построим уравнение множественной регрессии результативного признака Y на количественную переменную x_1 и качественную переменную d_1 методом пошаговой регрессии (рисунок 3.10).



Regression Summary for Dependent Variable: y (фиктив)
R= ,99625868 R?= ,99253136 Adjusted R?= ,99219942
F(2,45)=2990,1 p<0,0000 Std.Error of estimate: 3,0898

N=47	Beta	Std.Err. of Beta	B	Std.Err. of B	t(45)	p-level
x2	0,852737	0,021609	0,422539	0,010707	39,46214	0,000000
d1	0,172163	0,021609	7,299196	0,916158	7,96718	0,000000

Рисунок 3.10 – Результаты множественной регрессии

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что распределение регрессионных остатков не отличаются от нормального

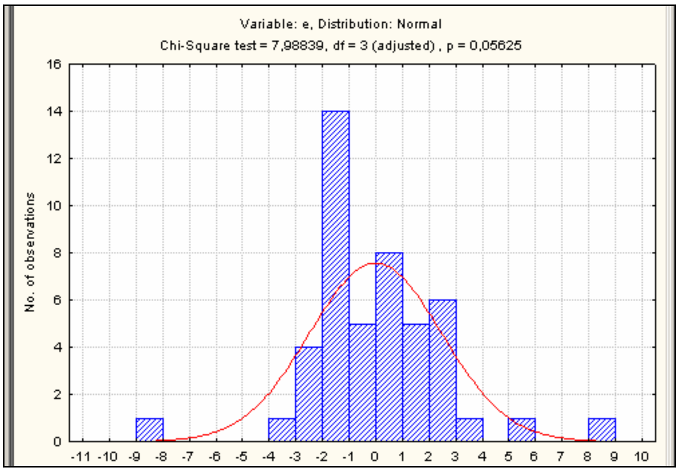


Рисунок 3.11 – Гистограмма регрессионных остатков

Следовательно, можно сделать выводы, что модель $\hat{y} = 0{,}42x_2 + 7{,}3d_1$ значима; существенное влияние на цену квартиры, оказывает количественная переменная – общая площадь квартиры и качественный признак – тип дома.

3.7 Вопросы для самоконтроля

1 В чем состоит проблема построения модели по неоднородным данным?

2 Опишите возможные подходы к построению линейных регрессионных моделей по неоднородным данным.

3 Правило введения манекенов.

4 Критерий Чоу. Назначение и техника использования.

5 Содержательная интерпретация значимых (незначимых) фиктивных переменных.

4 Нелинейные модели регрессии

4.1 Подходы к оцениванию параметров нелинейных моделей регрессии

Линейные модели регрессии могут служить неплохой аппроксимацией зависимостей между показателями социально-экономического процесса лишь в некоторой области изменения объясняющих переменных. Реальные связи безусловно имеют нелинейный характер. Примерами такого рода регрессионных моделей являются **производственные функции**, описывающие зависимости между объемом произведенной продукции и основными факторами производства – трудом, капиталом; **функции спроса** – зависимости между спросом на какой-либо товар и доходом и ценами на этот и другие товары.

Для оценки коэффициентов нелинейной модели регрессии используются следующие подходы [1]:

1) Подбираются такие преобразования к исходным переменным $y, x_1, ..., x_k$, которые позволили бы представить искомую зависимость в виде линейного соотношения между преобразованными переменными. Такие преобразования называются **процедурами линеаризации модели**.

2) Если не удается линеаризовать модель, то приходится исследовать нелинейную регрессионную зависимость в терминах исходных переменных $y_i = f(X_i; \beta) + \varepsilon_i$, то оценка параметров может быть осуществлена методом наименьших квадратов, то есть решением оптимизационной задачей:

$$\sum_{i=1}^{n} (y_i - f(X_i; \beta))^2 \underset{\beta}{\rightarrow} \min . \tag{4.1}$$

Наиболее полноценным является исследование нелинейных моделей регрессии на основе первого подхода, которому далее и уделяется основное внимание.

Приведем некоторые распространенные типы нелинейных моделей, допускающих непосредственную линеаризацию (рисунок 4.1) [1, 4]. Зависимости, нелинейные по оцениваемым параметрам подразделяются на внутренне линейные и внутренне нелинейные. Нелинейная модель называется внутренне линейной ($\widetilde{y} = \beta_0 x^{\beta_1}$), если с помощью линеаризации она может быть приведена к линейному виду. Нелинейная модель называется внутренне нелинейной ($\widetilde{y} = \beta_0 + \beta_1 x^{\beta_2}$), если она не может быть сведена к линейной функции.

Нелинейная регрессия

Нелинейные регрессии по объясняющим переменным

- $\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$;
- $\tilde{y} = \beta_0 + \beta_1 \dfrac{1}{x}$;

Нелинейные регрессии по оцениваемым параметрам

нелинейная модель внутренне линейна

- $\tilde{y} = \beta_0 \cdot x^{\beta_1}$;
- $\tilde{y} = e^{\beta_1 + \beta_2 \cdot x}$;
- $\tilde{y} = \beta_0 \cdot \beta_1^{\,x}$

нелинейная модель внутренне нелинейна

- $\tilde{y} = \beta_0 + \beta_1 x^{\beta_2}$;
- $\tilde{y} = \beta_0 (1 - \dfrac{1}{1 - x^{\beta_1}})$

Рисунок 4.1 – Виды нелинейных моделей регрессии

4.2 Некоторые виды нелинейных зависимостей, поддающихся непосредственной линеаризации

Зависимость гиперболического типа:

$$\text{а)} \quad \begin{aligned} &\tilde{y} = \beta_0 + \beta_1 \frac{1}{x}; \\ &y_i = \beta_0 + \beta_1 \frac{1}{x_i} + \varepsilon_i, \end{aligned} \qquad (4.2)$$

График функции представлен на рисунке 4.2

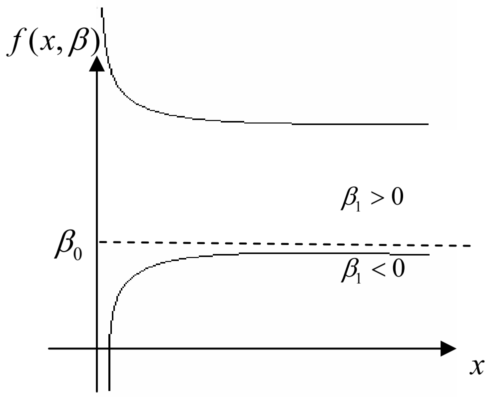


Рисунок 4.2 – График функции $\tilde{y} = \beta_0 + \beta_1 \dfrac{1}{x}$

С помощью преобразования объясняющей переменной $x^* = \frac{1}{x}$ эта зависимость приводится к линейному виду $y_i = \beta_0 + \beta_1 x_i^* + \varepsilon_i$, соответственно при вычислении МНК-оценок матрица X будет иметь вид: $X^* = \begin{bmatrix} 1 & \frac{1}{x_1} \\ \ldots & \ldots \\ 1 & \frac{1}{x_n} \end{bmatrix}$

б) $$\begin{aligned} \tilde{y} &= \frac{1}{\beta_0 + \beta_1 x}; \\ y_i &= \frac{1}{\beta_0 + \beta_1 x_i + \varepsilon_i}, \end{aligned} \tag{4.3}$$

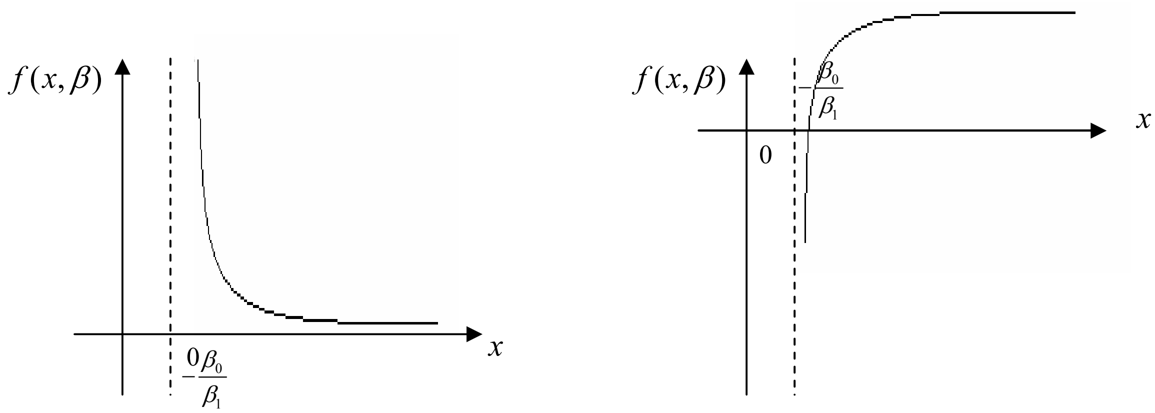


Рисунок 4.3 – График функции $\tilde{y} = \frac{1}{\beta_0 + \beta_1 x}$

С помощью преобразования результативной признака $y^* = \frac{1}{y}$ эта зависимость приводится к линейному виду $y^*{}_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, соответственно при вычислении МНК-оценок $Y^* = \begin{bmatrix} \frac{1}{y_1} \\ \ldots \\ \frac{1}{y_n} \end{bmatrix}$.

Замечание. Функции, изображенные на рисунке 4.2 (случай $\beta_1 < 0$) используются в определенных ситуациях при построении так называемых кривых Энгеля, которые описывают зависимость спроса на определенный вид товаров или услуг от уровня доходов потребителей; функции, изображенные на рисунке 4.2 (случай $\beta_1 > 0$) и рисунке 4.3 – при изучении спроса на товар в зависимости от цены.

Показательная (экспоненциальная) зависимость

Достаточно широкий класс экономических показателей характеризуется приблизительно постоянным темпом относительного прироста во времени. Этому соответствует показательная форма зависимости:

в) $\tilde{y} = \beta_0 \cdot e^{\beta_1 \cdot x}$;
$y_i = \beta_0 \cdot e^{\beta_1 \cdot x_i + \varepsilon_i}$ (4.4)

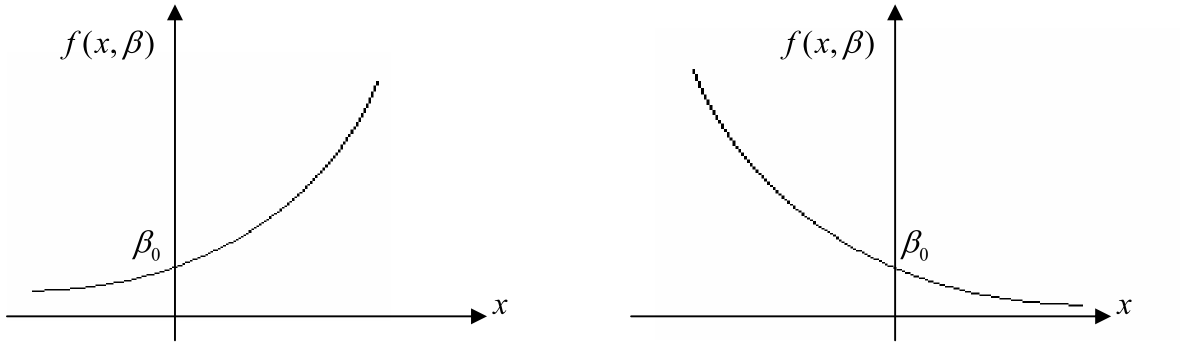


Рисунок 4.4 – График функции $\tilde{y} = \beta_0 \cdot e^{\beta_1 \cdot x}$

Переход к новой переменной $y^* = \ln y$, $Y^* = \begin{bmatrix} \ln(y_1) \\ \ln(y_2) \\ ... \\ \ln(y_n) \end{bmatrix}$ позволяет свести исследуемому зависимость к линейному виду: $y_i^* = \beta_0^* + \beta_1 x_i + \varepsilon_i$, где $\beta_0^* = \ln \beta_0$.

г) $\tilde{y} = \beta_0 e^{\frac{\beta_1}{x}}$;
$y_i = \beta_0 e^{\frac{\beta_1}{x_i} + \varepsilon_i}$ (4.5)

Линеаризация искомой зависимости достигается с помощью следующих преобразований переменных: $y^* = \ln y$, $Y^* = \begin{bmatrix} \ln(y_1) \\ \ln(y_2) \\ ... \\ \ln(y_n) \end{bmatrix}$; $x^* = \frac{1}{x}$,

$$X^* = \begin{bmatrix} 1 & \frac{1}{x_1} \\ ... & ... \\ 1 & \frac{1}{x_n} \end{bmatrix}, \ \beta_0^* = \ln \beta_0$$

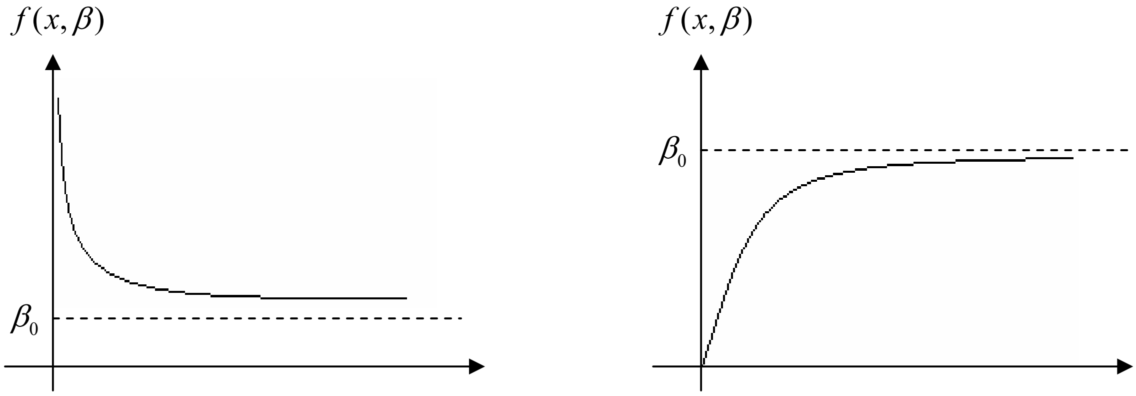


Рисунок 4.5 – График функции $\widetilde{y} = \beta_0 e^{\frac{\beta_1}{x}}$.

Зависимость степенного типа [1]

$$д)\ \begin{matrix} \widetilde{y} = \beta_0 x^{\beta_1} \\ y_i = \beta_0 \cdot x_i^{\beta_1} \cdot \varepsilon_i \end{matrix} \qquad (4.6)$$

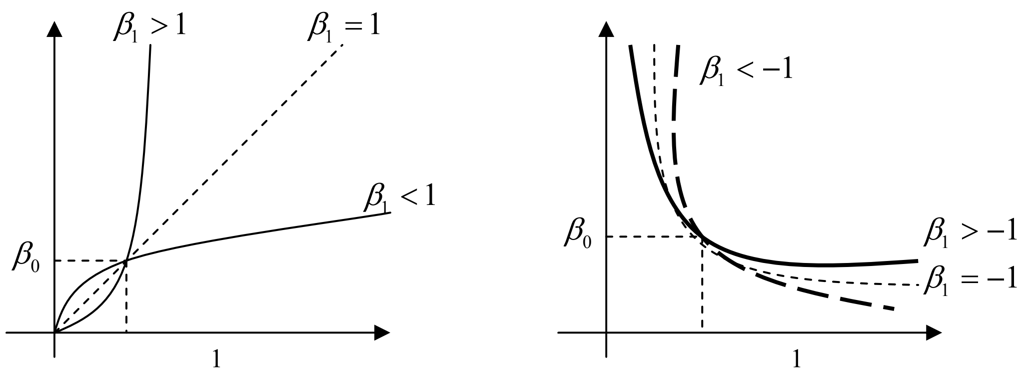


Рисунок 4.6 – График $\widetilde{y} = \beta_0 x^{\beta_1}$

Линеаризация искомой зависимости достигается с помощью следующих преобразований переменных: $y^* = \ln y$, $Y^* = \begin{bmatrix} \ln(y_1) \\ \ln(y_2) \\ ... \\ \ln(y_n) \end{bmatrix}$; $X^* = \ln(X)$,

$X^* = \begin{bmatrix} 1 & \ln(x_1) \\ ... & ... \\ 1 & \ln(x_n) \end{bmatrix}$, где $\beta_0^* = \ln \beta_0$.

Важную роль зависимости степенного типа играют в задачах построения и анализа производственных функций, функций спроса. При анализе степенных регрессионных зависимостей содержательную интерпретацию получает коэффициент β_1 как коэффициент эластичности.

Аналогично процедура линеаризации проводится для множественных регрессионных уравнений. Например, для степенной функции $\tilde{y} = \beta_0 x_1^{\beta_1} x_2^{\beta_2} ... x_k^{\beta_k}$ необходимо осуществить преобразования переменных:

$$y^* = \ln y, \qquad Y^* = \begin{bmatrix} \ln(y_1) \\ \ln(y_2) \\ ... \\ \ln(y_n) \end{bmatrix}; \qquad x_1^* = \ln(x_1), \qquad x_2^* = \ln(x_2), \qquad \ldots, \qquad x_k^* = \ln(x_k),$$

$$X^* = \begin{bmatrix} 1 & \ln(x_{11}).... & \ln(x_{1k}) \\ ... & ... & ... \\ 1 & \ln(x_{n1}).... & \ln(x_{nk}) \end{bmatrix}, \text{ где } \beta_0^* = \ln \beta_0 .$$

В итоге модель приводится к виду: $y_i^* = \beta_0^* + \sum_{l=1}^{k} \beta_l x_{il}^* + \varepsilon_i$.

4.3 Тестовые задания для самоконтроля

1) Указать, с помощью какого преобразования исходных переменных зависимость $\tilde{y} = \beta_0 e^{\frac{\beta_1}{x}}$ сводится к линейной:

а) $y^* = \ln y$, $x^* = \frac{1}{x}$ (где $\beta_0^* = \ln \beta_0$);

б) $x^* = \frac{1}{x}$, $y^* = \frac{1}{y}$;

в) $y^* = \frac{1}{y}$, $x^* = e^{-x}$;

г) $y^* = \ln y$ (где $\beta_0^* = \ln \beta_0$).

2) Кривая Филипса характеризует:

а) зависимость между объемом произведенной продукции и основными факторами производства;

б) зависимость между спросом на товар (услугу) и уровнем доходов потребителя;

в) нелинейное соотношение между нормой безработицы и процентом прироста заработной платы;

г) зависимость между спросом на товар (услугу) и его (ее) ценой.

3) Указать регрессионную модель, которая не поддается непосредственной линеаризации:

а) $y_i = \beta_0 + \beta_1 \frac{1}{x_i} + \varepsilon_i$;

б) $y_i = \beta_0 \cdot \left(1 - \frac{1}{1 - x_i^{\beta_1}}\right) + \varepsilon_i$;

в) $y_i = \beta_0 \cdot (x_i)^{\beta_1} \cdot \varepsilon_i$;

г) $y_i = \frac{x_i}{\beta_0 x_i + \beta_1 + x_i \varepsilon_i}$.

4) При анализе степенных регрессионных зависимостей $y_i = \beta_0 \cdot (x_i)^{\beta_1} \cdot \varepsilon_i$ содержательную интерпретацию получает параметр β_1 как:

а) множественный коэффициент корреляции;

б) частный коэффициент корреляции;

в) коэффициент эластичности;

г) коэффициент детерминации.

5) Требуется оценить функцию регрессии $\tilde{y} = \beta_0 e^{\frac{\beta_1}{x}}$ на основе информации: $Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}^T$ $X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T$. Указать вектор-столбец Y^* и матрицу X^*, участвующих в формулах МНК-оценок, после линеаризации зависимости:

а) $Y^* = \begin{pmatrix} \frac{1}{y_1} & \frac{1}{y_2} & \dots & \frac{1}{y_n} \end{pmatrix}^T$; $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \ln x_1 & \ln x_2 & \dots & \ln x_n \end{pmatrix}^T$;

б) $Y^* = \begin{pmatrix} \ln y_1 & \ln y_2 & \dots & \ln y_n \end{pmatrix}^T$; $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \ln x_1 & \ln x_2 & \dots & \ln x_n \end{pmatrix}^T$;

в) $Y^* = \begin{pmatrix} \frac{1}{y_1} & \frac{1}{y_2} & \dots & \frac{1}{y_n} \end{pmatrix}^T$; $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{-x_1} & e^{-x_2} & \dots & e^{-x_n} \end{pmatrix}^T$;

г) $Y^* = \begin{pmatrix} \ln y_1 & \ln y_2 & \dots & \ln y_n \end{pmatrix}^T$; $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \frac{1}{x_1} & \frac{1}{x_2} & \dots & \frac{1}{x_n} \end{pmatrix}^T$.

6) Указать регрессионную модель, которая не поддается непосредственной линеаризации:

а) $y_i = \beta_0 \cdot (x_i)^{\beta_1} \cdot \varepsilon_i$;

б) $y_i = \frac{x_i}{\beta_0 x_i + \beta_1 + x_i \varepsilon_i}$;

в) $y_i = \beta_0 \cdot (x_i)^{\beta_1} + \varepsilon_i$;

г) $y_i = \beta_0 + \beta_1 \frac{1}{x_i} + \varepsilon_i$.

7) Указать, с помощью какого преобразования исходных переменных зависимость $\tilde{y} = \beta_0 e^{\beta_1 x}$ сводится к линейной:

а) $x^* = \frac{1}{x}$, $y^* = \frac{1}{y}$;

б) $y^* = \ln y$ (где $\beta_0^* = \ln \beta_0$);

в) $y^* = \dfrac{1}{y}$;

г) $y^* = \ln y$, $x^* = \dfrac{1}{x}$ (где $\beta_0^* = \ln \beta_0$).

8) Указать, с помощью какого преобразования исходных переменных зависимость $\tilde{y} = \beta_0 + \beta_1 \dfrac{1}{x}$ сводится к линейной:

а) $y^* = \dfrac{1}{y}$;

б) $y^* = \ln y$ (где $\beta_0^* = \ln \beta_0$);

в) $x^* = \dfrac{1}{x}$;

г) $x^* = \dfrac{1}{x}$, $y^* = \dfrac{1}{y}$.

9) Указать, с помощью какого преобразования исходных переменных зависимость $\tilde{y} = \dfrac{1}{\beta_0 + \beta_1 x}$ сводится к линейной:

а) $y^* = \ln y$ (где $\beta_0^* = \ln \beta_0$);

б) $x^* = \dfrac{1}{x}$, $y^* = \dfrac{1}{y}$;

в) $x^* = \dfrac{1}{x}$;

г) $y^* = \dfrac{1}{y}$.

