КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ УПРАВЛЕНИЯ, ЭКОНОМИКИ И ФИНАНСОВ

Кафедра экономико-математического моделирования

И. И. ИСМАГИЛОВ, Е.И. КАДОЧНИКОВА

СПЕЦИАЛЬНЫЕ МОДЕЛИ ЭКОНОМЕТРИКИ В СРЕДЕ GRETL

Учебное пособие для студентов,
обучающихся по направлению 38.04.01 «Экономика»

Казань 2018

УДК 330.43

ББК Ув631я73-1

Рекомендовано к публикации на заседании кафедры

экономико-математического моделирования

Протокол № 3 от 24 ноября 2017 года

Рецензенты:

доктор экономических наук,

заведующий кафедрой бизнес-статистики и

математических методов в экономике

Казанского национального исследовательского технологического

университета

А. В. Аксянова

кандидат экономических наук,

доцент кафедры экономико-математического моделирования ИУЭиФ КФУ

Е. Л. Фесина

Исмагилов И.И., Кадочникова Е.И. Специальные модели эконометрики в среде Gretl: учебное пособие для студентов, обучающихся по направлению 38.04.01 «Экономика» / И.И. Исмагилов, Е.И. Кадочникова – Казань: Казан. ун-т, 2018. – 91 с.

Данное учебное пособие предназначено для использования на практических занятиях по дисциплине «Эконометрика (продвинутый уровень)» для магистерских программ направления 38.04.01 «Экономика» и других направлений. Цель учебно-методического пособия – развить практические умения и навыки построения эконометрических моделей средствами пакета Gretl.

Содержание

Введение

Учебное пособие по дисциплине «Эконометрика (продвинутый уровень)» предназначено для магистрантов направления подготовки 38.04.01 «Экономика», студентов других направлений, а также представителей бизнес-сообщества, составлено в соответствии с рабочей программой дисциплины и требованиями действующего Федерального государственного образовательного стандарта высшего образования.

Данное учебное пособие содержит практическое изложение специальных моделей эконометрики в примерах и задачах, предполагается, что недостаток теоретических знаний может быть восполнен читателем с помощью рекомендуемых в пособии учебников. В результате обучения магистранты должны уметь обрабатывать экономические и финансовые показатели, характеризующие микро- и макроэкономику, идентифицировать, верифицировать и интерпретировать модели, выполнять необходимые преобразования для улучшения качества моделей, применяя компьютерные технологии, в частности программный пакет Gretl. Ориентация на Gretl обусловлена следующими моментами. Во-первых, это бесплатный свободно распространяемый, достаточно удобный и универсальный пакет для выполнения эконометрических расчетов. Во-вторых, Gretl предоставляет возможность «почувствовать» все детали и тонкости изучаемых методов