10) Требуется оценить функцию регрессии $\tilde{y} = \beta_0 + \beta_1 \dfrac{1}{x}$ на основе информации: $Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}^T$ $X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T$. Указать вид матрицы X^*, участвующей в формуле МНК-оценок, после линеаризации зависимости:

а) $X^* = \begin{pmatrix} \dfrac{1}{x_1} & \dfrac{1}{x_2} & \dots & \dfrac{1}{x_n} \\ 1 & 1 & \dots. & 1 \end{pmatrix}^T$;

б) $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{-x_1} & e^{-x_2} & \dots. & e^{-x_n} \end{pmatrix}^T$;

в) $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \dfrac{1}{x_1} & \dfrac{1}{x_2} & \dots. & \dfrac{1}{x_n} \end{pmatrix}^T$;

г) $X^* = \begin{pmatrix} \dfrac{1}{x_1} & \dfrac{1}{x_2} & \dots. & \dfrac{1}{x_n} \end{pmatrix}^T$.

11) Указать, с помощью какого преобразования исходных переменных зависимость $\tilde{y} = \dfrac{x}{\beta_0 x + \beta_1}$ сводится к линейной:

а) $x^* = \dfrac{1}{x}$, $y^* = \dfrac{1}{y}$;

б) $y^* = \ln y$ (где $\beta_0^* = \ln \beta_0$);

в) $y^* = \dfrac{1}{y}$;

г) $x^* = \dfrac{1}{x}$.

12) Требуется оценить функцию регрессии $\tilde{y} = \beta_0 + \beta_1 \ln x$ на основе информации: $Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}^T$ $X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T$. Указать вид матрицы X^*, участвующей в формуле МНК-оценок, после линеаризации зависимости:

а) $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \dfrac{1}{x_1} & \dfrac{1}{x_2} & \dots\dots & \dfrac{1}{x_n} \end{pmatrix}^T$;

б) $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \ln x_1 & \ln x_2 & \dots\dots & \ln x_n \end{pmatrix}^T$;

в) $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{-x_1} & e^{-x_2} & \dots\dots & e^{-x_n} \end{pmatrix}^T$;

г) $X^* = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \dfrac{1}{\ln x_1} & \dfrac{1}{\ln x_2} & \dots\dots & \dfrac{1}{\ln x_n} \end{pmatrix}^T$.

13) Какая из приведенных зависимостей не поддается непосредственной линеаризации:

а) $y_i = \beta_0 + \beta_1 \cdot x_i^{\beta_2} + \varepsilon_i$;

б) $y_i = \beta_0 \cdot (x_i)^{\beta_1} \cdot \varepsilon_i$;

в) $y_i = \beta_0 + \beta_1 \dfrac{1}{x_i} + \varepsilon_i$;

г) $y_i = \dfrac{x_i}{\beta_0 x_i + \beta_1 + x_i \varepsilon_i}$.

14) Производственная функция Кобба-Дугласа характеризует:

а) нелинейное соотношение между нормой безработицы и процентом прироста заработной платы;

б) зависимость между спросом на товар (услугу) и его (ее) ценой;

в) зависимость между спросом на товар (услугу) и уровнем доходов потребителя;

г) нелинейную зависимость между объемом произведенной продукции и основными факторами производства.

4.4 Индивидуальное задание №5[8]

По данным Приложения А провести регрессионный анализ:

- из экономических или других соображений подобрать параметрический класс нелинейных зависимостей для модели регрессии;
- линеаризовать модель, оценить параметры и провести содержательный анализ.

4.5 Порядок выполнения индивидуального задания №5

Рассмотрим различные модели регрессии для описания зависимости между потреблением цыплят (y) и среднедушевым доход (x_1), стоимость 1 фунта цыплят (x_2) за n= 20 лет.

По результатам оценки линейной модели (рисунок 4.7 – 4.8) $\hat{y} = \underset{(0,947)}{29{,}696} + \underset{(0,0007)}{0{,}010}\, x_2$, можно сделать вывод, что модель значима; существенное влияние на потребление цыплят оказывает среднедушевой доход.

N=20	Regression Summary for Dependent Variable: y (11) R= ,95570976 R?= ,91338115 Adjusted R?= ,90856899 F(1,18)=189,81 p<,00000 Std.Error of estimate: 1,9774					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(18)	p-level
Intercept			29,69614	0,947734	31,33384	0,000000
x1	0,955710	0,069370	0,01024	0,000743	13,77704	0,000000

Рисунок 4.7 – Результаты множественной регрессии

Поскольку можно допустить, нормальный характер распределения регрессионных остатков (рисунок 4.8).

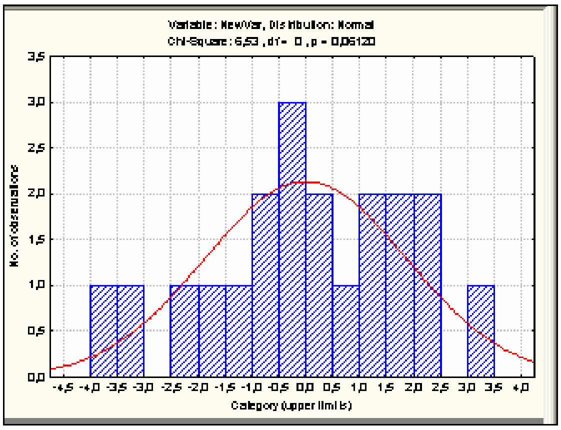


Рисунок 4.8 – Гистограмма регрессионных остатков

[8] Индивидуальные задания и пример их выполнения разработаны авторами совместно с преподавателем кафедры математических методов и моделей в экономике Жемчужниковой Ю.А.

Анализируя данные, можно предположить, что функцию спроса можно искать в форме Кобба-Дугласа: $\tilde{y} = \beta_0 \cdot (x_1)^{\beta_1} \cdot (x_2)^{\beta_2}$.

Модель нелинейной регрессии:

$$y_i = \beta_0 \cdot x_{i1}^{\beta_1} \cdot x_{i2}^{\beta_2} \cdot \delta_i,\ i = 1,2...,n, \qquad (4.7)$$

где δ_i - регрессионные остатки.

Линеаризуем модель логарифмированием:

$$\underbrace{\ln(y_i)}_{z_i} = \underbrace{\ln(\beta_0)}_{c_0} + \beta_1 \underbrace{\ln(x_{i1})}_{s_{i1}} + \beta_2 \underbrace{\ln(x_{i2})}_{s_{i2}} + \ln(\delta_i). \qquad (4.8)$$

Преобразуем наши данные, задав в строке формул необходимое преобразование (рисунок 4.9):

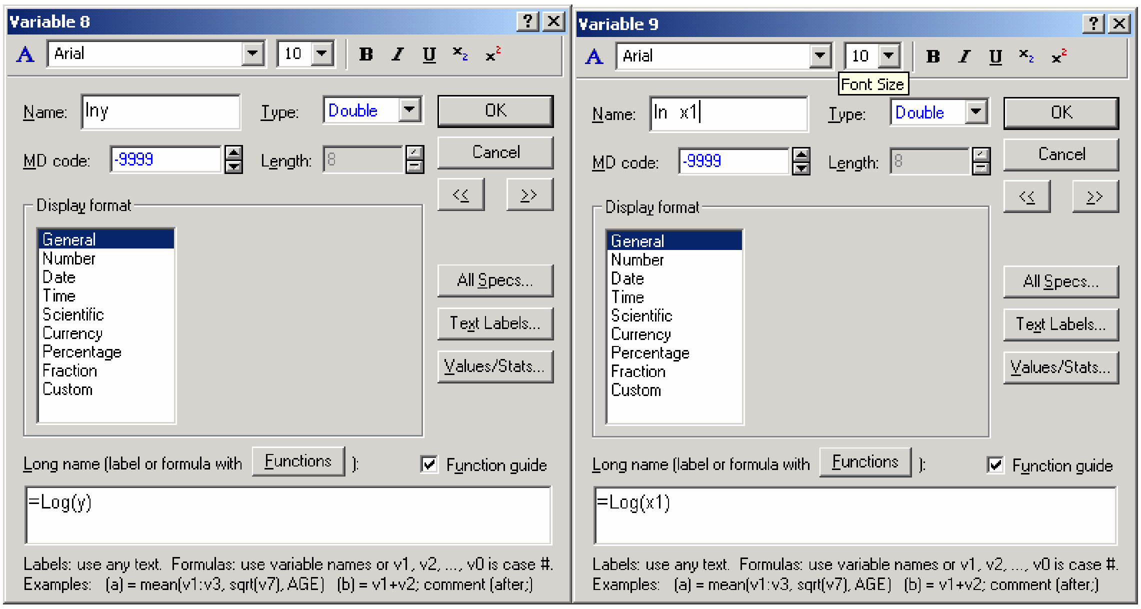


Рисунок 4.9 – Преобразование переменных

Отчет о построении регрессии ln y (z) на ln x1(s_1), ln x2(s_2) представлен на рисунке 4.10.

gression Summary for Dependent Variable: lny (11.sta)

Regression Summary for Dependent Variable: lny (11.sta)
R= ,98460432 R?= ,96944566 Adjusted R?= ,96585103
F(2,17)=269,69 p<,00000 Std.Error of estimate: ,02932

N=20	Beta	Std.Err. of Beta	B	Std.Err. of B	t(17)	p-level
Intercept			2,017103	0,124288	16,22923	0,000000
ln x1	1,409469	0,119609	0,427775	0,036301	11,78399	0,000000
lnx2	-0,469513	0,119609	-0,325470	0,082914	-3,92540	0,001090

Рисунок 4.10 - Отчет оценивания параметров регрессионной модели

На основании отчета (рисунок 4.10) можно сделать вывод, что модель регрессии значима; значимое влияние на результативный признак оказывают обе объясняющие переменные. В результате получили следующее уравнение регрессии: $\hat{z} = 2{,}017 + 0{,}428 s_1 - 0{,}325 s_2$ $\hat{R}^2 = 0{,}969$. Поскольку оценка коэффициента детерминации линеаризованной модели выше, чем линейной, то предпочтение следует отдать нелинейной модели (4.6), оценка которой имеет вид:

$$\hat{y} = 7{,}525 \cdot (x_1)^{0{,}428} \cdot (x_2)^{-0{,}325}$$

Из модели следует, что с ростом среднедушевого дохода на 1% при неизменной стоимости цыплят их потребление в среднем увеличится на 0,428%. Увеличение же стоимости цыплят на 1% при неизменном среднедушевом доходе приводит к уменьшению потребления в среднем на 0,325%.

4.6 Вопросы для самоконтроля

1 Приведите примеры линеаризуемых моделей и приемы их линеаризации.

2 МНК для нелинеаризуемых моделей – проблемы оценки и исследования

5 Временные ряды

5.1 Временной ряд – как случайный процесс

Допустим, рассматривается показатель У, характеризующий, к примеру, цену актива, урожайность зерновых конкретного хозяйства, величину инфляции и т.д. Ясно, что бессмысленно рассматривать этот показатель без привязки к конкретному моменту времени и , в этом плане, показатель есть функция от времени, но с другой стороны очевидно, что, в любой момент времени такой показатель является случайной величиной, определенной на некотором пространстве элементарных исходов Ω, то есть У является функцией $У(t,\omega)$. Такая функция называется случайной. Формально определение случайной функции приводится ниже.

Случайной функцией $У(t,\omega)$, t $\in$T называют измеримое отображение $У:\Omega \to R$ пространства элементарных исходов Ω в R, зависящее от параметра t.

Если $T=[a,b]$ - отрезок числовой оси, а параметр t $\in$T интерпретируют как время, то вместо термина «случайная функция» используют термин «случайный процесс».

Пример. Стоимость входного билета в игровой клуб составляет 100 р. до 14^{00}, 150 р. с 14^{00} до 18^{00} и 500 р. после 18^{00}. В любой момент посетитель может сыграть в кости. Бросаются две игральные кости и суммируются выпавшие очки. Выбрасывание двух очков оценивается номиналом в 15 р. до 14^{00}, 20 рублей с 14^{00} до 18^{00}, 100 р. с 18^{00}. Каждое следующее очко (после 1+1) дает надбавку на величину номинала. Размер «проигрыша» клуба за один сеанс в расчете на одного игрока есть случайный процесс осуществляющий отображение пространства $\Omega=\{(1,1),(1,2),(2,1)...(6,6)\}$в множество чисел $\{2,3,4,5,6,...12\}\subset R$, который для всякого фиксированного t является случайной величиной. Распределение случайной величины:

для $8 \le t \le 14$

Размер «проигрыша»	-65	-50	-35	-20	-5	10	25	40	55	70	85
p	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

для $14 \le t \le 18$

Размер «проигрыша»	-70	-50	-30	-10	10	30	50	70	90	110	130
p	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

для $18 \le t \le 24$

Размер «про-игрыша»	-425	-350	-275	-200	-125	-50	25	100	175	250	325
р	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Таким образом, рассматриваемый случайный процесс описан своими законами распределения для всех моментов времени. В общем случае в разных временных сечениях будем иметь случайные величины с разными законами распределения.

Если множество Т является **дискретным**, то будем говорить о случайном процессе с дискретным временем, иначе будем говорить о случайном процессе с непрерывным временем (в том случае, когда T=N=$\{1,2,....n,....\}$ или T=$\{t_0, t_1....t_n,....\}$ говорят о случайной последовательности).

При фиксированном $\omega_0 \in \Omega$ - У(t, ω_0)- неслучайная функция параметра t$\in$T, называемая траекторией случайного процесса или его реализацией (рисунок 5.1).

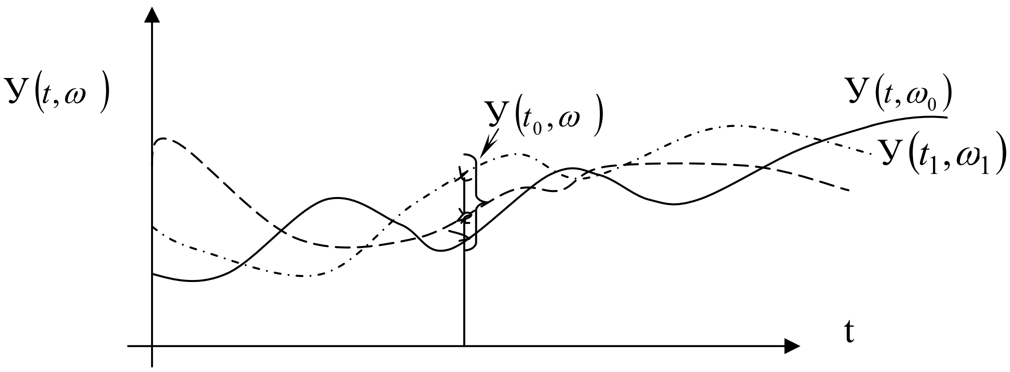


Рисунок 5.1 – Траектория случайного процесса

Под **временным рядом** принято понимать упорядоченную во времени последовательность величин У(0),У(1),…У(n), которая является реализацией некоторого случайного процесса У(t,ω) в дискретные, обычно равноотстоящие моменты времени $t_0, t_1....t_n,....$

В конкретной ситуации (ω- фиксировано) мы наблюдаем $y_1, y_2,..., y_n$-траекторию временного ряда (фиксированные уровни временного ряда).

Следует отметить, что понятие временного ряда, как последовательности случайных величин, не связано с природой случайного процесса, который, вообще говоря, является процессом с непрерывным временем, что, кстати, позволяет выбирать последовательность $t_0, t_1....t_n,....$, сообразуясь с целью и задачами исследования. С этих позиций встречающаяся в литературе терминология «дискретный временной ряд» не соответствует сути изучаемых процессов.

5.2 Описание случайных процессов

5.2.1 Конечномерные законы распределения

Зафиксируем любое $t_1 \in T, Y(t_1,\omega)$ - случайная величина. Закон распределения случайной величины $Y(t_1,\omega)$, $t_1 \in T$ будем называть **одномерным законом распределения случайного процесса**, тогда:

$F_Y(y/t_1) = P(Y(t_1)<y)$ – одномерная функция распределения случайного процесса;

$f_Y(y/t_1)$ – одномерная плотность распределения случайного процесса, где $F_Y(y/t_1) \int\limits_{-\infty}^{y} f_Y(s/t_1)ds$;

$P_Y(y/t_1)$ – одномерное распределение вероятностей случайного процесса, где $P_Y(y/t_1)$ – вероятность того, что $Y(t_1,\omega)$=y, причем $\sum\limits_{y} P_Y(y/t_1) = 1$

Если зафиксировать любые t_1,t_2, то в соответствующих сечениях получим случайные величины $Y(t_1,\omega)\,Y(t_2,\omega)$, рассматривая которые совместно, как двумерный случайный вектор $(Y(t_1,\omega)\,Y(t_1,\omega))^T$ будем говорить о двумерном законе распределения случайного процесса.

К примеру:

$F_Y(y_1,y_2/t_1,t_2) = P(Y(t_1,\omega)<y_1,Y((t_2,\omega)<y_2))$ – двумерная функция распределения случайного процесса;

$f_Y(y_1,y_2/t_1,t_2)$ – двумерная плотность случайного процесса, если

$F_Y(y_1,y_2/t_1,t_2)= \int\limits_{-\infty}^{y_1} \int\limits_{-\infty}^{y_2} f_Y(s_1,s_2/t_1,t_2)ds_1ds_2$;

$P_Y(y_1,y_2/t_1,t_2)$- двумерное распределение вероятностей случайного процесса, где $P_Y(y_1,y_2/t_1,t_2)$- вероятность того, что $Y(t_1,\omega)=y_1$, $Y(t_2,\omega)=y_2$ и $\sum\limits_{y_1} \sum\limits_{y_2} P_Y(y_1,y_2/t_1,t_2)=1$.

Аналогичным образом можно определить любой конечномерный закон распределения.

5.2.2 Числовые характеристики случайных процессов

Совокупность всех конечномерных законов распределения случайного процесса является полной его характеристикой, но зачастую ограничиваются одно и двумерными законами распределения и числовыми характеристиками (моментами в основном 1-го и 2-го порядков). При этом под моментами k-го порядка случайного процесса понимают соответствующие моменты его временных сечений.

Математическим ожиданием случайного процесса $Y(t,\omega)$, называют неслучайную функцию $m_Y(t)$ значение которой, при каждом $t \in T$, равно математическому ожиданию случайной величины $Y(t,\omega)$, являющейся сечением случайного процесса при фиксированном t:

$$m_y(t) = \int_{-\infty}^{\infty} y\, f_Y(y/t)dy\,. \qquad (5.1)$$

Дисперсией случайного процесса $Y(t,\omega)$, называют неслучайную функцию $\sigma_Y^2(t)$ значение которой, при каждом $t \in T$:

$$D_Y(t) = M\left[Y(t,\varpi) - m_y(t)\right]^2. \qquad (5.2)$$

$$D_Y(t) = \int_{-\infty}^{\infty} (y - m_y(t))^2 f_Y(y/t)dy$$

Ковариационной функцией случайного процесса $Y(t,\omega)$, $t \in T$, называют функцию $K_Y(t_1,t_2)$ двух скалярных переменных t_1 и t_2, значение которой при фиксированных $t_1,t_2 \in T$ равно ковариации двух случайных величин $Y(t_1,\omega)$ и $Y(t_2,\omega)$, определяемой следующим образом:

$$K_Y(t_1,t_2) = M\left[\left(Y(t_1,\omega) - m_Y(t_1)\right)\left(Y(t_2,\omega) - m_Y(t_2)\right)\right] = \mathrm{cov}\left[Y(t_1,\varpi), Y(t_2,\varpi)\right]$$

$$K_Y(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y_1 - m_y(t_1))(y_2 - m_y(t_2)) f_Y(y_1,y_2/t_1t_2)dy_1dy_2 \qquad (5.3)$$

Свойства ковариационной функции:

1) $K_Y(t_1,t_2) = K_Y(t_2,t_1)$;
2) $K_Y\ (t_1,t_1) = D_Y(t_1)$.

По аналогии с коэффициентом корреляции двух скалярных случайных величин в теории случайных процессов используют понятие корреляционной функции:

$$k_{Y_1Y_2}(t_1,t_2) = \frac{K_{Y_1Y_2}(t_1,t_2)}{\sqrt{\sigma_{Y_1}^2(t_1)}\sqrt{\sigma_{Y_2}^2(t_2)}} \qquad (5.4)$$

Методы и модели анализа временных рядов подразделяются в зависимости от того, является ли исследуемый временной ряд стационарным.

Случайный процесс y_t называется **стационарным в узком смысле**, если все конечномерные законы распределения вероятностей любой размерности инварианты относительно сдвига во времени, то есть при $\forall n, \forall t^* : t_i + t^* \in T,\ i = 1,2,..,n$:

$$F_Y(y_1, y_2, ..., y_n; t_1, t_2, ..., t_n) =$$
$$= F_Y(y_1, y_2, ..., y_n; t_1 + t^*, t_2 + t^*, ..., t_n + t^*) \quad (5.5)$$

или

$$f_Y(y_1, y_2, ..., y_n; t_1, t_2, ..., t_n) = f_Y(y_1, y_2, ..., y_n; t_1 + t^*, t_2 + t^*, ..., t_n + t^*) \quad (5.5.а)$$

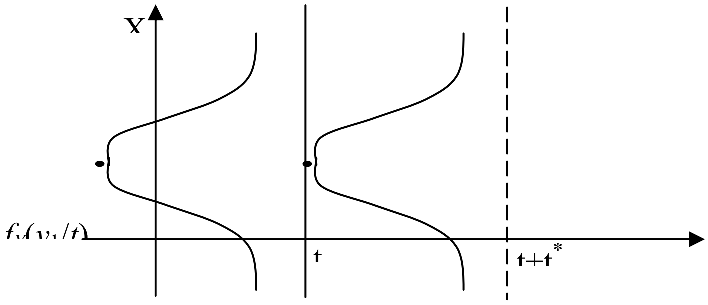


Рисунок 5.2 – График функции плотности вероятностей $f_Y(y_1/t)$

При анализе временных рядов моменты времени $t_1, t_2, ..., t_n$ принято, как правило, брать равностоящими, то есть $t_\tau = t_0 + \tau h,\ \tau = 1,2...$, где h - постоянный шаг (интервал) времени между соседними моментами наблюдениями, а величина τ задает количество тактов (шагов) между моментами наблюдения. Мы будем говорить об автоковариации первого порядка, то есть о ковариации величин разделенных τ=1 тактом времени. При τ=2 будем говорить говорят об автоковариации 2 порядка.

Автокорреляционной функцией называется функция

$R = \begin{pmatrix} \rho(1) \\ \rho(2) \\ ... \\ \rho(k) \end{pmatrix}$, где $\rho(i) = \dfrac{K_Y(t_\tau, t_{\tau+i})}{\sqrt{\sigma_Y^2(t_\tau)} \cdot \sqrt{\sigma_Y^2(t_{\tau+i})}}$ - коэффициент автокорреляции i-го порядка (i=1,…,k), который характеризует зависимость уровней временного ряда, разделенных i тактами времени.

Процессы, не удовлетворяющие определению, принято называть нестационарными в узком смысле.

Из определения стационарного временного ряда вытекает ряд свойств:

1) Математическое ожидание стационарного в узком смысле процесса является постоянной величиной.

2) Дисперсия стационарного в узком смысле процесса является постоянной величиной.

3) Ковариационная функция стационарного в узком смысле процесса не зависит от моментов времени t_i, t_j, а зависит от расстояния $\tau = t_i - t_j$ между ними.

Помимо процессов, стационарных в узком смысле, выделяют класс процессов, стационарных в широком смысле.

Случайный процесс y_t называется **стационарным в широком смысле**, если его математическое ожидание и дисперсия постоянны, а ковариационная функция зависит только от расстояния между сечениями [1, 7, 6].

Замечание. Для случайного процесса с дискретным временем при вычислении характеристик по формулам (5.1) – (5.3) знак интеграла заменяется на сумму.

5.3 Компонентный состав временных рядов

В практике прогнозирования принято считать, что значения уровней временных рядов экономических показателей складываются из следующих компонент: тренда, сезонной, циклической и случайной составляющих.

Под трендом понимают изменение, определяющее общее направление развития, основную тенденцию временного ряда. Это систематическая составляющая долговременного действия.

Наряду с долговременными тенденциями во временных рядах часто имеют место более или менее регулярные колебания – периодические составляющие рядов динамики.

Если период колебания не превышает 1 года, то их называют сезонными, чаще всего причиной их возникновения считаются природно-климатические условия.

При большем периоде колебания, считают, что во временных рядах имеет место циклическая составляющая.

Если из временного ряда удалить тренд и периодические составляющие, то останется нерегулярная компонента.

Случайные факторы разделяются на 2 вида в зависимости от факторов, под действием которых формируется случайная компонента:

- факторы резкого, внезапного действия;
- текущие факторы.

Первый тип факторов, как правило, вызывает значительные отклонения, факторы второго типа вызывают случайные колебания.

Если временной ряд представляется в виде суммы соответствующих компонент, то полученная модель носит название аддитивной, если в виде произведения – мультипликативной или смешанного типа [1, 6, 7]:

$$\begin{gathered} Y_t = u_t + s_t + v_t + \varepsilon_t, \\ Y_t = u_t * s_t * v_t * \varepsilon_t, \\ Y_t = u_t * s_t * v_t + \varepsilon_t, \end{gathered} \tag{5.6}$$

где Y_t - уровень временного ряда;

u_t - трендовая составляющая;

s_t - сезонная составляющая;

v_t - циклическая составляющая;

ε_t - случайная компонента.

Случайная компонента ε_t - «белый шум» - то есть стационарный нормально распределенный процесс с нулевым математическим ожиданием постоянной дисперсией, то есть:

$$\mu = M(\varepsilon_t) = 0;$$

$$D(\varepsilon_t) = \sigma^2;$$

$$\operatorname{cov}(\varepsilon_t \varepsilon_{t-k}) = 0, \forall k \neq 0.$$

При разложении рядов динамики на отдельные компоненты следует принимать во внимание, что компоненты исходного временного ряда по существу ненаблюдаемые и являются только теоретическими величинами. Но, не смотря на это, такой подход к разбиению фактический уровней временного ряда может оказаться довольно полезным для решения проблем анализа и прогнозирования на базе временного ряда.

Часто на практике возникают ситуации, когда на основании лишь визуального анализа не удается обнаружить наличие трендовой или периодических составляющих, что может быть вызвано, например, значительными колебаниями или как говорят «зашумленностью» временного ряда. В таких случаях для выявления наличия трендовой или периодической составляющей используют специальные критерии. Перечень критериев для проверки наличия трендовой и периодической составляющей представлен в таблице 5.1 [1, 6, 7].

Таблица 5.1 – Критерии проверки гипотез о постоянстве среднего и дисперсии временного ряда

постоянство средней	постоянство дисперсии
Параметрические	
- метод сравнения средних уровней ряда - использование критерия Стьюдента	- критерий Кокрена - критерий Бартлетта
Непараметрические	
- тест Манна-Уитни - метод Кокса-Стюарта - критерий серий, основанный на медиане выборки - критерий «восходящих» и «нисходящих» серий - критерий Фостера-Стюарта	- тест Сиджела-Тьюки - критерий Фостера-Стюарта

Рассмотрим подробнее некоторые из них. Наиболее удобными при проверке гипотезы об отсутствии тренда среднего являются так называемые сериальные критерии к ним относятся например, критерий восходящих и нисходящих серий и критерий серий, основанный на медиане выборки [1, 6, 7].

Выдвигается нулевая гипотеза о постоянстве среднего временного ряда (отсутствии тренда):

H_0: $M(y_t)=a=const$

H_1: $M(y_t) \neq const$.

Этапы реализации критерия серий, основанного на медиане выборки:

1) Исходные данные ранжируются в порядке возрастания или убывания для определения медианы.

2) Значения исходного ряда сравниваются с медианным значением, и ставится знак “+” или “–”: если $y_t > y_{med}$, то ставится “+”, если $y_t < y_{med}$, то ставится “-”, если $y_t = y_{med}$ - пропускается уровень и ставится 0. Таким образом, получается ряд из “+” и “–”. Последовательность “+” и “–” называется серией.

3) Определяется критерий $k_{\max}(T)$, то есть длина наибольшей серии.

4) Определяется критерий $V(T)$, то есть число серий.

5) Гипотеза не отвергается, если одновременно выполняются следующие неравенства:

$$\begin{cases} k_{\max}(T) < [3{,}3\ln(T+1)] \\ V(T) > [0{,}5(T+1-1{,}96\sqrt{T-1})] \end{cases}. \tag{5.7}$$

Если хотя бы одно из неравенств нарушено, то гипотеза об отсутствии тренда отвергается и делается вывод о наличии тренда.

Критерий «восходящих» и «нисходящих» серий реализуется в виде следующей последовательности шагов [1, 6, 7]:

- Определяется последовательность исходя из следующих условий:

$$\delta_i = \begin{cases} +, y_t - y_{t-1} > 0 \\ -, y_t - y_{t-1} < 0 \end{cases} \tag{5.8}$$

подсчитывается ν_T - число серий в совокупности δ_i, где под серией понимается подряд идущие «+» или «-»

- определяется продолжительность самой длинной серии $\tau_{\max}(T)$:

- проверка гипотез основана на том, что при условии случайности ряда протяженность самой длинной серии не должна быть слишком большой, а общее число серий не должно быть слишком маленьким.

$$\begin{cases} \nu(T) > [\frac{1}{3}(2T-1) - 1.96\sqrt{\frac{16T-29}{90}}] \\ \tau_{\max}(T) \le \tau_0(T) \end{cases}, \tag{5.9}$$

где $\tau_0(T)$ находится по таблице [1]:

T	n<26	26-153	153-170
$\tau_0(T)$	5	6	7

Если не выполняется одно из условий данной системы, следовательно, H_0 отвергается, то есть существует трендовая составляющая.

Для проверки наличия периодичности (сезонности)существует несколько методов: визуальный метод, критерий пиков и ям, дисперсионный критерий, гармонический критерий, анализ автокорреляционной и частной автокорреляционной функций.

Рассмотрим более подробно критерий пиков и ям.

Выдвигается нулевая гипотеза

H_0: случайный характер временного ряда, нет сезонности

H_1: присутствие сезонности

Условимся говорить, что в точке *к* временной ряд y_j , j=1,…,T имеет пик, если одновременно $y_{k-1}>y_k$, $y_{k+1}>y_k$ и имеет яму, если значение y_k меньше обеих соседних. Будем говорить, что *к* – экстремальная точка ряда, если в этой точке пик или яма.

Для трех последовательных значений y_i, y_{i+1}, y_{i+2}, i=1,…,(T-2) определим

$$x_i = \begin{cases} 1, & \text{если среди них есть экстремальная точка} \\ 0 & \text{иначе} \end{cases}$$

Тогда число экстремальных точек $e = \sum_{i=1}^{T-2} x_i$.

Алгоритм проверки гипотезы [1, 6, 7]:

1) Определяется число e экстремальных точек временного ряда по T имеющимся в распоряжении значениям.

2) Вычисляется t- статистика

$$t = \frac{3e - 2T + 4}{\sqrt{16T - 29}}\sqrt{10} \qquad (5.10)$$

В случае справедливости нулевой гипотезы статистика t распределена по стандартному нормальному закону. Расчетное значение t сравнивается с двусторонней критической точкой стандартного нормального распределения. Если $|t| < t_{\frac{\alpha}{2}}$ то гипотеза H_0 не отвергается.

5.4 Аналитическое выравнивание временного ряда

Для моделирования и прогнозирования одномерных временных рядов применяют методы, основанные на равнозначном и не равнозначном учете исходной информации. К методам, основанным на равнозначном учете исходной информации относятся методы основанные на использовании кривых роста (видов трендов) или тренд - сезонных моделей в случае наличия периодических колебаний во временных рядах. Суть данных методов заключается в том, что на основе анализа исходной информации в качестве тренда подбирается та или иная форма аналитической зависимости (f_t) и определяются показатели сезонности (в относительном выражении (индексами) (i_t)для мультипликативной сезонности, в абсолютном выражении (I_t) для аддитивной сезонности). Прогнозная модель имеет вид: $\hat{y}_t = f_t + I_t$.

Преимуществом использования данных методов является простота, а также менее жесткие требования относительно большой длины временного ряда.

К методам, основанным на неравнозначном учете исходной информации, относятся адаптивные методы прогнозирования – методы экспоненциального сглаживания и прогнозирование на основе моделей Бокса-Дженкинса. Преимуществом данных методов является возможность учета закономерности изменения явления в динамике по наиболее существенным, последним уровням, следовательно, получение точных и надежных прогнозных оценок, так как наиболее ценной является информация последних уровней.

Удобным средством описания одномерных временных рядов является их выравнивание с помощью тех или иных функций времени, называемых кривыми роста. Кривые роста позволяют получить выровненные или теоретические значения динамического ряда. Кривые роста хорошо себя зарекомендовали при прогнозировании социально-экономических явлений.

Процедура разработки прогноза и использованием кривых роста включает в себя следующие этапы [1, 6]:

1) выбор одной или нескольких кривых, формы которых соответствуют характеру изменения временного ряда;
2) оценка параметров выбранных кривых;
3) проверка адекватности выбранных кривых прогнозируемому процессу и окончательный выбор кривой роста;
4) расчет точечного и интервального прогнозов.

Изучая тренды, следует иметь в виду, что существует, вообще говоря, несколько их разновидностей. Первым и самим очевидным типом тренда представляется **тренд среднего**, когда временной ряд выглядит как колебания около медленно возрастающей или убывающей величины. Второй тип трендов — это **тренд дисперсии**. В этом случае во времени меняется амплитуда колебаний переменной. Иными словами, процесс гетероскедастичен. Часто экономические процессы с возрастающим средним имеют и возрастающую дисперсию.

Третий и более тонкий тип тренда, визуально не всегда наблюдаемый, изменение значимости одной из компонент временного ряда.

Проводя разложение ряда на компоненты, мы, как правило, подразумеваем под трендом изменение среднего уровня переменной.

В настоящее время в литературе описывается несколько десятков кривых роста, многие из которых широко применяются для выравнивания временных рядов. Условно их можно разделить на 3 класса в зависимости от того, какой тип динамики развития они хорошо описывают [1, 7]:

1) кривые, используемые для описания процессов с монотонным характером развития и отсутствием пределов роста.

2) кривые, описывающие процессы, имеющие предел роста в исследуемом периоде – кривые насыщения.

3) кривые насыщения, имеющие точку перегиба - S-образные кривые.

Среди кривых роста 1 типа прежде всего следует выделить класс полиномов, описывающихся уравнением вида

$$y_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_p t^p + \varepsilon_t, \tag{5.11}$$

где $a_i, i = \overline{1, p}$ - параметры многочлена,

t - независимая переменная времени.

Коэффициенты полиномов невысоких степеней могут иметь конкретную интерпретацию в зависимости от содержания динамического ряда. Так, a_1 можно трактовать как скорость роста, a_2 как ускорение роста, a_3 - изменение ускорения и a_0 как начальный уровень ряда динамики. Оценка параметров в модели осуществляется с помощью МНК.

Введем в рассмотрение вектор оценок параметров полинома:

$$\widehat{a} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \ldots \\ a_p \end{pmatrix} = (Z^T Z)^{-1} Z^T Y, \tag{5.12}$$

где $У = \begin{pmatrix} у_1 \\ у_2 \\ \ldots \\ у_Т \end{pmatrix}$, $Z = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 2^2 & \ldots & 2^p \\ 1 & 3 & 3^2 & \ldots & 3^p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & T & T^2 & \ldots & T^p \end{pmatrix}$

Для класса экспоненциальных кривых, в отличие от полиномов, характерной является зависимость приростов от величины самой функции. Эти кривые хорошо описывают лавинообразные процессы, когда прирост зависит от уже достигнутого уровня.

Простая экспоненциальная или показательная кривая имеет вид

$$y_t = a \cdot b^t \tag{5.13}$$

Причем, если $b > 1$, то кривая растет вместе с увеличением t и наоборот. Параметр a характеризует начальное условие развития, а параметр b - постоянный темп роста.

Более сложным вариантом экспоненциальной кривой является логарифмическая парабола, уравнение которой:

$$y_t = a \cdot b^t \cdot c^{t^2} \tag{5.14}$$

Данное выражение можно прологарифмировать, а оценку параметров осуществить с помощью МНК.

В рамках 2 класса кривых следует рассматривать кривые, имеющие отличную от 0 асимптоту. Примером такой кривой служит модифицированная экспоненциальная кривая:

$$y_t = k + a \cdot b^t. \tag{5.15}$$

Наиболее известными кривыми 3 типа являются кривая Гомперца и логистическая кривая Перля-Рида. Особенностью уравнений этих кривых является то, что их параметры могут быть определены МНК лишь приближенно, поэтому используется ряд искусственных методов, основанных на разбиении исходного временного ряда на промежутки.

Уравнение кривой Гомперца имеет следующий вид:

$$y_t = k \cdot a^{b^t}. \tag{5.16}$$

Данная кривая несимметрична и для решения экономических задач наиболее интересный вариант, когда $\log a < 0$ и $b < 1$.

Кривая Перля-Рида выражает геометрическую прогрессию, в которой возрастание затухает по мере приближения к некоторому определенному пределу. Максимальный предел устанавливается, прежде всего, на основании конкретного изучения исследуемого социально-экономического явления.

Кривая Перля-Рида имеет следующий вид:

$$\frac{1}{y_t} = k + a \cdot b^t. \tag{5.17}$$

На практике к выбору формы кривой подходят исходя из значения критерия минимума суммы квадратов отклонений фактических значений уровней от расчетных, полученных выравниванием.

Существует несколько методов для оценки сезонных вариаций. Основная идея всех этих методов заключается в том, что в реальном ряду сначала оценивается и убирается тренд, а потом сглаживается возможная нерегулярная компонента. Оставшиеся данные будут содержать только сезонные вариации. Сезонные величины собираются и усредняются для получения числа (а точнее числового индекса) для каждого наблюдаемого интервала года (недели, месяца, квартала).

При использовании аддитивной декомпозиции, оценки трендовой, сезонной и нерегулярной компонент суммируются, что в результате дает исходный ряд. Если используется мультипликативная декомпозиция, то для того, чтобы восстановить исходную последовательность, отдельные компоненты перемножаются. В этом случае сезонная компонента представляется набором числовых индексов. Эти числа показывают, какие периоды в году характеризуются относительно низкими показателями, а какие – относительно высокими.

5.5 Экспоненциальные методы сглаживания временных рядов

Суть метода экспоненциального сглаживания заключается в том, что уровни исходного временного ряда взвешиваются с помощью скользящей средней, с экспоненциальным характером изменения весов. Общая формула (рекуррентная формула) расчета экспоненциальной средней имеет вид [1, 6, 7]:

$$\mathrm{S}_{\mathrm{t}}(y) = \alpha y_t + (1-\alpha) S_{t-1}(y) = \mathrm{S}_{\mathrm{t}} = \alpha y_t + \beta S_{t-1}, \tag{5.18}$$

где S_{t} - значение экспоненциальной средней;

α - параметр сглаживания, α - постоянная величина, $0 < \alpha < 1$;

$\beta = 1-\alpha$ - параметр затухания;

$t = 1, 2, ..., T$;

S_0 - величина, характеризующая начальное условие.

Величина S_{t} - взвешенная сумма всех членов ряда. Причем веса отдельных уровней ряда убывают по мере их удаления в прошлое (в зависимости от возраста наблюдений). Экспоненциальная средняя играет роль «фильтра», поглощающего колебания временного ряда. С одной стороны, следует увеличивать вес более свежих наблюдений, что может быть достигнуто повышение α, с другой стороны, для сглаживания случайных отклонений величину α нужно уменьшить.

Если α стремится к 1 – это означает, что при прогнозе в основном учитывается влияние только последних уровней временного ряда.

Если α стремится к 0 – это означает, что при прогнозе учитывается влияние прошлых уровней временного ряда.

Автор метода простого экспоненциального сглаживания Р.Г. Браун предложил следующую формулу расчета α.

$\alpha = 2/(T+1)$

где Т- число уровней временного ряда, вошедших в интервал сглаживания.

Модель экспоненциального сглаживания с аддитивным сезонным эффектом имеет вид [1, 6, 7]:

$$y_t = f_t + g_t + \varepsilon_t, \tag{5.19}$$

где f_t — некоторый усредненный уровень временного ряда в момент t после устранения сезонного эффекта;

g_t – аддитивный показатель сезонности;

$t = 1, 2, ..., T$.

Модель экспоненциального сглаживания с мультипликативным сезонным эффектом имеет вид:

$$y_t = f_t \cdot m_t + \varepsilon_t, \tag{5.20}$$

где m_t – мультипликативный показатель сезонности; $t = 1, 2, ..., T$.

Множество комбинаций различных типов тенденций с циклическими эффектами аддитивного и мультипликативного характера можно представить в виде обобщенной формулы:

$$f_t = \alpha d_{1t} + (1-\alpha) d_{2t}, \tag{5.21}$$

где α - параметр сглаживания, причем $0 < \alpha < 1$;

d_{1t}, d_{2t} - характеристики модели;

$$d_{1t} = \begin{cases} y_t \text{ - если сезонный эффект отсутствует} \\ y_t - g_{t-k} \text{ - в случае аддитивного сезонного эффекта} \\ \dfrac{y_t}{m_{t-k}} \text{ - в случае мультипликативного сезонного эффекта} \end{cases}$$

g_{t-k} и m_{t-k} - аддитивный и мультипликативный показатели сезонности с периодом колебания k;

$t = k, k+1, ..., T$;

g_0, m_0 - начальные условия.

Таким образом, d_1 представляет собой текущую оценку процесса y_t или очищенную от сезонных колебаний (при их наличии) с помощью коэффициентов сезонности g_{t-k} или m_{t-k}, рассчитанных для предшествующего цикла.

$$d_{2t} = \begin{cases} f_{t-1} \text{ - при отсутствии тенденции} \\ f_{t-1} + c_{t-1} \text{ - в случае аддитивного роста} \\ f_{t-1} \cdot r_{t-1} \text{ - в случае экспоненциального роста} \end{cases}$$

В этом выражении c_{t-1} - абсолютный прирост, характеризующий изменение среднего уровня процесса, или аддитивный коэффициент роста, r_{t-1} - коэффициент экспоненциального роста.

Адаптация всех перечисленных параметров осуществляется с помощью экспоненциального сглаживания:

$$g_t = \gamma \cdot (y_t - f_t) + (1-\gamma) \cdot g_{t-k}, \tag{5.22}$$

$$m_t = \gamma \cdot \frac{y_t}{f_t} + (1-\gamma) \cdot m_{t-k}, \tag{5.23}$$

$$c_t = \beta \cdot (f_t - f_{t-1}) + (1-\beta) \cdot c_{t-1}, \tag{5.24}$$

$$r_t = \beta \cdot \frac{f_t}{f_{t-1}} + (1-\alpha_r) \cdot r_{t-1}, \tag{5.25}$$

где $0 < \alpha, \beta, \gamma < 1$; $t = k, k+1, ..., T$.

Уинтерс предлагает находить оптимальные значения парамтров адаптации экспериментальным путем, с помощью сетки значений α, β, γ (например, (0,2; 0,1; 0,1) , (0,3; 0,1; 0,2) , . . .). В качестве критерия сравнения вариантов рекомендуется стандартное отклонение ошибки.

5.6 Тестовые задания для самоконтроля

1) В сериальных критериях стационарности предполагается, что при условии случайности ряда:

а) протяженность самой длинной серии должна быть большой, а общее число серий не должно быть слишком маленьким

б) протяженность самой длинной серии не должна быть слишком большой, а общее число серий должно быть маленьким

в) протяженность самой длинной серии не должна быть слишком большой, а общее число серий не должно быть слишком маленьким

г) протяженность самой длинной серии должна быть большой, а общее число серий должно быть маленьким

2) Для определения того, во сколько раз в среднем за единицу времени изменяется уровень временного ряда, используется:

а) абсолютный прирост

б) средний абсолютный прирост

в) средний темп роста

г) средний относительный прирост

3) К кривым, описывающим процессы, имеющим предел роста в исследуемом периоде относятся:

а) полином второго порядка, логарифмический тренд и гиперболический тренд

б) логистическая кривая, логарифмический тренд и гиперболический тренд в) логарифмический тренд, кривая Перля-Рида

г) кривая Перля-Рида, равносторонняя гипербола

4) Временной ряд описывается моделью $y_t = 2.3 + 0.1t + 1.2t^2 + e_t$. Среднее ускорение ряда составит:

а) 1,2

б) 2,4

в) 0,6

г) 0,1

5) Пусть Y_t - уровни временного ряда, u_t - трендовая составляющая, s_t - сезонная составляющая, v_t - циклическая составляющая, e_t - случайная компонента. Тогда модель временного ряда смешанного типа:

а) имеет вид $Y_t = u_t + s_t + v_t + e_t$

б) имеет вид $Y_t = u_t * s_t * v_t * e_t$

в) имеет вид $Y_t = u_t * s_t * v_t + e_t$

г) имеет вид $Y_t = u_t + s_t + v_t$

6) Отдельные наблюдения временного ряда называются

а) уровнями

б) темпами

в) приростами

г) показателями

7) График автокорреляционной функции называется

а) автограммой

б) автокоррелограммой

в) коррелограммой

г) у него нет специального названия

8) Множество комбинаций различных типов тенденций с циклическими эффектами аддитивного и мультипликативного характера можно представить в виде обобщенной формулы:

а) $f_t = \alpha_f d_1 + (1-\alpha_f)d_2$

б) $f_t = (1-\alpha_f)d_1 + \alpha_f d_2$

в) $f_t = (1-\alpha_f)g_t + \alpha_f d_2$

г) $f_t = (1-\alpha_f)d_1 + \alpha_f g_t$

9) Первый уровень временного ряда равен 10, последний уровень 100000. Всего во временном ряду 5 наблюдений. Средний темп роста на базисной основе составит:

а) 100

б) 3,2

в) 10

г) 32

10) Частный коэффициент автокорреляции 3 порядка показывает связь между:

а) y_t и y_{t-3}

б) y_t и y_{t-3} при фиксированном влиянии $y_{t-4}, y_{t-5,...}$

в) y_{t-2} и y_{t-3} при фиксированном влиянии $y_{t-4}, y_{t-5,...}$

г) y_t и y_{t-3} при фиксированном влиянии $y_{t-1}, y_{t-2,...}$

11) К систематическим составляющим временного ряда относятся

а) трендовая и случайная составляющая

б) трендовая и циклическая составляющая

в) сезонная и случайная составляющая

г) циклическая и случайная составляющая

12) В критерии «восходящих и нисходящих» серий ставится знак если:

а) если $\begin{cases} +, y_t > y_{t-1}, y_{t-2}, ..., y_1 \\ -, иначе \end{cases}$

б) если $\begin{cases} +, y_t > y_{t-1} \\ -, иначе \end{cases}$

в) если $\begin{cases} +, y_t < y_{t-1}, y_{t-2}, ..., y_1 \\ -, иначе \end{cases}$

г) если $\begin{cases} +, y_t < y_{t-1} \\ -, иначе \end{cases}$

13)При реализации критерия серий, основанного на медиане выборки получены 10 серий, с длиной наибольшей серии равной 7. При этом величина $\{[3{,}3\ln(T+1)]=8$. Можно сделать вывод, что:

а) во временном ряду присутствует тренд среднего

б) во временном ряду отсутствует тренд среднего

в) во временном ряду присутствует тренд дисперсии

г) во временном ряду отсутствует тренд дисперсии

14) Уравнение кривой Гомперца имеет вид:

а) имеет вид $y_t = k \cdot a^{b^t}$

б) имеет вид $\frac{1}{y_t} = k + a \cdot b^t$

в) имеет вид $y_t = k + a \cdot b^t$

г) имеет вид $y_t = a \cdot b^t \cdot c^{t^2}$

15) Для определения того, на сколько в среднем в единицу времени изменяется уровень временного ряда, используется

а) абсолютный прирост

б) средний абсолютный прирост

в) средний темп роста

г) средний относительный прирост

5.7 Индивидуальное задание №6

По данным Приложения В:

1) провести анализ компонентного состава временного ряда;

2) определить тип модели (аддитивный, мультипликативный, смешанный);

3) выбрать значения параметров адаптации;

4) исследовать адекватность модели;

5) осуществить прогнозирование на 3 периода.

5.8 Порядок выполнения индивидуального задания №6

Рассмотрим процедуру моделирования и прогнозирования на основе сезонных адаптивных моделей, используя поквартальную информацию о среднедушевых денежных доходах населения Оренбургской области (y_t) за период 1997-2005 гг.

Окно с частью данных для анализа представлено на рисунке 5.3.

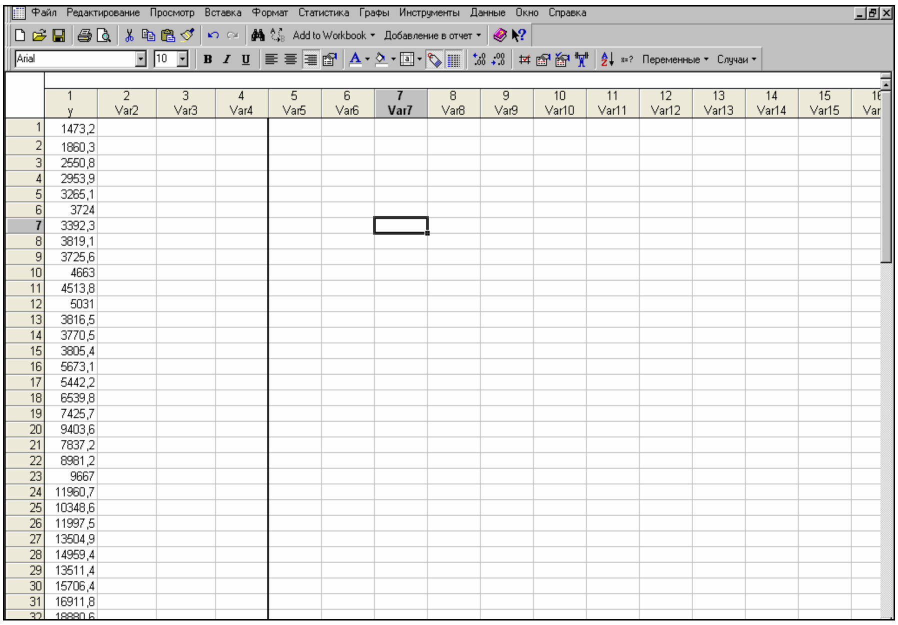


Рисунок 5.3 – Исходные данные

Первым этапом при определении компонентного состава временного ряда является построение графика исходного временного ряда. Для построения графика в меню системы открыть **Statistics –Критерии, Дополнительные линейные/нелинейные модели** и выбрать в появившемся меню (рисунок 5.4) строку **Time Series Analysis/ Forecasting - Анализ временных рядов и прогнозирование.**

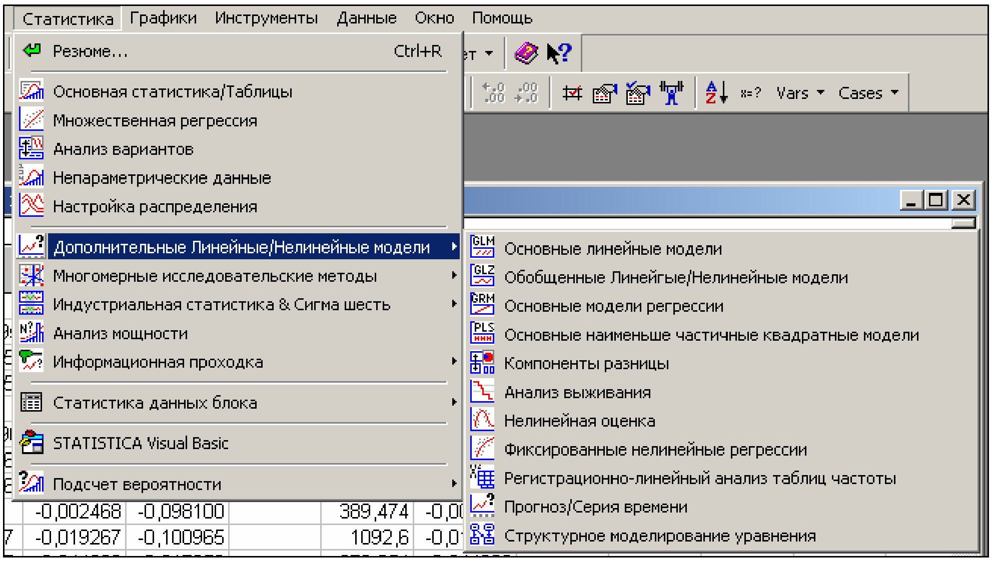


Рисунок 5.4 – Выбор пункта меню для проведения экспоненциального сглаживания

На экране откроется окно рисунок 5.5.

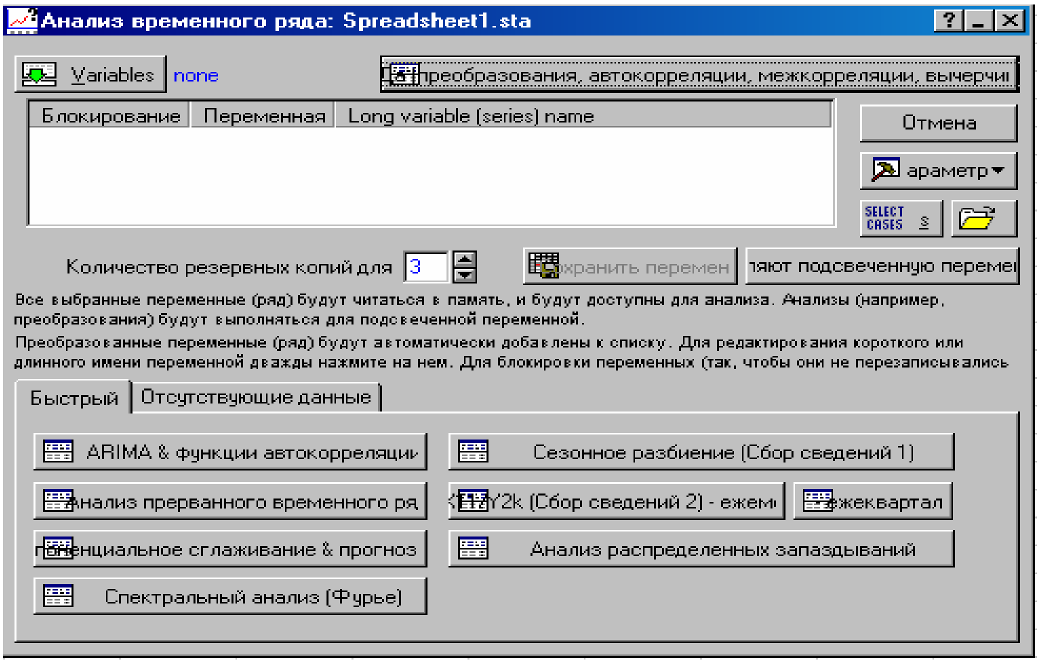


Рисунок 5.5 – Выбор пунктов меню для экспоненциального сглаживания

Выбирается пункт **Exponsmooting/Экспоненциальное сглаживание и прогноз**. Для задания переменных воспользуемся кнопкой **Variables/Переменные** из панели **Экспоненциальное сглаживание и прогноз** 5.6.

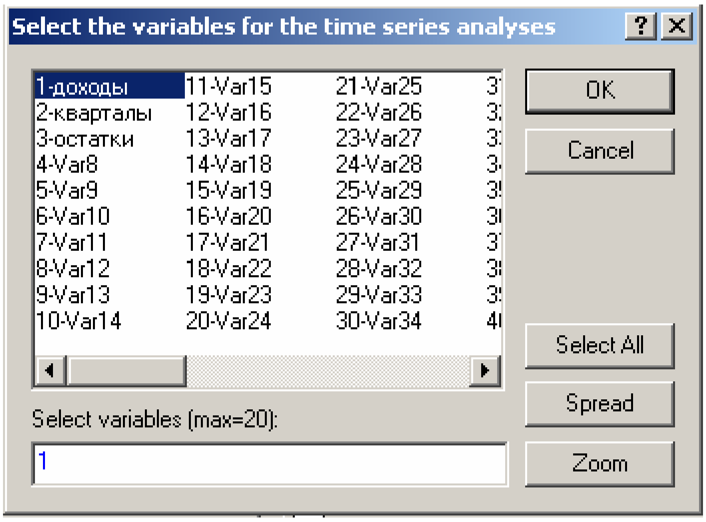


Рисунок 5.6 – Выбор переменной для проведения экспоненциального сглаживания

После выбора переменной необходимо щелкнуть на кнопке **OK**, вновь окажемся в панели модуля **Экспоненциальное сглаживание**.

Для построения графика, отображающего динамику изменения показателя выберем опцию **Review series/Показ переменной** - нажав кнопку **Review highlighted variable/Показ высвеченной переменной –** получим график ряда y_t (рисунок 5.7):

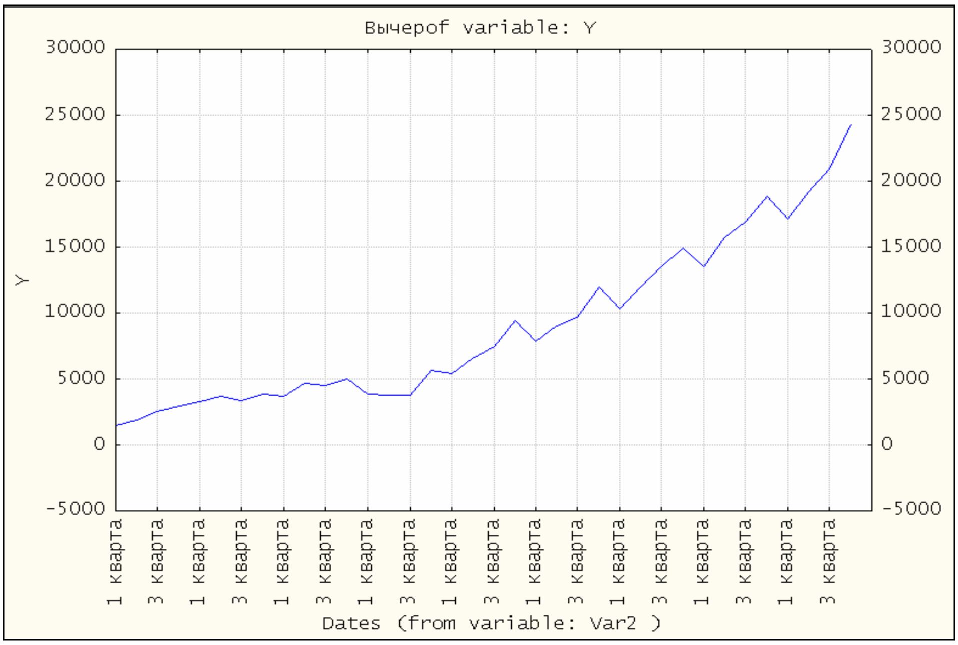


Рисунок 5.7 – Динамика среднемесячных доходов по кварталам

Нажатие на кнопку **Advanced** позволяет перейти к окну функциональных возможностей модуля **Экспоненциальное сглаживание** (рисунок 5.8).

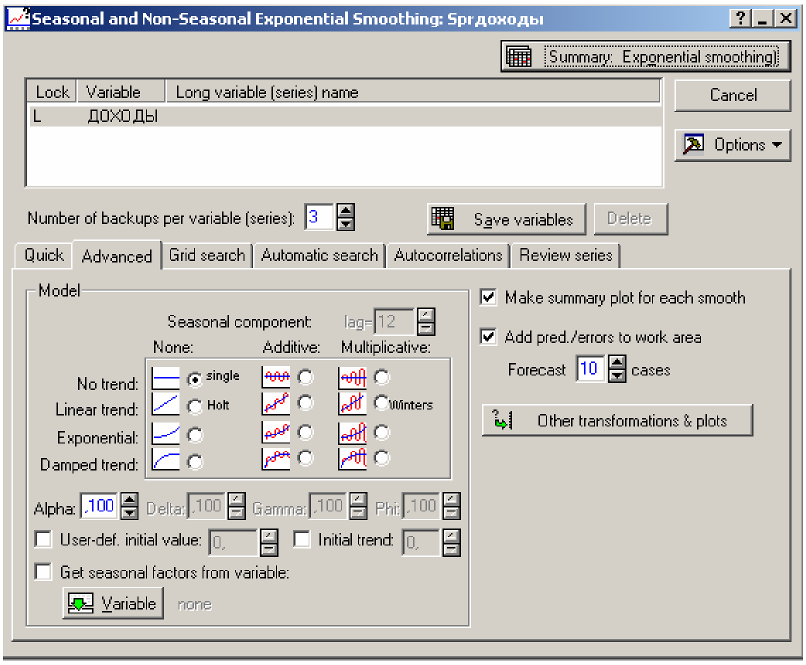


Рисунок 5.8 – Выбор пунктов меню для экспоненциального сглаживания

В зависимости от вида модели аддитивная либо мультипликативная, по наличию либо отсутствию тренда и сезонности выбирается один из пунктов. В нашем случае присутствует нелинейный тренд и аддитивная сезонность (размах сезонности не возрастает). Выберем соответствующую закладку (рисунок 5.9). Сезонность квартальная, количество лагов равно 4, лаг = 12 предлагается автоматически.

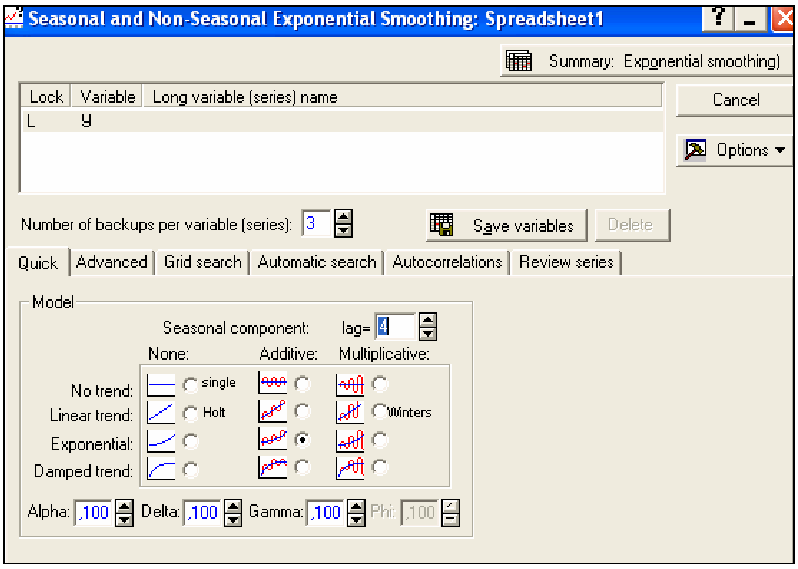


Рисунок 5.9 – Выбор вида экспоненциального сглаживания

Определить значения параметров адаптации можно автоматически, воспользовавшись опцией **Automatic search/ Автоматический поиск** либо вручную используя кнопку **Grid search/Поиск по сетке**.

Наилучшим значениям параметров адаптации соответствуют минимальные значения мер ошибок (рисунок 5.10).

Model Number	Parameter grid search (Smallest abs. errors are highlighted) (Sprдоходы) Model: Expon. trend, add.season (4); S0=1743, T0=1,126 ДОХОДЫ								
	Alpha	Delta	Gamma	Mean Error	Mean Abs Error	Sums of Squares	Mean Squares	Mean % Error	Mean Abs % Error
559	0,700000	0,900000	0,100000	-173,421	603,0428	19162898	532302,7	-3,41992	11,03045
478	0,600000	0,900000	0,100000	-202,659	615,1921	19280268	535563,0	-3,88349	11,25722
550	0,700000	0,800000	0,100000	-171,681	608,3797	19408946	539137,4	-3,40260	11,10932
469	0,600000	0,800000	0,100000	-200,217	616,5596	19418188	539394,1	-3,86010	11,21793
460	0,600000	0,700000	0,100000	-197,775	616,9567	19641177	545588,3	-3,83681	11,17075
541	0,700000	0,700000	0,100000	-170,070	614,5142	19665871	546274,2	-3,38570	11,19192
479	0,600000	0,900000	0,200000	-110,244	599,6633	19696960	547137,8	-2,74375	11,06955
640	0,800000	0,900000	0,100000	-153,459	612,7212	19840647	551129,1	-3,09432	11,20740
470	0,600000	0,800000	0,200000	-108,864	604,2068	19859283	551646,7	-2,71930	11,04654
532	0,700000	0,600000	0,100000	-168,628	619,8066	19914512	553180,9	-3,36924	11,26764

Рисунок 5.10 – Результаты определения оптимальных значений параметров адаптации методом поиска на сетке

В данном окне модуля представлены оценки мер ошибок **Mean Error/Средняя ошибка -** вычисляется простым усреднением ошибок на каждом шаге, **Mean Abs Error/Средняя абсолютная ошибка-** вычисляется как среднее абсолютных ошибок, **Sums of squares /Сумма квадратов ошибок** и **Mean squares/ среднеквадратическая ошибка -** вычисляются как сумма (или среднее) квадратов ошибок, **Mean % Error/Средняя относительная ошибка -** вычисляется как среднее относительных ошибок,
Mean Abs % Error/Средняя абсолютная относительная ошибка - вычисляется как среднее абсолютных относительных ошибок. Это наиболее часто используемые индексы качества подгонки. Минимальные значения мер ошибок соответствуют параметру сглаживания **Alpha/**α=0.7, параметру сезонного сглаживания **Delta/**$\delta = 0.9$**,** параметру сглаживания тренда **Gamma/**$\gamma = 0.1$.

Оценка модели экспоненциального сглаживания с мультипликативным ростом и аддитивным сезонным эффектом:

$$\hat{f}_t = 0.7(y_t - \hat{g}_{t-4}) + 0.3\hat{f}_{t-1} * \hat{r}_{t-1}$$

$$\hat{g}_t = 0.9(y_t - \hat{f}_t) + 0.1\hat{g}_{t-4}$$

$$\hat{r}_t = 0.1\frac{\hat{f}_t}{\hat{f}_{t-1}} + 0.9\hat{r}_{t-1}$$

Нажатие на кнопку **Advanced** позволит установить оптимальные значения параметров сглаживания и определить период прогнозирования в опции **Forecast/ Прогнозирование** (рисунок 5.11).

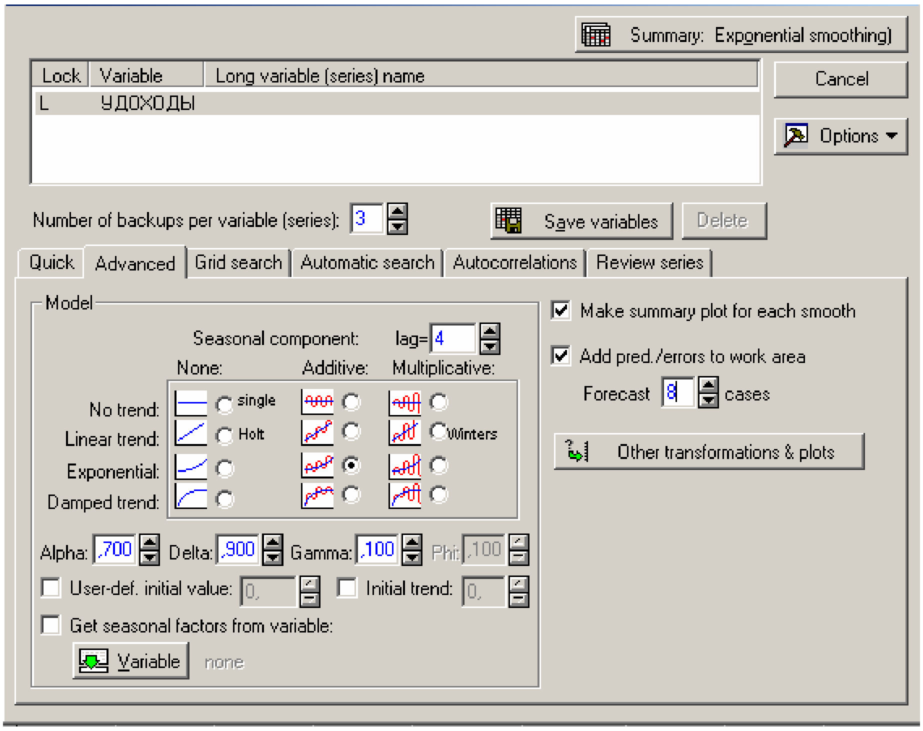


Рисунок 5.11 – Модуль Экспоненциальное сглаживание

После определения всей необходимой информации для экспоненциального сглаживания, щелкните по кнопке **Summary. Exponential Smoothing** в правом верхнем углу окна. Результаты расчетов приведены в виде отчета на рисунке 5.12.

Exp. smoothing: Additive season (4) S0=1743, T0=1,126 (Sprдоходы)
Expon.trend, add.season; Alpha=,700 Delta=,900 Gamma=,100
ДОХОДЫ

Case	ДОХОДЫ	Smoothed Series	Resids	Seasonal Factors
1	1473,20	1274,34	198,86	-688,199
2	1860,30	2304,92	-444,62	-78,133
3	2550,80	2283,06	267,74	-35,381
4	2953,90	3628,54	-674,64	801,714
5	3265,10	1977,28	1287,82	
6	3724,00	3833,37	-109,37	
7	3392,30	4566,74	-1174,44	
8	3819,10	4789,12	-970,02	
9	3725,60	3574,60	151,00	
10	4663,00	4172,88	490,12	
11	4513,80	5022,87	-509,07	
12	5031,00	5850,54	-819,54	
13	3816,50	5159,26	-1342,76	
14	3770,50	4725,89	-955,39	
15	3805,40	4003,64	-198,24	
16	5673,10	4681,78	991,32	

Рисунок 5.12 – Наблюденные, сглаженные значения ряда динамики показателя, значения остатков и показателей сезонности

Для проведения теста на нормальный характер распределения остатков, скопируем столбец **Residual** в окно с исходными данными. Затем в меню системы **Statistica** выберем пункт **Distribution Fitting** (рисунок 5.13). На экране появится окно:

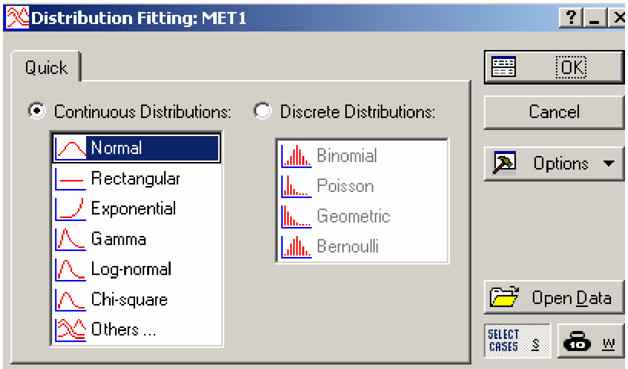


Рисунок 5.13 - Выбор вида распределения остатков

В появившемся окне выберем распределение **Normal – Нормальное** и щелкнем по кнопке **OK.** После чего на экране появится окно (рисунок 5.14):

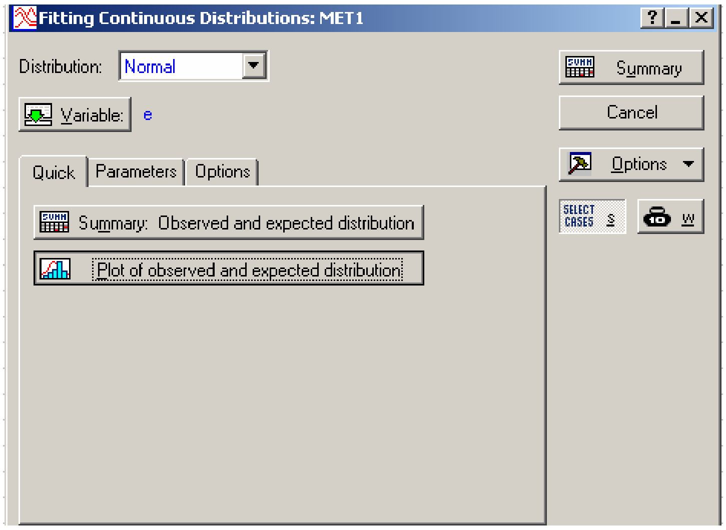


Рисунок 5.14 - Выбор пунктов для построения гистограммы остатков

В данном окне сначала необходимо выбрать переменные, используя кнопку **Variable.** Кроме того, в данном модуле, используя кнопку **Parameters – Параметры,** можно изменить количество интервалов, верхнюю и нижнюю границы интервалов и т.д. Для получения графика нормального распределения, нажмем по кнопке **Plot of observed and expected distribution.**

На экране появится окно (рисунок 5.15), содержащее гистограмму распределения, значение χ^2 – критерия, степени свободы, значимость нулевой гипотезы.

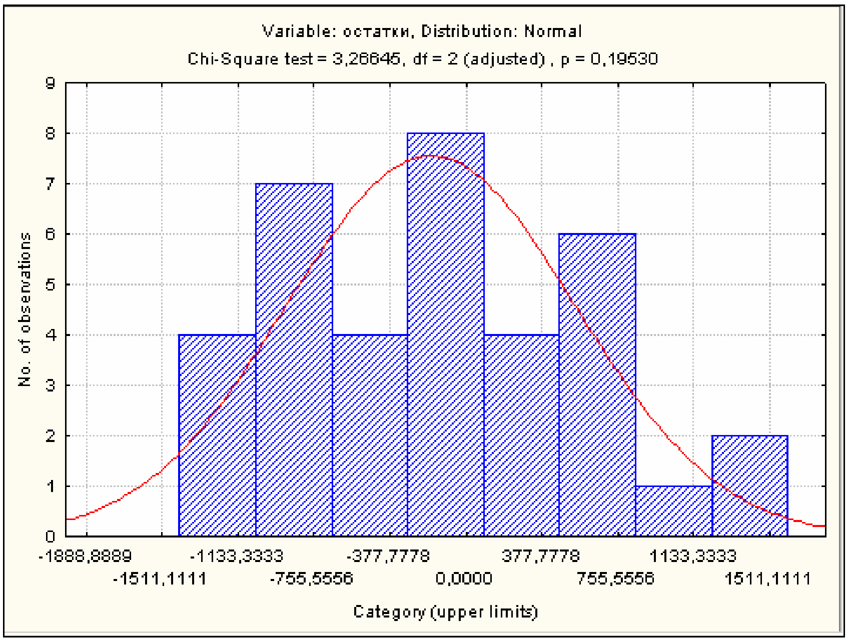


Рисунок 5.15 - График распределения остатков

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что распределение остатков не отличаются от нормального, так как значимость нулевой гипотезы (P=0,19) больше, чем заданный.

Далее можно приступить к исследованию некоррелированности остатков модели. Некоррелированность остатков исследуются в специальном окне **Autocorrelations –Автокорреляция**. Для этого необходимо щелкнуть мышкой по кнопке **Autocorrelations** в окне рисунка 5.11. В появившемся окне можно установить уровень значимости в опции **p-level for highlighting** и порядок автокорреляции в опции **Number of lags.** Нажатие на кнопку **Autocorrelations – Автокорреляция** даст оценку автокорреляционной функции (рисунок 5.16).

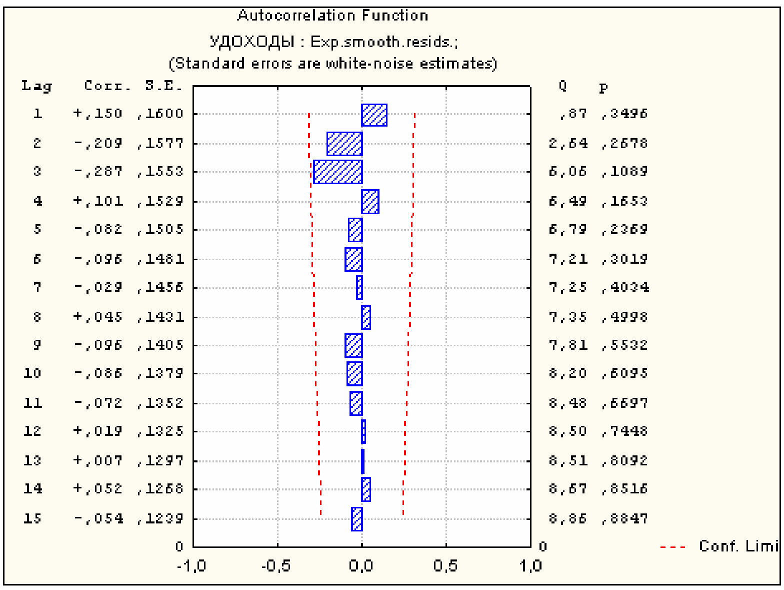


Рисунок 5.16 – Оценка автокорреляционной функции остатков

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что остатки некоррелированы. Значимость коэффициентов автокорреляции проверяется на основе расчета Q-статистики Бокса-Льюнга, значения которого приводятся вместе со значениями значимости нулевой гипотезы. Нажатие на кнопку **Partial Autocorrelations – Частная Автокорреляция** даст оценку частной автокорреляционной функции (5.17).

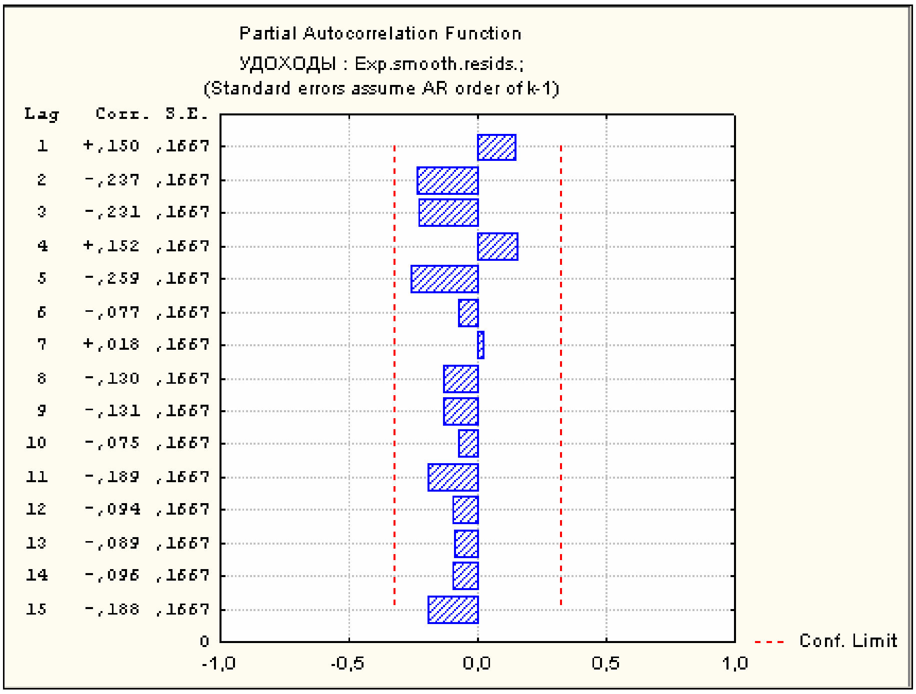


Рисунок 5.17 – Оценка частной автокорреляционной функции остатков

На уровне значимости 0,05 можно принять нулевую гипотезу о том, что остатки некоррелированы. Так как остатки нормально распределены и некоррелированы, то можно переходить к прогнозированию.

Вернемся к окну **Exponential Smoothing/Экспоненциальное сглаживание**. В опции **Forecast/ Прогнозирование** устанавливается период упреждения, в данном случае период упреждения – 2 года или 8 кварталов. График прогнозных значений можно получить, нажав на кнопку **Summary. Exponential Smoothing**.

На рисунке 5.18 представлен прогноз исходного временного ряда на 2 года вперед.

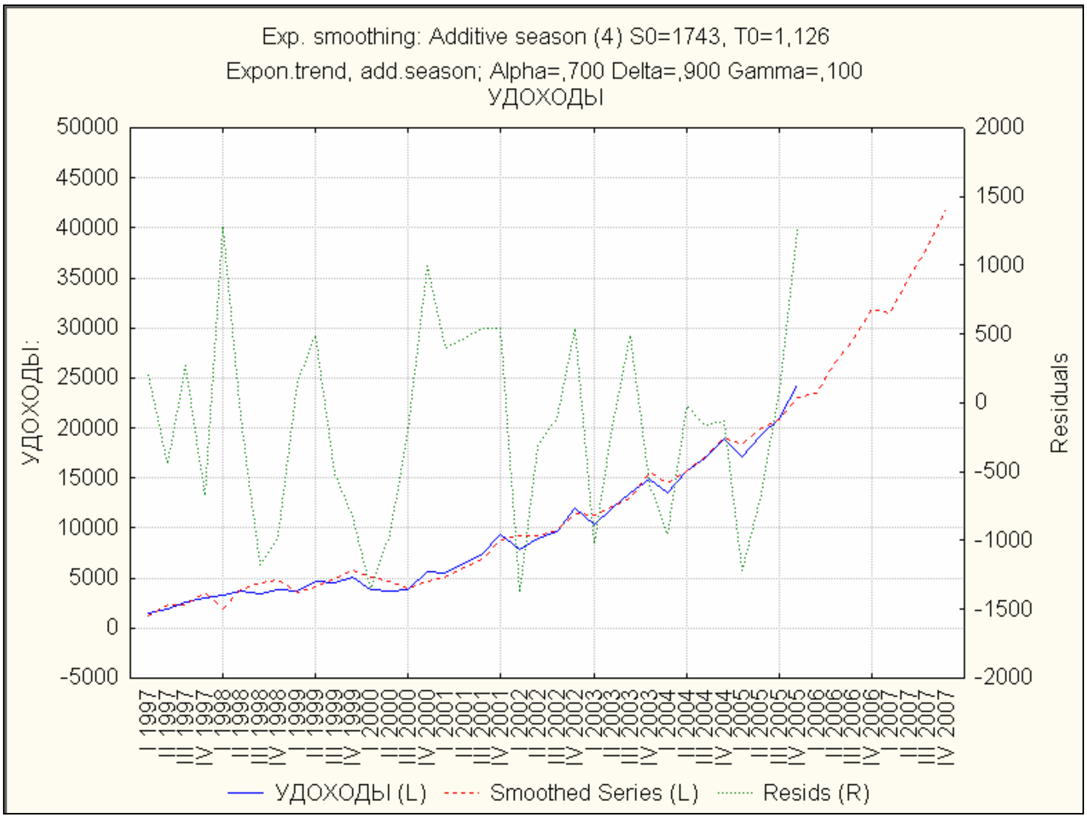


Рисунок 5.18 – Результат прогнозирования методом экспоненциального сглаживания

Оценка модели экспоненциального сглаживания с мультипликативным ростом и аддитивным сезонным эффектом:

$$\hat{y}_t = \hat{f}_t + \hat{g}_t$$

$$\hat{f}_t = 0.7(y_t - \hat{g}_{t-4}) + 0.3\hat{f}_{t-1} * \hat{r}_{t-1}$$

$$\hat{g}_t = 0.9(y_t - \hat{f}_t) + 0.1\hat{g}_{t-4}$$

$$\hat{r}_t = 0.1\frac{\hat{f}_t}{\hat{f}_{t-1}} + 0.9\hat{r}_{t-1}$$

Согласно прогнозу в четвертом квартале 2007 г. среднедушевые денежные доходы населения Оренбургской области составят 41878,69 рублей.

6 Системы одновременных регрессионных уравнений

6.1 Основные понятия системы одновременных регрессионных уравнений. Косвенный метод наименьших квадратов

Эконометрическая модель в виде системы регрессионных уравнений, примером которой может служить модель (2.1), наиболее типична, поскольку позволяет отразить многообразие связей между показателями, характеризующими изучаемое явление (процесс). Общий вид такой модели, называемой системой одновременных регрессионных уравнений в структурной форме [1]

$$\begin{cases} \beta_{11}y_{1t}+\beta_{12}y_{2t}+...+\beta_{1m}y_{mt}+C_{11}x_{1t}+C_{12}x_{2t}+...+C_{1k}x_{kt}=\varepsilon_{1t} \\ \beta_{21}y_{1t}+\beta_{22}y_{2t}+...+\beta_{2m}y_{mt}+C_{21}x_{1t}+C_{22}x_{2t}+...+C_{2k}x_{kt}=\varepsilon_{2t} \\ .... \\ \beta_{m1}y_{1t}+\beta_{m2}y_{2t}+...+\beta_{mm}y_{mt}+C_{m1}x_{1t}+C_{m2}x_{2t}+...+C_{mk}x_{kt}=\varepsilon_{mt} \end{cases} \quad (6.1)$$

$t=1,2...n$,

Матричная форма записи СОУ:

$$BY_t+CX_t=\varepsilon_t,\ t=1,n, \quad (6.2)$$

где: $y_t=(y_{1t},...y_{mt})^T$ - вектор эндогенных переменных, измеренных в момент времени t;

$x_t=(x_{1t},...x_{kt})^T$ - вектор предопределенных переменных, отнесенных к моменту времени t;

$B=\begin{pmatrix} \beta_{11}...\beta_{1m} \\ .... \\ \beta_{m1}...\beta_{mm} \end{pmatrix}$ - матрица неизвестных коэффициентов при эндогенных переменных структурной формы СОУ, причем $\det B\neq 0$;

$C=\begin{pmatrix} C_{11}...C_{1k} \\ .... \\ C_{m1}...C_{mk} \end{pmatrix}$ - матрица неизвестных коэффициентов при предопределенных переменных структурной формы СОУ;

$\varepsilon_t=(\varepsilon_{1t}...\varepsilon_{mt})^T$ - вектор регрессионных остатков в момент времени t компоненты которого некоррелированы между собой и для разных t, гомоскедастичны при каждом t.

Не нарушая общности можно считать β_{ii}=1, что позволяет каждое уравнение СОУ представить в виде

$$y_{it} = -\beta_{i1} y_{1t} \ldots - \beta_{i\,i-1} y_{i-1t} - \beta_{i\,i+1} y_{i+1\,t} - \ldots - \beta_{im} y_{mt} -$$
$$- C_{i1} x_{1t} - \ldots - C_{ik} x_{kt} + \varepsilon_{it}, \quad t = \overline{1,n} \tag{6.3}$$

Выражение (6.3) – стандартная форма линейной модели множественной регрессии относительно y_i, но в ее правой части присутствуют стохастические регрессоры $y_{1,t}$,…, в этом случае известно, что оценки неизвестных коэффициентов могут оказаться смещенными.

Учитывая невырожденность матрицы B, мы можем перейти от (6.2) к так называемой приведенной форме СОУ:

$$Y_t = -B^{-1} C x_t + B^{-1} \varepsilon_t = \Pi x_t + \delta_t, \tag{6.4}$$

где: $\Pi = \{\Pi_{ij}\}_{\substack{i=1,m \\ j=1,k}} \equiv -B^{-1} C$ - матрица приведенной формы СОУ.

$\delta_t = B^{-1} \varepsilon_t$ - вектор регрессионных остатков приведенной формы, кстати, уже не являющейся белым шумом.

Оценив построчно матрицу приведенной формы из ЛММР,

$$y_{it} = \Pi_{i1} x_{1t} + \Pi_{i2} x_{2t} + \ldots + \Pi_{ik} x_{kt} + \delta_{it}, \; t = \overline{1,n}; \; i = \overline{1,m}; \tag{6.5}$$

можно поставить задачу восстановления коэффициентов структурной формы из системы

$$B \hat{\Pi} = -C \tag{6.6}$$

или

$$\begin{array}{l}
\beta_{11} \hat{\Pi}_{11} + \beta_{12} \hat{\Pi}_{21} + \ldots + \beta_{1m} \hat{\Pi}_{m1} = -C_{11} \\
\beta_{21} \hat{\Pi}_{11} + \beta_{22} \hat{\Pi}_{21} + \ldots + \beta_{2m} \hat{\Pi}_{m1} = -C_{21} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
\beta_{m1} \hat{\Pi}_{11} + \beta_{m2} \hat{\Pi}_{21} + \ldots + \beta_{mm} \hat{\Pi}_{m1} = -C_{m1} \\
\beta_{11} \hat{\Pi}_{22} + \beta_{12} \hat{\Pi}_{22} + \ldots + \beta_{1m} \hat{\Pi}_{m2} = -C_{12} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
\beta_{m1} \hat{\Pi}_{12} + \beta_{m2} \hat{\Pi}_{22} + \ldots + \beta_{mm} \hat{\Pi}_{m2} = -C_{m2} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
\beta_{11} \hat{\Pi}_{1k} + \beta_{12} \hat{\Pi}_{2k} + \ldots + \beta_{1m} \hat{\Pi}_{mk} = -C_{1k} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
\beta_{m1} \hat{\Pi}_{1k} + \beta_{12} \hat{\Pi}_{2k} + \ldots + \beta_{mm} \hat{\Pi}_{mk} = -C_{mk}
\end{array} \tag{6.6a}$$

Система линейных, относительно элементов матриц B и C уравнений содержит m^2+mk неизвестных и состоит из mk уравнений. В общем случае такая система не разрешима. Однако, практика эконометрического моделирования показывает, что матрица B и C имеют достаточно разреженную структуру (большое количество нулевых элементов) и поэтому возможны ситуации, в которых система (6.6а) разрешима [1]. Рассмотрим на примерах возможные ситуации.

Модель, описывающая зависимость спроса и предложения некоторого товара от его цены и дохода в условиях равновесия:

$$\begin{aligned} y_{1t} &= \alpha_2 y_{2t} + \varepsilon_{1t} \text{ - предложение,} \\ y_{1t} &= \beta_2 y_{2t} + \beta_3 x_{1t} + \varepsilon_{2t} \text{ - спрос,} \end{aligned} \tag{6.7}$$

где: y_{1t} – спрос(предложение); y_{2t} – цена – эндогенные переменные; x_{1t} – доход – экзогенные переменные.

Перейдем от (6.7) к СОУ в приведенной форме, т.е. разрешим систему относительно эндогенных переменных y_{1t} и y_{2t}:

$$\begin{aligned} y_{1t} &= \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} x_{1t} + \frac{\alpha_2 \varepsilon_{2t} - \beta_2 \varepsilon_{1t}}{\alpha_2 - \beta_2}, \\ y_{1t} &= \frac{\beta_3}{\alpha_2 - \beta_2} x_{1t} + \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\alpha_2 - \beta_2}. \end{aligned} \tag{6.8}$$

Перепишем (6.8) в виде

$$\begin{aligned} y_{1t} &= \Pi_{11} x_{1t} + \delta_{1t}, \\ y_{2t} &= \Pi_{21} x_{1t} + \delta_{2t}, \end{aligned}$$

где:

$$\Pi_{11} = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2},\ \Pi_{21} = \frac{\beta_3}{\alpha_2 - \beta_2},\ \delta_{1t} = \frac{\alpha_2 \varepsilon_{2t} - \beta_2 \varepsilon_{1t}}{\alpha_2 - \beta_2},\ \delta_{2t} = \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\alpha_2 - \beta_2}, \tag{6.9}$$

x_{1t} некоррелирована с δ_{1t} и δ_{2t}, если только x_{1t} некоррелирована с регрессионными остатками исходной модели.

Оценки $\hat{\Pi}_{11}, \hat{\Pi}_{21}$, полученные методом наименьших квадратов имеют вид:

$$\hat{\Pi}_{11} = \frac{\sum_{t=1}^{n} y_{1t} x_{1t}}{\sum_{t=1}^{n} x_{1t}^2};\ \hat{\Pi}_{21} = \frac{\sum_{t=1}^{n} y_{2t} x_{1t}}{\sum_{t=1}^{n} x_{1t}^2}; \tag{6.10}$$

Оценки (6.10) являются состоятельными для Π_{11}, Π_{21}, но, как видно из (6.9), $\alpha_2 = \dfrac{\Pi_{11}}{\Pi_{21}}$. Поэтому, в силу теоремы Слуцкого, оценка $\hat{\alpha}_2 = \dfrac{\hat{\Pi}_{11}}{\hat{\Pi}_{21}}$ будет состоятельной оценкой структурной формы коэффициента α_2. В то же время по оценкам $\hat{\Pi}_{11}, \hat{\Pi}_{21}$ не предоставляется возможным найти оценки коэффициентов β_2 и β_3 структурной формы.

Определение 1. Уравнение структурной формы СОУ называется точно идентифицируемым, если его коэффициенты однозначно определяются по оценкам коэффициентов приведенной формы СОУ [1].

Определение 2. Уравнение структурной формы СОУ называется неидентифицируемым, если его коэффициенты нельзя определить по оценкам коэффициентов приведенной формы СОУ [1].

Таким образом, в приведенном выше примере, первое уравнение СОУ является точно идентифицируемым, а второе неидентифицируемым.

Рассмотрим более сложную модель, описывающую предложение и спрос в условиях равновесия, включив в модель для спроса процентную ставку x_{2t}. В итоге модель спроса-предложения имеет вид:

$$\begin{aligned} y_{1t} &= \alpha_2 y_{2t} + \varepsilon_{1t} \\ y_{1t} &= \beta_2 y_{2t} + \beta_3 x_{1t} + \beta_4 x_{2t} + \varepsilon_{2t} \end{aligned} \tag{6.11}$$

Приведенная форма этой системы одновременных уравнений:

$$\begin{aligned} y_{1t} &= \Pi_{11} x_{1t} + \Pi_{12} x_{2t} + \delta_{1t}, \\ y_{2t} &= \Pi_{21} x_{1t} + \Pi_{22} x_{2t} + \delta_{2t}, \end{aligned}$$

где:

$$\begin{aligned} &\Pi_{11} = \frac{\alpha_2\beta_3}{\alpha_2 - \beta_2},\ \Pi_{12} = \frac{\alpha_2\beta_4}{\alpha_2 - \beta_2},\ \Pi_{21} = \frac{\beta_3}{\alpha_2 - \beta_2},\ \Pi_{22} = \frac{\beta_4}{\alpha_2 - \beta_2} \\ &\delta_{1t} = \frac{\alpha_2\varepsilon_{2t} - \beta_2\varepsilon_{1t}}{\alpha_2 - \beta_2},\ \delta_{2t} = \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\alpha_2 - \beta_2}. \end{aligned} \tag{6.12}$$

Очевидно, что $\alpha_2 = \frac{\Pi_{11}}{\Pi_{21}}$ и $\alpha_2 = \frac{\Pi_{12}}{\Pi_{22}}$, но тогда оценив предварительно $\hat{\Pi}_{11}, \hat{\Pi}_{12}$ из первого уравнения (6.12) и $\hat{\Pi}_{21}, \hat{\Pi}_{22}$ из второго уравнения (6.12), найдем $\hat{\alpha}_2$: $\hat{\alpha}_2 = \frac{\hat{\Pi}_{11}}{\hat{\Pi}_{21}}$ и $\hat{\alpha}_2 = \frac{\hat{\Pi}_{12}}{\hat{\Pi}_{22}}$.

Определение 3. Уравнение структурной формы СОУ называется сверхидентифицируемым, если его коэффициенты оцениваются по оценкам коэффициентов приведенной формы СОУ не единственным образом [1].

Так, в предыдущем примере, первое уравнение структурной формы сверхидентифицируемо, а второе неидентифицируемо.

Определение 4. Метод оценивания коэффициентов точно – идентифицируемого уравнения структурной формы, по оценкам коэффициентов приведенной формы, называется косвенным методом наименьших квадратов [1].

В частности, оценка коэффициента α_2 в первом уравнении модели (6.9) получена косвенным методом наименьших квадратов.

Используя необходимые и достаточные условия идентифицируемости уравнения структурной формы [1], можно выявить точно идентифицируемые уравнения и затем оценить их коэффициенты косвенным методом наименьших квадратов, но это лишь частный подход к оцениванию коэффициентов структурной формы. Поэтому остановимся на общем подходе к оцениванию коэффициентов, так называемых рекурсивных систем и общем методе оценки коэффициентов уравнения СОУ – двухшаговом методе наименьших квадратов.

6.2 Оценка (идентификация) коэффициентов системы одновременных уравнений рекурсивного вида

Определение 5. Говорят, что имеется СОУ рекурсивного вида, если первое уравнение содержит только одну эндогенную переменную (будем считать это y_{1t}), а остальные предопределенные. Второе уравнение может содержать не более двух эндогенных переменных (y_{2t} и возможно y_{1t}) и т.д. При этом число, не включенных в i-е уравнение структурной формы, предопределенных переменных не меньше числа включенных в него эндогенных переменных без единицы $(k - k_i) \geq m_i - 1$ [1].

Рекурсивная СОУ имеет, таким образом, вид

$$BY_t + CX_t = \delta_t, \tag{6.13}$$

где B- нижняя треугольная матрица с «1» на главной диагонали.

Если регрессионные остатки δ_{it} и δ_{jt} некоррелированные для всех t и для всех $i \neq j$ (такая СОУ называется часто рекурсивной), то коэффициенты структурной формы оцененные последовательно в каждом уравнении с помощью метода наименьших квадратов $\left(\hat{\beta}^{(i)^T}, \hat{C}^{(i)^T}\right)^T = \left(Z^T Z\right)^{-1} Z^T Y^{(i)}$ будут состоятельным, а при нормальности δ_{jt} –асимптотически эффективными. В данном слечае $\beta^{(i)}$- вектор столбец коэффициентов при эндогенных переменных в правой части i-го уравнения структурной формы, разрешенного относительно y_{it}; $C^{(i)}$ – вектор столбец коэффициентов при экзогенных переменных в этом же уравнении; Z – матрица из векторов столбцов наблюденных значений эндогенных и предопределенных переменных стоящих в правой части i-го уравнения и $Y^{(i)}$ – вектор наблюденных значений y_i.

Например, для модели рекурсивной СОУ:

$$\begin{matrix} y_{1t} = \alpha_1 x_{2t} + \alpha_2 x_{2t} + \varepsilon_{1t}, \\ y_{1t} = \beta_1 y_{1t} + \beta_2 x_{3t} + \varepsilon_{2t}, \end{matrix} \qquad \begin{matrix} M\left(\varepsilon_{1t}\,\varepsilon_{2t}\right) = 0, \ \forall t \\ M\left(\varepsilon_{1t}\,\varepsilon_{2\tau}\right) = \begin{cases} 0, \ t \neq \tau \\ \sigma_i^2, \ t = \tau, \ i = 1,2... \end{cases} \end{matrix} \tag{6.14}$$

МНК- оценки коэффициентов первого уравнения (6.14):

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \left(Z^{(1)} \cdot Z^{(1)T}\right)^{-1} \left(Z^{(1)} \cdot Y^{(1)}\right), \qquad (6.15)$$

где $Z^{(1)} = \left(X^{(1)} X^{(2)}\right)$, $X^{(1)} = \begin{pmatrix} x_{11} \\ x_{12} \\ \dots \\ x_{1n} \end{pmatrix}$, $X^{(2)} = \begin{pmatrix} x_{21} \\ x_{22} \\ \dots \\ x_{2n} \end{pmatrix}$, $Y^{(1)} = \begin{pmatrix} y_{11} \\ y_{12} \\ \dots \\ y_{1n} \end{pmatrix}$.

МНК- оценки коэффициентов второго уравнения (6.14) имеют вид:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(Z^{(2)} \cdot Z^{(2)T}\right)^{-1} \left(Z^{(2)} \cdot Y^{(2)}\right), \qquad (6.16)$$

где $Z^{(2)} = \left(Y^{(1)} X^{(3)}\right)$, $Y^{(1)} = \begin{pmatrix} y_{11} \\ y_{12} \\ \dots \\ y_{1n} \end{pmatrix}$, $X^{(3)} = \begin{pmatrix} x_{31} \\ x_{32} \\ \dots \\ x_{3n} \end{pmatrix}$, $Y^{(2)} = \begin{pmatrix} y_{21} \\ y_{22} \\ \dots \\ y_{2n} \end{pmatrix}$.

6.3 Двухшаговый метод наименьших квадратов

Представим *i*-е уравнение СОУ в виде:

$$y_{it} = -\beta_{ij} y_{j_1,t} - \dots - \beta_{ij_{m_{i-1}}} y_{j_{m_{i-1}},t} - C_{il_1} x_{l_1 t} - \dots - C_{il_{ki}} x_{l_{ki} t} + \varepsilon_{it}, \; t = \overline{1,n}. \qquad (6.17)$$

Алгоритм двухшагового метода наименьших квадратов:

1) оцениваем функции регрессии каждой из эндогенных переменных $y_{j_1,t}, y_{j_2,t}, \dots y_{j_{m_{i-1}},t}$ на все предопределенные переменные, т.е.

$\hat{y}_{j_s,t} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{1t} + \dots + \hat{\gamma}_k x_{kt}, \;\; j_s = j_1, j_2, \dots j_{m_{i-1}}$

2) находим оценки модельных значений $\hat{Y}^{(j_s)} = \left(\hat{y}_{j_s,1} \dots \hat{y}_{j_s,n}\right)^T$

3) находим оценку $\Theta = \left(-\beta^{(i)T}, -C^{(i)T}\right)^T$ из модели не содержащей стохастических регрессоров

$Y^{(i)} = Z\Theta + \varepsilon$,

где $Z = \left(\hat{Y}^{(j_1)}, \hat{Y}^{(j_2)} \dots \hat{Y}^{(j_{m_{i-1}})} X^{(l_i)} \dots X^{(l_{k_i})}\right)$.

6.4 Тесты для самоконтроля

1) Переменные, задаваемые «извне», автономно, в определенной степени управляемые (планируемые) называются

а) экзогенными;

б) эндогенными;
в) лаговыми;
г) предопределенными.

2) Переменные, значения которых формируются в процессе функционирования анализируемой социально-экономической системы, называются
а) экзогенными;
б) эндогенными;
в) лаговыми;
г) предопределенными.

3) Множество всех экзогенных и лаговых эндогенных переменных называется
а) предопределенными переменными;
б) зависимыми переменными;
в) лаговыми переменными;
г) фиктивными переменными.

4) Система уравнений, описывающая соотношения между спросом и предложением, $\begin{cases} q_t = \alpha_0 + \alpha_1 p_t + \varepsilon_t \\ q_t = \beta_0 + \beta_1 p_t + \beta_2 x_t + \delta_t \end{cases}$, где q_t - объем предложения (спроса) в период t, который совпадает со спросом (предложением) на товар в условиях равновесия; p_t - средняя цена за единицу товара, зафиксированная на рынке в период t; x_t - среднедушевой доход в период t, ε_t и δ_t - регрессионные остатки соответствующих уравнений, представляет собой
а) рекурсивную систему регрессионных уравнений;
б) структурную форму системы одновременных уравнений;
в) приведенную форму системы одновременных уравнений;
г) систему алгебраических уравнений.

5) Уравнение структурной формы системы одновременных уравнений, для которого все участвующие в нем неизвестные коэффициенты однозначно восстанавливаются (определяются) по коэффициентам приведенной формы называется
а) сверх идентифицируемым;
б) точно идентифицируемым;
в) неидентифицируемым;
г) линейным.

6) Уравнение структурной формы системы одновременных уравнений, для которого все участвующие в нем неизвестные коэффициенты восстанавливаются (определяются) по коэффициентам приведенной формы, причем некоторые из них могут принимать одновременно несколько числовых значений, соответствующих одной и той же приведенной форме называется

а) сверх идентифицируемым;
б) точно идентифицируемым;
в) неидентифицируемым;
г) линейным.

7) Уравнение структурной формы системы одновременных уравнений, для которого хотя бы один из участвующих в нем неизвестных коэффициентов не может быть восстановлен по коэффициентам приведенной формы называется
а) сверх идентифицируемым;
б) точно идентифицируемым;
в) неидентифицируемым;
г) линейным.

8) Косвенный метод наименьших квадратов возможно применять для оценки коэффициентов
а) сверх идентифицируемого уравнения системы;
б) точно идентифицируемого уравнения системы;
в) неидентифицируемого уравнения системы;
г) алгебраического уравнения.

9) Для оценивания коэффициентов сверх идентифицируемого уравнения, в случае некоррелированности регрессионных остатков разных уравнений системы одновременных уравнений, рекомендуется применять
а) косвенный метод наименьших квадратов;
б) двухшаговый метод наименьших квадратов;
в) двухшаговый и трехшаговый метод наименьших квадратов;
г) метод наименьших квадратов.

10) Связь между коэффициентами структурной ($B_{m\times m}, C_{m\times k}$ - матрицы неизвестных коэффициентов при эндогенных и предопределенных переменных) и приведенной формы ($\Pi_{m\times к}$ - матрица неизвестных коэффициентов) системы одновременных уравнений имеет вид
а) $\Pi = -B^{-1} \cdot C$;
б) $\Pi = B^{-1} \cdot C$;
в) $\Pi \cdot B = C$;
г) $B \cdot C = -\Pi$.

11) При реализации двухшагового метода наименьших квадратов наблюдаемые значения стохастических регрессоров заменяются
а) наблюдаемыми значениями эндогенных переменных;
б) наблюдаемыми значениями предопределенных переменных;
в) наблюдаемыми значениями функции регрессии эндогенных переменных на все преопределенные переменные;

г) наблюдаемыми значениями зависимых переменных.

12) Система регрессионных уравнений, имеющая следующий вид:

$$\begin{cases} y_{1t} = c_{11}x_{1t} + c_{12}x_{2t} + ... + c_{1k}x_{kt} + \varepsilon_{1t} \\ y_{2t} = \beta_{21}y_{1t} + c_{21}x_{1t} + c_{22}x_{2t} + ... + c_{2k}x_{kt} + \varepsilon_{2t} \\ y_{3t} = \beta_{31}y_{1t} + \beta_{32}y_{2t} + c_{31}x_{1t} + c_{32}x_{2t} + ... + c_{3k}x_{kt} + \varepsilon_{3t} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_{mt} = \beta_{m1}y_{1t} + \beta_{m2}y_{2t} + \beta_{mm-1}y_{m-1t} + c_{m1}x_{1t} + c_{m2}x_{2t} + ... + c_{mk}x_{kt} + \varepsilon_{mt} \end{cases},$$

где $y_{1t},...,y_{mt}$ – эндогенные переменные, $x_{1t},...,x_{kt}$ – предопределенные переменные, $\beta_{ij}, c_{is} - i = 1,m;\ j = 1,m;\ s = 1,k$ – неизвестные коэффициенты системы, ε_{it} - регрессионные остатки i-го уравнения, называется

а) системой внешне не связанных регрессионных уравнений;

б) рекурсивной системой регрессионных уравнений;

в) системой алгебраических уравнений.

13) Для системы одновременных уравнений $\begin{cases} y_{1t} = \beta_{12}y_{2t} + c_{11}x_{1t} + \varepsilon_{1t} \\ y_{2t} = \beta_{21}y_{1t} + c_{22}x_{2t} + \varepsilon_{2t} \end{cases}$, (где y_{1t} - производительность труда на предприятии в период t; y_{2t} - средняя оплата труда на предприятии в период t; x_{1t} - энерговооруженность в период t; x_{2t} - средний по предприятию размер тарифной ставки в период t) оценена матрица коэффициентов приведенной формы $\hat{\Pi} = \begin{bmatrix} 1{,}0388 & -0{,}4773 \\ 0{,}1070 & 1{,}9700 \end{bmatrix}$. Необходимое и достаточное (ранговое) условие идентифицируемости первого и второго уравнения имеет вид

а) $\begin{matrix} rang\hat{\Pi}_x(1) > m_1 - 1;\ rang\hat{\Pi}_x(2) > m_2 - 1 \\ где\ m_1 - 1 = 1;\ m_2 - 1 = 1 \end{matrix}$,

б) $\begin{matrix} rang\hat{\Pi}_x(1) > m_1 - 1;\ rang\hat{\Pi}_x(2) < m_2 - 1 \\ где\ m_1 - 1 = 1;\ m_2 - 1 = 1 \end{matrix}$

в) $\begin{matrix} rang\hat{\Pi}_x(1) = m_1 - 1;\ rang\hat{\Pi}_x(2) = m_2 - 1 \\ где\ m_1 - 1 = 1;\ m_2 - 1 = 1 \end{matrix}$

г) $\begin{matrix} rang\hat{\Pi}_x(1) > m_1 - 1;\ rang\hat{\Pi}_x(2) = m_2 - 1 \\ где\ m_1 - 1 = 1;\ m_2 - 1 = 1 \end{matrix}$

14) Для системы одновременных уравнений $\begin{cases} y_{1t} = \beta_{12}y_{2t} + c_{11}x_{1t} + c_{12}x_{2t} + \varepsilon_{1t} \\ y_{2t} = \beta_{21}y_{1t} + c_{23}x_{3t} + \varepsilon_{2t} \end{cases}$, (где y_{1t} - валовой региональный продукт в период t; y_{2t} - розничный товарооборот в период t; x_{1t} - объем основных фондов в период t; x_{2t} - объем промышленности в период t; x_{3t} - средний душевой доход в период t) получена оценка приведенной формы

$\begin{cases} \hat{y}_1 = 0{,}0716x_1 + 0{,}376x_2 + 0{,}00494x_3 \\ \hat{y}_2 = 0{,}0488x_1 + 0{,}0406x_2 + 0{,}00434x_3 \end{cases}$. Проверка порядкового и рангового условий идентифицируемости, позволяет сделать вывод, что первое уравнение структурной формы

а) сверх идентифицируемое;

б) точно идентифицируемое;

в) неидентифицируемое;

г) нелинейное.

.

15) По 63 наблюдениям для системы одновременных уравнений $\begin{cases} y_{1t} = \beta_{12} y_{2t} + \varepsilon_{1t} \\ y_{2t} = \beta_{21} y_{1t} + c_{21} x_{1t} + \varepsilon_{2t} \\ y_{3t} = y_{2t} + x_{2t} \end{cases}$ вычислены оценки коэффициентов приведенной формы: $\begin{cases} \hat{y}_1 = \underset{(1,5)}{6{,}23}\, x_1 \\ \hat{y}_2 = \underset{(2,52)}{10{,}5}\, x_2 \\ \hat{y}_3 = \underset{(1,35)}{12{,}5}\, x_1 + \underset{(3,65)}{8{,}2}\, x_2 \end{cases}$. Матрица приведенной формы имеет вид

а) $\hat{\Pi} = \begin{bmatrix} 6{,}23 & 0{,}00 \\ 0{,}00 & 10{,}5 \\ 12{,}5 & 8{,}2 \end{bmatrix}$;

б) $\hat{\Pi} = \begin{bmatrix} 6{,}23 & 0{,}67 \\ -4{,}12 & 10{,}5 \\ 12{,}5 & 8{,}2 \end{bmatrix}$;

в) $\hat{\Pi} = \begin{bmatrix} 0{,}00 & 0{,}67 \\ -4{,}12 & 0{,}00 \\ 12{,}5 & 8{,}2 \end{bmatrix}$;

г) $\hat{\Pi} = \begin{bmatrix} 6{,}23 & 0{,}67 \\ 12{,}5 & 8{,}2 \end{bmatrix}$.

7 Ответы к тестовым заданиям

№ задания	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ответы к разделу №1	б	а	в	в	в	В	в	г	а	г	а	а	б	б	в
Ответы к разделу №2	а	а	г	а	б	Б	б	в	а	в	в	б	в	а	г
Ответы к разделу №3	г	а	в	в	г	В	б	г	в	г	в	б	-	-	-
Ответы к разделу №4	а	в	б	в	г	В	б	в	г	в	а	б	а	г	-
Ответы к разделу №5	в	в	г	б	в	А	в	а	в	г	б	б	б	г	б
Ответы к разделу №6	а	б	а	б	б	А	в	б	б	а	в	б	в	б	а

Список использованных источников

1 **Айвазян, С.А.** Прикладная статистика. Основа эконометрики: в 2т.: учеб. для вузов / С.А. Айвазян, В.С. Мхитарян. - М.: ЮНИТИ-ДАНА, 2001. - Т.2. – 432 с. – ISBN 5-238-00304-82.

2 **Доугерти, Кр.** Введение в эконометрику: учебник для вузов/ Кр. Доугерти. - М.: МГУ, ИНФРА-М, 2004. – 402 с. – ISBN 5-16-001463-2.

3 **Магнус, Я.Р.** Эконометрика. Начальный курс: учеб. / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 576 с. – ISBN 5-7779-0055-X.

4 Эконометрика: учебник / под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2007. – 576 с. – ISBN 978-5-279-02786-6.

5 **Реннер, А.Г.** Математическая статистика: уч. пособие / А.Г. Реннер, Г.Г. Аралбаева. – Оренбург: ОГУ, 2003. – 175 с.

6 **Тихомиров, Н.П.** Эконометрика: учеб. для вузов / Н.П. Тихомиров, Е.Ю. Дорохина. - М.: Экзамен, 2003. - 512 с. – ISBN 5-94692-438-9.

7 **Домбровский, В.В.** Эконометрика: учебник/ В.В.Домбровский. – М.: Нов. учебник, 2004. – 342 с. – ISBN 5-8393-0400-X.

8 **Новак, Эдвард.** Введение в методы эконометрики: сб. задач: пер. с польск.; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2004. – 248с. – ISBN 5-279-02927-0.

9 **Катышев, П.К.** Сборник задач к начальному курсу эконометрики / П.К. Катышев, А.А Пересецкий. – М.: Дело, 2003. – 208 с. – ISBN 5-7749-0137-8.

10 Сборник задач по эконометрике: учебное пособие для студентов экономических вузов / Е.Ю. Дорохина, Л.Ф. Преснякова, Н.П. Тихомиров – М.: Издательство "Экзамен", 2003. – 224 с. – ISBN 5-94692-206-8.

11 Практикум по эконометрике: учебное пособие / под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2006. – 344 с.

12 **Айвазян, С.А.** Прикладная статистика в задачах и упражнениях: учеб. для вузов / С.А. Айвазян, В.С. Мхитарян. - М.: ЮНИТИ-ДАНА, 2001. -270 с.- ISBN 5-238-00303-X.

13 **Замков, О.О.** Математические методы в экономике: учебник/ О.О. Замков, А.В. Толстопятенко, Ю.Н. Черемных. – М.: Дело и Сервис, 2001. – 368 с. – ISBN 5-86509-054-2.

14 **Кулинич, Е.И.** Эконометрия / Е.И. Кулинич. - М.: Финансы и статистика, 2001. - 304 с. – ISBN 5-279-02090-7.

15 **Сошникова, Л.А.** Многомерный статистический анализ в экономике: учебное пособие для вузов / Л.А. Сошникова; под ред. проф. В.Н. Тамашевича. - М.: ЮНИТИ-ДАНА, 1999. - 598 с. – ISBN 5-238-00099-5.

16 **Реннер, А.Г.** Математическая статистика: учебное пособие / А.Г. Реннер, Г.Г. Аралбаева. – Оренбург: ГОУ ВПО ОГУ, 2002. – 175 с.

17 **Ниворожкина, Л.И.** Многомерные статистические методы в экономике: учебник / Л.И. Ниворожкина, С.В. Арженовский. – М.: Издательско-торговая корпорация «Дашков и К», 2008. – 224 с. – ISBN 978-5-91131-565-8.

18 **Просветов, Г.И.** Эконометрика: задачи и решения: учебно-практическое пособие / Г.И. Просветов. – М.: Издательство «Альфа-Пресс», 2008. – 192 с. – ISBN 978-5-94280-349-0.
19 **Орлов, А.И.** Эконометрика: учебник для вузов / А.И. Орлов. – М.: Изд-во «Экзамен», 2003. – 576 с. – ISBN 5-94692-452-4.
20 **Бывшев, В.А.** Эконометрика: учебн. пособие для студентов, обучающихся по специальностям "Финансы и кредит", "Бухгалтерский учет, анализ и аудит", "Мировая экономика", "Налоги и налогообложение"/ В.А. Бывшев. - Москва: Финансы и статистика, 2008.- 480 с. – ISBN 978-5-279-03274-7.
21 **Гладилин, А.В.** Эконометрика: учеб. пособие/ А. В. Гладилин, А. Н. Герасимов, Е. И. Громов. - М.: КноРус, 2006. -232 с. – ISBN 5-85971-118-2.
22 **Кремер, Н.Ш.** Эконометрика: учеб. для вузов/ Н.Ш. Кремер, Б.А. Путко. - М.: ЮНИТИ-ДАНА, 2003. - 311 с. – ISBN 5-238-00333-1.
23 **Дуброва, Т.А.** Статистические методы прогнозирования /учебн.-практ. пособие/ Т.А. Дуброва. - М.: МГУ экономики ,статистики и информатики, 1998. -92 с.
24 **Чураков, Е.П.** Математические методы обработки экспериментальных данных в экономике: учебн. пособие. – М.: Финансы и статистика, 2004. – 240с. – ISBN 5-279-02745-6.
25 **Большаков, А.А.** Методы обработки многомерных данных и временных рядов: учеб. пособие для вузов/А. А. Большаков, Р. Н. Каримов. - М.: Горячая линия - Телеком, 2007. - 522 с. – ISBN 5-93517-287-9.
26 **Реннер, А.Г.** Математические методы моделирования социально-экономических процессов (региональный аспект)/ А.Г. Реннер, О.И. Бантикова, О.С. Бравичева, О.И. Стебунова, Л.М. Туктамышева. – Самара: Изд-во СамНЦ РАН, 2008. – 182с. – ISBN 978-5-93424-363-1.
27 **Берндт, Э.Р.** Практика эконометрики: классика и современность: учеб. для вузов: пер. с англ./Э. Р. Берндт; под ред. С. А. Айвазяна.-М.: ЮНИТИ-ДАНА, 2005. - 863 с. – ISBN 5-238-00859-7.
28 **Прокофьев, В.А.** Статистические методы анализа социально-экономического развития административно-территориальных образований/ В.А. Прокофьев, В.А. Динес, Н.Б. Телятников. - Саратовский государственный социально-экономический университет. – Саратов, 2008. – 288 с. – ISBN 978-5-87309-747-0.
29 **Боровиков, В.П.** STATISTICA - Статистический анализ и обработка данных в среде Windows. / В.П. Боровиков, И.П. Боровиков - М.: Информационно-издательский дом "Филинъ", 1998. - 608 с. – ISBN 5-279-01980-1.
30 **Садовникова Н.А., Шмойлова Р.А.** Анализ временных рядов и прогнозирование: метод. указ. и тематика к/р по курсу "Анализ временных рядов и прогнозирование-М.: МЭСИ, 1998. - 22 с.

Приложение А

(обязательное)

Информационная база для реализации индивидуальных заданий

Таблица А.1 – Данные о деятельности сельскохозяйственных предприятий

Y	X_1	X_2	X_3	X_4	X_5
9,7	1,59	0,26	2,05	0,32	0,14
8,4	0,34	0,28	0,46	0,59	0,66
9,0	2,53	0,31	2,46	0,30	0,31
9,9	4,63	0,40	6,44	0,43	0,59
9,6	2,16	0,26	2,16	0,39	0,16
8,6	2,16	0,30	2,69	0,32	0,17
12,5	0,68	0,29	0,73	0,42	0,23
7,6	0,35	0,26	0,42	0,21	0,08
6,9	0,52	0,24	0,49	0,20	0,08
13,5	3,42	0,31	3,02	1,37	0,73
9,7	1,78	0,30	3,19	0,73	0,17
10,7	2,40	0,32	3,30	0,25	0,14
12,1	9,36	0,40	11,51	0,39	0,38
9,7	1,72	0,28	2,26	0,82	0,17
7,0	0,59	0,29	0,60	0,13	0,35
7,2	0,28	0,26	0,30	0,09	0,15
8,2	1,64	0,29	1,41	0,20	0,08
8,4	0,09	0,22	0,05	0,43	0,20
13,1	0,08	0,25	0,03	0,73	0,20
8,7	1,36	0,26	0,17	0,99	0,42

Таблица А.2 – Матрица парных коэффициентов корреляции

Variable	X1	X2	X3	X4	X5
X1	1,0000	,8543	,9778	,1104	,3410
	p= ---	p=,000	p=,000	p=,643	p=,141
X2	,8543	1,0000	,8818	,0269	,4596
	p=,000	p= ---	0	p=,911	p=,041
X3	,9778	,8818	1,0000	,0303	,2784
	0	0	p= ---	p=,899	p=,235
X4	,1104	,0269	,0303	1,0000	,5706
	p=,643	p=,911	p=,899	p= ---	p=,009
X5	,3410	,4596	,2784	,5706	1,0000
	p=,141	p=,041	p=,235	p=,009	p= ---

Таблица А.3 - Расчет критерия Спирмена

Номер семьи	Доход	Расходы на питание	e_i	Ранг X (R_x)	Ранг /e/ ($R_{/e/}$)	$D_i = R_X - R_{/e/}$
1	30	34	-12	1	14	-13
2	36	36,7	-11,7	2	13	-11
3	40	38,4	-5,4	3	8	-5
4	45	40,6	-5,6	4	9	-5
5	50	42,8	-2,8	5	3	2
6	60	47,2	0,8	6	1	5
7	70	51,6	-1,6	7	2	5
8	80	56	-4	8	5.5	2.5
9	85	58,2	-6,2	9	10	-1
10	90	60,4	6,6	10	11	-1
11	92	61,3	13,7	11	16	-5
12	100	64,8	12,2	12	15	-3
13	120	73,6	4,4	13	7	6
14	130	78	4	14	5.5	8.5
15	145	84,6	3,4	15	4	11
16	150	86,8	23,2	16	19	-3
17	200	108,8	16,2	17	17	0
18	250	130,8	-16,8	18	18	0
19	300	152,8	-27,8	19	20	-1
20	360	179,2	9,8	20	12	8

Таблица А.4 – Данные о доходе и потреблении бананов

Номер семьи	Потребление бананов в год (в фунтах)	Семейный доход (в 10000$)
1	1,93	1
2	7,13	2
3	8,78	3
4	9,69	4
5	10,09	5
6	10,42	6
7	10,62	7
8	10,71	8
9	10,79	9
10	11,13	10

Таблица А.5 – Варианты индивидуальных заданий

№ варианта	Результативный признак	Номера факторных признаков X	№ варианта	Результативный признак	Номера факторных признаков X
1	X_{11}	1,2,3,6,9	17	X_6	1,2,7,8,9
2	X_{11}	1,2,4,6,9	18	X_6	1,2,7,9,10
3	X_{11}	1,2,6,7,9	19	X_6	1,3,4,7,9
4	X_{11}	1,2,6,8,9	20	X_6	1,3,7,8,9
5	X_{11}	1,2,6,9,10	21	X_6	1,3,7,9,10
6	X_{11}	1,3,4,6,9	22	X_6	1,4,7,8,9
7	X_{11}	1,3,6,7,9	23	X_6	1,4,7,9,10
8	X_{11}	1,3,6,8,9	24	X_6	1,7,8,9,10
9	X_{11}	1,3,6,9,10	25	X_6	1,2,3,8,9
10	X_{11}	1,4,6,7,9	26	X_6	1,2,4,8,9
11	X_{11}	1,4,6,8,9	27	X_6	1,2,8,9,10
12	X_{11}	1,4,6,9,10	28	X_6	1,3,4,8,9
13	X_{11}	1,6,8,9,10	29	X_6	1,4,8,9,10
14	X_{11}	1,6,7,9,10	30	X_6	1,2,3,4,9
15	X_6	1,2,3,7,9	31	X_6	1,2,3,4,10
16	X_6	1,2,4,7,9	32	X_6	1,3,4,9,10

Таблица А.6 – Показатели уровня жизни населения

Страны	Потребление продуктов на душу населения						Число врачей на 10000ч.	Смерт-ность на 100000 насел.	ВВП по ППС	Расходы на здраво-охр.	Урожай-ность зернов.
	Мясо	Масло животн.	Сахар	Алко-голь	Фрукты	Хлебо-продукты					
	кг	кг	кг	л	Кг	кг			в % США	в % к ВВП	ц/га
	x1	x2	x3	x4	x8	x9	x5	x11	x6	x7	x10
Россия	55	3.9	30	5	28	124	44.5	84.98	20.4	3.2	14.4
Австралия	100	2.6	47	8.2	121	87	32.5	30.58	71.4	8.5	11.6
Австрия	93	5.3	37	12	146	74	33.9	38.42	78.7	9.2	56.1
Азербайджан	20	4.1	12.4	7.9	52	141	38.8	60.34	12.1	3.3	16.4
Армения	20	3.7	4.3	6.5	72	134	34.4	60.22	10.9	3.2	13.5
Белоруссия	72	3.6	28	5.4	38	120	43.6	60.79	20.4	5.4	22.4
Бельгия	85	6.9	48	11	83	72	41	29.82	79.7	8.3	65.5
Болгария	65	3	18	9.5	92	156	36.4	70.57	17.3	5.4	27.8
Великобри-тания	67	3.5	39	8.8	91	91	17.9	34.51	69.7	7.1	62.3
Венгрия	73	1.7	40	10.9	73	106	32.1	64.73	24.5	6	39.8
Германия	88	6.8	35	8.1	138	73	38.1	36.63	76.2	8.6	56.9
Греция	83	1	24	8.8	99	108	41.5	32.84	44.4	5.7	37.4
Грузия	21	3.8	36	9.8	55	140	55	62.64	11.3	3.5	18.6
Дания	98	5	38	10.3	89	77	36.7	34.07	79.2	6.7	54.4
Ирландия	99	3.3	31	9.6	87	102	15.8	39.27	57	6.7	64.2
Испания	89	0.4	26	8.95	103	72	40.9	28.46	54.8	7.3	22.6
Италия	84	2.2	27	9.6	169	118	49.4	30.27	72.1	8.5	46
Казахстан	61	4.2	19.2	7.2	10	191	38.1	69.04	13.4	3.3	7.9
Канада	98	3.1	44	7.4	123	77	27.6	25.42	79.9	10.2	25.4
Киргизия	46	4.1	23.5	6.7	20	134	33.2	53.13	11.2	3.4	17
Нидерланды	86	3,4	37	8,5	176	59	30,1	28,0	72,4	8,7	70,2
Португалия	73	3,2	27	9,7	150	83	28,4	38,79	48,6	7,3	17,6
США	115	1,9	29	8,1	99	103	20,6	32,04	100	14,1	55,2
Финляндия	62	5,8	36	6,8	82	94	33,8	38,58	63,9	8,8	35,9
Франция	91	8,8	36	12,3	84	85	36,7	18,51	77,5	9,8	64,3
Чехия	82	8,2	45	9,4	65	114	32,2	57,62	34,7	1,9	40,2
Япония	40	0,7	20	3,7	60	119	23,1	20,80	83,5	7,3	63,1

Приложение Б

(обязательное)

Информационная база для моделирования стоимости квартир

Таблица Б.1 - Варианты для самостоятельной работы, наименование показателей и исходные данные для эконометрического моделирования стоимости квартир в г. Коврове

№ Варианта	Результативный признак, у	Номер факторных признаков, х
1	2	3
1	1,2	1,2,3,4,5
2	1,2	4,5,6,7,8
3	1,2	5,6,7,8,9
4	1,2	1,3,5,6,7
5	1,2	1,4,6,8,9
6	1,2	1,5,6,7,8
7	1,2	1,6,7,8,9
8	1,2	1,2,6,7,8
9	1,2	2,4,5,7,8
10	1,2	2,5,6,8,9
11	1,2	2,4,6,8,9
12	1,2	3,4,6,7,8
13	1,2	3,5,7,8,9
14	1,3	1,2,3,4,5
15	1,3	4,5,6,7,8
16	1,3	5,6,7,8,9
17	1,3	1,3,5,6,7
18	1,3	1,4,6,8,9
19	1,3	1,5,6,7,8
20	1,3	1,6,7,8,9
21	1,3	1,2,6,7,8
22	1,3	2,4,5,7,8
23	1,3	2,5,6,8,9
24	1,3	2,4,6,8,9
25	1,3	3,4,6,7,8
26	1,3	3,5,7,8,9
27	2,3	1,2,3,4,5
28	2,3	4,5,6,7,8
29	2,3	5,6,7,8,9
30	2,3	1,3,5,6,7
31	2,3	1,4,6,8,9
32	2,3	1,5,6,7,8
33	2,3	1,6,7,8,9
34	2,3	1,2,6,7,8
35	2,3	2,4,5,7,8
36	2,3	2,5,6,8,9
37	2,3	2,4,6,8,9
38	2,3	3,4,6,7,8
39	2,3	3,5,7,8,9

Таблица Б.2 – Наименование показателей

Наименование показателя	Обозначение
1 Дом улучшенной планировки	X_1
Дом «хрущёвка»	
2 Квартира расположенная на одном из промежуточных этажей	X_2
Квартира расположена на первом (последнем) этаже	
3 Дом панельный (блочный)	X_3
Дом кирпичный	
4 Жилая площадь, кв.м	X_4
5 Общая площадь, кв.м	X_5
6 Площадь кухни, кв.м	X_6
7 Квартира «угловая»	X_7
Квартира «неугловая»	
8 В квартире есть балкон (лоджия)	X_8
В квартире нет балкона (лоджии)	
9 Коэффициент зонирования (коэфф)	X_9
10 Стоимость однокомнатной квартиры (тыс.руб)	$У_1$
11 Стоимость двухкомнатной квартиры (тыс.руб)	$У_2$
12 Стоимость трёхкомнатной квартиры (тыс.руб)	$У_3$

Таблица Б.3 - Исходные данные для однокомнатной квартиры

№	x1	x2	x3	x4	x5	x6	x7	x8	x9	y1
1	2	3	4	5	6	7	8	9	10	11
1	Улуч. планировка	Промеж. этаж	Панельный	18,5	33	7,5	Не «угловая»	Есть балкон	0,6	52
2	Улуч. планировка	Первый этаж	Панельный	19	38	9	Не «угловая»	Есть балкон	0,5	50
3	Улуч. планировка	Промеж. этаж	Панельный	20	37	9	Не «угловая»	Нет балкона	0,6	44
4	Улуч. планировка	Первый этаж	Панельный	19,8	38	9	Не «угловая»	Нет балкона	0,4	50
5	Улуч. планировка	Последний этаж	Панельный	19	38	9	Не «угловая»	Нет балкона	0,5	50
6	Улуч. планировка	Первый этаж	Кирпичный	16,6	32	8	«Угловая»	Нет балкона	0,7	45
7	«Хрущёвка»	Последний этаж	Кирпичный	17	32	6	Не «угловая»	Нет балкона	0,1	45
8	«Хрущёвка»	Первый этаж	Панельный	18	31	6	«Угловая»	Есть балкон	0,3	40
9	Улуч. планировка	Промеж. этаж	Панельный	17,5	34	9	Не «угловая»	Есть балкон	0,8	45
10	Улуч. планировка	Последний этаж	Панельный	19,6	39	10	Не «угловая»	Есть балкон	0,4	49
11	Улуч. планировка	Промеж. этаж	Панельный	19,8	37	11	Не «угловая»	Есть балкон	0,6	44
12	Улуч. планировка	Последний этаж	Панельный	19,8	38	9	Не «угловая»	Есть балкон	0,4	50
13	Улуч. планировка	Первый этаж	Панельный	19	38	9	Не «угловая»	Есть балкон	0,5	47
14	Улуч. планировка	Промеж. этаж	Панельный	19	38	9	Не «угловая»	Есть балкон	0,7	48
15	Улуч. планировка	Промеж. этаж	Панельный	19,3	43	9	Не «угловая»	Есть балкон	0,6	50
16	Улуч. планировка	Последний этаж	Панельный	19	37	9	Не «угловая»	Есть балкон	0,5	42
17	Улуч. планировка	Промеж. этаж	Панельный	20	37	9	Не «угловая»	Есть балкон	0,6	44
18	Улуч. планировка	Промеж. этаж	Панельный	20	38	9	Не «угловая»	Есть балкон	0,9	55
19	Улуч. планировка	Первый этаж	Панельный	19,3	43	9	«Угловая»	Есть балкон	0,7	50
20	«Хрущёвка»	Последний этаж	Кирпичный	18	31	6	«Угловая»	Есть балкон	0,4	42
21	Улуч. планировка	Первый этаж	Панельный	19,8	38	9	Не «угловая»	Есть балкон	0,4	43
22	«Хрущёвка»	Промеж. этаж	Кирпичный	16	30	6	Не «угловая»	Есть балкон	0,5	40
23	«Хрущёвка»	Промеж. этаж	Кирпичный	15	33	6	Не «угловая»	Есть балкон	0,6	45
24	«Хрущёвка»	Последний этаж	Кирпичный	18	31	6	«Угловая»	Нет балкона	0,7	35

Продолжение таблицы Б.3

1	2	3	4	5	6	7	8	9	10	11
26	Улуч. планировка	Промеж. этаж	Панельный	19	37	9	Не «угло-вая»	Есть балкон	0,5	53
27	Улуч. планировка	Промеж. этаж	Панельный	19	38	9	Не «угло-вая»	Есть балкон	0,6	48
28	Улуч. планировка	Промеж. этаж	Панельный	19,8	38	9	Не «угло-вая»	Есть балкон	0,4	58
29	Улуч. планировка	Первый этаж	Панельный	19	43	9	Не «угло-вая»	Есть балкон	0,5	50
30	«Хрущёвка»	Последний этаж	Кирпичный	18.янв	34	7	Не «угло-вая»	Нет балкона	0,5	40
31	Улуч. планировка	Первый этаж	Панельный	18	34	7	Не «угло-вая»	Нет балкона	0,8	43
32	«Хрущёвка»	Последний этаж	Кирпичный	14	20	5	«Угловая»	Нет балкона	0,8	32
33	Улуч. планировка	Промеж. этаж	Панельный	20	38	9	Не «угло-вая»	Есть балкон	0,9	53
34	«Хрущёвка»	Промеж. этаж	Кирпичный	19	38	9	Не «угло-вая»	Нет балкона	0,5	52
35	Улуч. планировка	Промеж. этаж	Панельный	17	32	6,5	Не «угло-вая»	Есть балкон	0,7	40
36	«Хрущёвка»	Последний этаж	Кирпичный	18	32	6	«Угловая»	Есть балкон	0,4	44
37	Улуч. планировка	Последний этаж	Панельный	19,1	38	9,3	Не «угло-вая»	Есть балкон	0,5	45
38	Улуч. планировка	Первый этаж	Кирпичный	15	34,7	8	«Угловая»	Есть балкон	0,6	43
39	Улуч. планировка	Последний этаж	Панельный	19,3	39	9	Не «угло-вая»	Есть балкон	0,2	50
40	Улуч. планировка	Промеж. этаж	Панельный	19	39	9	Не «угло-вая»	Есть балкон	0,8	45
41	Улуч. планировка	Промеж. этаж	Кирпичный	18	34	8	Не «угло-вая»	Есть балкон	0,7	45
42	улуч. Планировка	Промеж. этаж	Панельный	18	34	9	Не «угло-вая»	Есть балкон	0,8	45
43	Улуч. планировка	Промеж. этаж	Панельный	20	38	9	Не «угло-вая»	Есть балкон	0,4	58
44	Улуч. планировка	Первый этаж	Панельный	17	32	8	«Угловая»	Есть балкон	0,9	45
45	Улуч. планировка	Последний этаж	Панельный	20	38	9	Не «угло-вая»	Есть балкон	0,6	43
46	Улуч. планировка	Промеж. этаж	Панельный	19	33	8	Не «угло-вая»	Есть балкон	0,6	50
47	Улуч. планировка	Первый этаж	Панельный	19	43	9	Не «угло-вая»	Есть балкон	0,5	50
48	Улуч. планировка	Промеж. этаж	Панельный	20	38	9	Не «угло-вая»	Есть балкон	0,9	55
49	«Хрущёвка»	Последний этаж	Кирпичный	18	32	6	Не «угло-вая»	Нет балкона	0,9	38

Продолжение таблицы Б.3

1	2	3	4	5	6	7	8	9	10	11
50	«Хрущёвка»	Промеж. этаж	Кирпичный	17	32	6	Не «угловая»	Есть балкон	0,6	40
51	Улуч. планировка	Первый этаж	Кирпичный	19	38	9	Не «угловая»	Есть балкон	0,1	52
52	«Хрущёвка»	Последний этаж	Кирпичный	17	32	6	«Угловая»	Есть балкон	0,4	39
53	Улуч. планировка	Промеж. этаж	Панельный	16,5	32	7	Не «угловая»	Есть балкон	0,9	40
54	Улуч. планировка	Первый этаж	Панельный	19,6	36	9,8	Не «угловая»	Есть балкон	0,4	50
55	«Хрущёвка»	Последний этаж	Кирпичный	18	32	6	Не «угловая»	Нет балкона	0,8	40
56	Улуч. планировка	Промеж. этаж	Панельный	17,5	34	9	Не «угловая»	Есть балкон	0,8	45
57	Улуч. планировка	Промеж. этаж	Панельный	14	31	10	Не «угловая»	Есть балкон	0,7	40
58	«Хрущёвка»	Промеж. этаж	Кирпичный	18	32	6	Не «угловая»	Есть балкон	0,7	42
59	«Хрущёвка»	Первый этаж	Кирпичный	18	34	6	«Угловая»	Нет балкона	0,4	35
60	Улуч. планировка	Последний этаж	Панельный	19,8	38	9	Не «угловая»	Есть балкон	0,6	48
61	«Хрущёвка»	Первый этаж	Панельный	14	20	5	«Угловая»	Нет балкона	0,6	35
62	Улуч. планировка	Последний этаж	Панельный	19,3	43	9	Не «угловая»	Есть балкон	0,6	50

Таблица Б.4 – Исходные данные для двухкомнатной квартиры

№	x1	x2	x3	x4	x5	x6	x7	x8	x9	y2
1	Улуч. планировка	Промеж. этаж	Панельный	30	53	9	"Не угло-вая"	Есть балкон	0,6	72
2	Улуч. планировка	Первый этаж	Панельный	28,5	52	7	"Угловая"	Есть балкон	0,3	62
3	Улуч. планировка	Последний этаж	Панельный	30	53	9	"Не угло-вая"	Есть балкон	0,3	60
4	Улуч. планировка	Первый этаж	Панельный	35	60	11	"Не угло-вая"	Есть балкон	0,9	75
5	Улуч. планировка	Последний этаж	Панельный	30	50	10	"Не угло-вая"	Есть балкон	0,9	65
6	Улуч. планировка	Первый этаж	Панельный	32	58	9	"Не угло-вая"	Есть балкон	0,4	78
7	Улуч. планировка	Последний этаж	Панельный	28	50	9	"Не угло-вая"	Есть балкон	0,8	60
8	Улуч. планировка	Первый этаж	Панельный	33	53	9	"Не угло-вая"	Есть балкон	0,6	67
9	Улуч. планировка	Последний этаж	Панельный	30	53	9	"Не угло-вая"	Есть балкон	0,3	65
10	Улуч. планировка	Промеж. этаж	Панельный	31	52,9	8,5	"Не угло-вая"	Есть балкон	0,9	68
11	Улуч. планировка	Промеж. этаж	Панельный	30	50	11	"Угловая"	Есть балкон	0,9	65
12	Улуч. планировка	Последний этаж	Панельный	28	50	9	"Угловая"	Есть балкон	0,7	67
13	Улуч. планировка	Первый этаж	Панельный	28	50	9	"Угловая"	Есть балкон	0,6	70
14	Улуч. планировка	Промеж. этаж	Панельный	28	47	9	"Угловая"	Есть балкон	0,4	65
15	Улуч. планировка	Последний этаж	Панельный	28	50	9	"Угловая"	Есть балкон	0,6	70
16	Улуч. планировка	Первый этаж	Панельный	28	50	10	"Не угло-вая"	Есть балкон	0,7	65
17	Улуч. планировка	Последний этаж	Панельный	29,4	50	9	"Не угло-вая"	Есть балкон	0,6	70
18	Улуч. планировка	Первый этаж	Панельный	33,1	52,9	8	"Не угло-вая"	Есть балкон	0,7	73
19	Улуч. планировка	Последний этаж	Панельный	34	60	9	"Не угло-вая"	Есть балкон	0,8	70
20	Улуч. планировка	Первый этаж	Панельный	30	46	7	"Не угло-вая"	Есть балкон	0,9	68
21	Улуч. планировка	Последний этаж	Панельный	32	52	7	"Не угло-вая"	Есть балкон	0,4	75
22	Улуч. планировка	Первый этаж	Панельный	29	60	9	"Не угло-вая"	Есть балкон	0,5	70
23	Улуч. планировка	Последний этаж	Кирпичный	28	54	6	"Не угло-вая"	Есть балкон	0,9	65
24	Улуч. планировка	Промеж. этаж	Кирпичный	33	52	6	"Не угло-вая"	Есть балкон	0,7	60

Продолжение таблицы Б.4

1	2	3	4	5	6	7	8	9	10	11
25	Улуч. планировка	Последний этаж	Кирпичный	25	47	6	"Не угло-вая"	Есть балкон	0,7	62
26	Улуч. планировка	Первый этаж	Кирпичный	25	41,5	6	"Не угло-вая"	Есть балкон	0,6	65
27	Улуч. планировка	Последний этаж	Кирпичный	31	43	6	"Не угло-вая"	Есть балкон	0,8	60
28	"Хрущёвка"	Первый этаж	Кирпичный	29,5	43	6	"Не угло-вая"	Есть балкон	0,8	63
29	"Хрущёвка"	Последний этаж	Кирпичный	29,5	45	6	"Не угло-вая"	Есть балкон	0,7	60
30	"Хрущёвка"	Промеж. этаж	Кирпичный	26	40	6	"Не угло-вая"	Есть балкон	0,7	58
31	"Хрущёвка"	Последний этаж	Кирпичный	32	52	6	"Не угло-вая"	Есть балкон	0,7	60
32	"Хрущёвка"	Первый этаж	Кирпичный	30	50	6,5	"Не угло-вая"	Есть балкон	0,6	65
33	"Хрущёвка"	Последний этаж	Кирпичный	28,8	44	6	"Не угло-вая"	Нет балкона	0,5	60
34	"Хрущёвка"	Первый этаж	Кирпичный	26	45	8	"Не угло-вая"	Есть балкон	0,6	60
35	"Хрущёвка"	Последний этаж	Кирпичный	31	43	6	"Не угло-вая"	Есть балкон	0,4	65
36	"Хрущёвка"	Первый этаж	Кирпичный	30	45	6	"Не угло-вая"	Есть балкон	0,6	58
37	"Хрущёвка"	Промеж. этаж	Кирпичный	26,4	39,8	6	"Угловая"	Есть балкон	0,3	58
38	"Хрущёвка"	Последний этаж	Кирпичный	31	41	6	"Угловая"	Есть балкон	0,8	55
39	"Хрущёвка"	Промеж. этаж	Кирпичный	30	46	6	"Не угло-вая"	Есть балкон	0,8	55
40	Улуч. планировка	Первый этаж	Кирпичный	31	50	9	"Не угло-вая"	Есть балкон	0,3	79
41	Улуч. планировка	Промеж. этаж	Кирпичный	30	49	8	"Угловая"	Есть балкон	0,7	60
42	Улуч. планировка	Последний этаж	Кирпичный	28	54	7	"Не угло-вая"	Есть балкон	0,1	68
43	Улуч. планировка	Промеж. этаж	Кирпичный	30	47	6	"Не угло-вая"	Есть балкон	0,7	63
44	Улуч. планировка	Первый этаж	Панельный	25	45	6	"Не угло-вая"	Есть балкон	0,8	60
45	Улуч. планировка	Промеж. этаж	Кирпичный	28	48	6	"Не угло-вая"	Нет балкона	0,2	66
46	Улуч. планировка	Промеж. этаж	Кирпичный	30	50	8	"Не угло-вая"	Есть балкон	0,7	62
47	Улуч. планировка	Последний этаж	Кирпичный	28	45	8	"Угловая"	Есть балкон	0,9	65
48	Улуч. планировка	Первый этаж	Кирпичный	22	40	9	"Угловая"	Есть балкон	0,5	62

Продолжение таблицы Б.4

1	2	3	4	5	6	7	8	9	10	11
49	Улуч. планировка	Промеж. этаж	Кирпичный	30	49	6	"Не угло-вая"	Есть балкон	0,8	59
50	Улуч. планировка	Промеж. этаж	Кирпичный	31	46	6	"Не угло-вая"	Нет балкона	0,6	60
51	Улуч. планировка	Последний этаж	Панельный	33,1	52,9	8	"Не угло-вая"	Есть балкон	0,7	70
52	Улуч. планировка	Промеж. этаж	Кирпичный	30	44	8	"Угловая"	Есть балкон	0,9	58
53	"Хрущёвка"	Первый этаж	Кирпичный	30	45	6	"Не угло-вая"	Есть балкон	0,6	60
54	"Хрущёвка"	Промеж. этаж	Кирпичный	26	40	6	"Угловая"	Есть балкон	0,4	62
55	Улуч. планировка	Промеж. этаж	Панельный	31	53	8,5	"Не угло-вая"	Есть балкон	0,9	68
56	Улуч. планировка	Последний этаж	Панельный	30	60	11	"Не угло-вая"	Есть балкон	0,8	72
57	Улуч. планировка	Промеж. этаж	Панельный	28	47	9	"Угловая"	Есть балкон	0,5	65
58	"Хрущёвка"	Первый этаж	Кирпичный	30	43	6	"Не угло-вая"	Есть балкон	0,8	63
59	Улуч. планировка	Последний этаж	Панельный	25	42	6	"Не угло-вая"	Есть балкон	0,7	60
60	Улуч. планировка	Первый этаж	Панельный	33	53	8	"Не угло-вая"	Есть балкон	0,6	73
61	Улуч. планировка	Последний этаж	Панельный	30	53	9	"Не угло-вая"	Есть балкон	0,4	65
62	Улуч. планировка	Промеж. этаж	Кирпичный	30,4	52,4	8	"Угловая"	Есть балкон	0,3	70
63	"Хрущёвка"	Первый этаж	Кирпичный	29,5	45	6	"Не угло-вая"	Есть балкон	0,7	65
64	Улуч. планировка	Последний этаж	Панельный	29,4	45	9	"Не угло-вая"	Нет балкона	0,5	70
65	Улуч. планировка	Первый этаж	Кирпичный	33	52	8	"Угловая"	Есть балкон	0,2	65
66	Улуч. планировка	Последний этаж	Панельный	28	50	9	"Угловая"	Есть балкон	0,6	70
67	"Хрущёвка"	Первый этаж	Кирпичный	26	45	8	"Не угло-вая"	Нет балкона	0,8	58
68	Улуч. планировка	Последний этаж	Панельный	29	54	9	"Не угло-вая"	Есть балкон	0,6	65
69	Улуч. планировка	Промеж. этаж	Кирпичный	28	50	7	"Не угло-вая"	Нет балкона	0,2	66
70	Улуч. планировка	Первый этаж	Кирпичный	24,8	38,6	6	"Угловая"	Нет балкона	0,2	65
71	Улуч. планировка	Последний этаж	Кирпичный	28,3	44,7	6	"Не угло-вая"	Нет балкона	0,8	60
72	Улуч. планировка	Первый этаж	Кирпичный	30	50	6	"Не угло-вая"	Есть балкон	0,3	68

Продолжение таблицы Б.4

1	2	3	4	5	6	7	8	9	10	11
73	Улуч. планировка	Промеж. этаж	Кирпичный	31	46	6	"Не угло-вая"	Нет балкона	0,7	59
74	Улуч. планировка	Последний этаж	Кирпичный	28	54	6	"Не угло-вая"	Нет балкона	0,2	66
75	"Хрущёвка"	Промеж. этаж	Кирпичный	30	47	6	"Не угло-вая"	Есть балкон	0,7	63
76	"Хрущёвка"	Первый этаж	Кирпичный	26	45	8	"Не угло-вая"	Нет балкона	0,8	58
77	Улуч. планировка	Промеж. этаж	Кирпичный	30	50	6	"Не угло-вая"	Есть балкон	0,7	59
78	Улуч. планировка	Последний этаж	Кирпичный	29	50	8,5	"Не угло-вая"	Есть балкон	0,6	70
79	Улуч. планировка	Первый этаж	Панельный	36	54	6,8	"Не угло-вая"	Есть балкон	0,4	72
80	"Хрущёвка"	Последний этаж	Кирпичный	29	45	6	"Не угло-вая"	Есть балкон	0,7	58
81	"Хрущёвка"	Первый этаж	Кирпичный	28	40	5	"Угловая"	Есть балкон	0,7	57
82	Улуч. планировка	Последний этаж	Кирпичный	30	50	8,5	"Не угло-вая"	Есть балкон	0,5	68
83	Улуч. планировка	Первый этаж	Кирпичный	30	50	8,5	"Не угло-вая"	Есть балкон	0,2	72

Таблица Б.5 – Исходные данные для трехкомнатной квартиры

№	x1	x2	x3	x4	x5	x6	x7	x8	x9	y3
1	Улуч. планировка	Первый этаж	Панельный	39,9	64	9	"Угловая"	Есть балкон	0,4	83
2	"Хрущёвка"	Промеж. этаж	Кирпичный	34,1	50	6	"Не угло-вая"	Есть балкон	0,8	72
3	"Хрущёвка"	Последний этаж	Кирпичный	47	55	6	"Не угло-вая"	Нет балкона	0,2	95
4	Улуч. планировка	Промеж. этаж	Панельный	39	57	9	"Не угло-вая"	Есть балкон	0,3	100
5	Улуч. планировка	Первый этаж	Панельный	43,8	63,8	9	"Не угло-вая"	Нет балкона	0,9	78
6	Улуч. планировка	Промеж. этаж	Панельный	44,7	60	7	"Не угло-вая"	Есть балкон	0,8	80
7	Улуч. планировка	Промеж. этаж	Панельный	44	60	6	"Не угло-вая"	Есть балкон	0,8	80
8	Улуч. планировка	Промеж. этаж	Панельный	40	62	8	"Не угло-вая"	Есть балкон	0,5	90
9	Улуч. планировка	Промеж. этаж	Панельный	36	58	6,8	"Не угло-вая"	Есть балкон	0,6	80
10	"Хрущёвка"	Последний этаж	Кирпичный	43,4	62	6,5	"Не угло-вая"	Нет балкона	0,6	75

Продолжение таблицы Б.5

1	2	3	4	5	6	7	8	9	10	11
11	Улуч. планировка	Промеж. этаж	Панельный	39	65	9	"Не угловая"	Есть балкон	0,5	100
12	Улуч. планировка	Промеж. этаж	Панельный	38	62	9	"Угловая"	Есть балкон	0,5	85
13	Улуч. планировка	Промеж. этаж	Кирпичный	37	64	8	"Не угловая"	Есть балкон	0,4	110
14	Улуч. планировка	Промеж. этаж	Панельный	39	65	9	"Не угловая"	Есть балкон	0,4	110
15	Улуч. планировка	Промеж. этаж	Панельный	44	78	10	"Не угловая"	Есть балкон	0,4	105
16	Улуч. планировка	Промеж. этаж	Панельный	39,9	63	11	"Не угловая"	Есть балкон	0,5	93
17	Улуч. планировка	Промеж. этаж	Панельный	40	64	9	"Не угловая"	Нет балкона	0,6	88
18	Улуч. планировка	Промеж. этаж	Панельный	42	64	9	"Не угловая"	Есть балкон	0,4	105
19	Улуч. планировка	Первый этаж	Кирпичный	40	58	6	"Не угловая"	Есть балкон	0,6	78
20	Улуч. планировка	Последний этаж	Кирпичный	44	76	9	"Не угловая"	Есть балкон	0,1	120
21	Улуч. планировка	Промеж. этаж	Кирпичный	40	64	9	"Не угловая"	Есть балкон	0,2	120
22	Улуч. планировка	Промеж. этаж	Кирпичный	38	60	9	"Не угловая"	Есть балкон	0,4	100
23	Улуч. планировка	Первый этаж	Панельный	38	62	8	"Не угловая"	Есть балкон	0,8	83
24	Улуч. планировка	Промеж. этаж	Панельный	39,9	64	9	"Не угловая"	Есть балкон	0,6	90
25	Улуч. планировка	Промеж. этаж	Панельный	40	64	9	"Не угловая"	Есть балкон	0,6	105
26	Улуч. планировка	Промеж. этаж	Панельный	39	63	9	"Угловая"	Есть балкон	0,6	93
27	Улуч. планировка	Последний этаж	Панельный	40	64	9	"Угловая"	Есть балкон	0,6	90
28	Улуч. планировка	Промеж. этаж	Панельный	39	64,4	9	"Не угловая"	Есть балкон	0,6	85
29	Улуч. планировка	Промеж. этаж	Панельный	40	64	9	"Угловая"	Есть балкон	0,6	95
30	Улуч. планировка	Промеж. этаж	Панельный	39,9	63	9	"Не угловая"	Есть балкон	0,6	105
31	Улуч. планировка	Промеж. этаж	Панельный	40	64	9	"Угловая"	Есть балкон	0,6	96
32	Улуч. планировка	Промеж. этаж	Панельный	39	63	9	"Не угловая"	Есть балкон	0,6	100
33	Улуч. планировка	Промеж. этаж	Кирпичный	39	65	9	"Не угловая"	Есть балкон	0,5	100
34	Улуч. планировка	Первый этаж	Панельный	40	64	9	"Не угловая"	Есть балкон	0,6	90

Продолжение таблицы Б.5

1	2	3	4	5	6	7	8	9	10	11
35	Улуч. планировка	Промеж. этаж	Панельный	40	64	9	"Угловая"	Есть балкон	0,5	100
36	Улуч. планировка	Промеж. этаж	Панельный	39,9	64	9	"Не угло-вая"	Есть балкон	0,7	90
37	Улуч. планировка	Промеж. этаж	Панельный	36	62	6,8	"Не угло-вая"	Есть балкон	0,5	92
38	"Хрущёвка"	Промеж. этаж	Кирпичный	40	64	6	"Не угло-вая"	Есть балкон	0,8	85
39	"Хрущёвка"	Последний этаж	Кирпичный	40	59	8	"Не угло-вая"	Нет балкона	0,7	85
40	Улуч. планировка	Промеж. этаж	Панельный	39	63	9	"Угловая"	Есть балкон	0,6	100
41	Улуч. планировка	Промеж. этаж	Кирпичный	46	78	14	"Не угло-вая"	Нет балкона	0,1	125
42	Улуч. планировка	Первый этаж	Панельный	40	64	9	"Не угло-вая"	Есть балкон	0,3	110
43	Улуч. планировка	Промеж. этаж	Панельный	41	64	8,5	"Не угло-вая"	Есть балкон	0,6	90
44	Улуч. планировка	Последний этаж	Панельный	38,8	62,6	8	"Не угло-вая"	Есть балкон	0,5	90
45	Улуч. планировка	Промеж. этаж	Панельный	39	64	9	"Не угло-вая"	Есть балкон	0,6	90
46	Улуч. планировка	Промеж. этаж	Кирпичный	37	64	8	"Не угло-вая"	Есть балкон	0,5	95
47	Улуч. планировка	Первый этаж	Кирпичный	40	58	6	"Не угло-вая"	Нет балкона	0,6	78
48	Улуч. планировка	Промеж. этаж	Панельный	40	62	8	"Не угло-вая"	Есть балкон	0,5	92
49	"Хрущёвка"	Последний этаж	Кирпичный	40	59	9	"Не угло-вая"	Нет балкона	0,5	75
50	"Хрущёвка"	Первый этаж	Кирпичный	40	56	6	"Не угло-вая"	Нет балкона	0,4	78
51	Улуч. планировка	Промеж. этаж	Панельный	39	63	9	"Не угло-вая"	Есть балкон	0,7	90
52	Улуч. планировка	Промеж. этаж	Панельный	39	64	10	"Не угло-вая"	Есть балкон	0,6	85
53	Улуч. планировка	Промеж. этаж	Панельный	38	60	8,5	"Не угло-вая"	Есть балкон	0,6	85
54	Улуч. планировка	Промеж. этаж	Панельный	39	64	9	"Не угло-вая"	Есть балкон	0,6	100
55	Улуч. планировка	Промеж. этаж	Панельный	39	64	9	"Не угло-вая"	Есть балкон	0,8	90
56	Улуч. планировка	Промеж. этаж	Панельный	39	63	8,5	"Угловая"	Есть балкон	0,8	90

Приложение В

(обязательное)

Информационная база для моделирования курса ценных бумаг

Таблица В.1 – Выборочные данные по курсам ценных бумаг

Кварталы	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
I 1997	37,26	17,92	48,73	17,69	5	73,20	73,20	239	239,11	22,06
II 1997	73,11	32,58	84,55	30,95	24	93,77	60,00	258	232,83	34,16
III 1997	38,93	51,36	45,76	33,72	60	99,17	46,50	280	210,83	32,29
IV 1997	70,39	113,09	60,38	32,16	32	64,40	70,20	262	262,04	-29,99
I 1998	10,99	137,24	35,70	93,02	53	123,83	54,40	274	227,14	2,23
II 1998	66,99	151,01	87,01	86,57	97	64,98	52,00	329	245,74	-0,14
III 1998	64,81	157,07	85,94	101,94	137	150,11	49,10	353	233,99	-1,09
IV 1998	107,93	236,49	143,51	118,75	118	110,68	76,00	276	237,94	20,97
I 1999	45,85	223,48	33,28	106,07	113	113,11	63,20	382	233,33	6,43
II 1999	99,16	282,30	133,64	124,24	137	110,92	57,90	437	234,13	13,29
III 1999	87,87	284,93	112,82	125,72	158	72,66	72,66	400	315,21	10,61
IV 1999	88,45	328,65	72,74	126,31	126	95,85	91,20	297	212,45	1,20
I 2000	63,85	406,55	103,65	134,17	134	119,01	78,90	440	211,86	14,68
II 2000	105,66	367,87	124,23	157,67	157	114,24	72,80	500	266,59	4,43
III 2000	105,62	370,08	124,49	173,40	189	120,64	66,70	445	249,84	-4,29
IV 2000	129,29	430,08	170,27	158,34	158	86,98	100,00	413	295,14	-28,64
I 2001	132,70	395,57	161,13	230,67	211	135,56	89,50	492	249,39	15,39
II 2001	137,93	428,27	181,21	235,01	253	132,91	72,80	532	292,30	-23,44
III 2001	137,84	490,12	150,95	213,47	263	121,16	93,90	492	262,41	5,84
IV 2001	158,59	502,39	197,75	285,03	240	112,34	112,34	450	311,71	30,60
I 2002	154,43	528,37	176,88	199,19	284	119,13	102,60	468	272,09	46,37
II 2002	174,30	592,22	199,82	268,18	316	126,83	86,80	524	256,52	24,95
III 2002	183,70	594,53	244,20	287,37	316	102,94	102,94	605	286,45	-0,86
IV 2002	162,16	599,75	189,02	292,67	292	112,14	128,90	497	272,77	28,89
I 2003	191,31	625,82	208,40	307,94	307	138,96	111,40	445	260,80	28,40
II 2003	226,16	681,13	211,19	310,86	353	103,24	93,00	504	316,45	31,72
III 2003	262,08	731,02	258,91	347,52	397	160,03	116,70	562	303,12	31,21
IV 2003	291,44	694,18	321,57	355,10	355	138,28	138,28	454	355,63	23,02
I 2004	272,92	745,36	293,26	357,05	380	98,78	116,00	457	354,30	12,64
II 2004	302,23	790,22	333,60	348,87	426	149,00	104,00	510	380,25	13,69
III 2004	323,66	772,85	283,81	389,70	460	124,27	124,27	467	362,29	-4,33
IV 2004	382,30	869,02	375,38	383,26	383	148,74	148,74	424	402,19	13,62
I 2005	377,94	871,44	363,82	402,61	413	142,25	93,00	516	368,14	46,50
II 2005	409,12	878,85	404,27	432,87	460	170,11	86,80	537	375,11	10,23
III 2005	418,23	898,02	434,47	432,19	480	157,76	104,00	494	403,62	-5,08
IV 2005	475,85	932,06	482,71	426,01	426	133,88	153,50	446	446,38	-4,35
I 2006	452,06	938,44	479,89	468,04	468	154,38	124,60	494	407,30	13,92
II 2006	528,85	944,70	542,70	470,03	505	141,17	100,00	553	431,81	13,03
III 2006	561,82	1013,30	589,11	483,11	555	158,00	136,80	574	502,04	78,10
IV 2006	582,46	1052,46	607,87	495,30	495	178,50	185,10	487	487,48	11,39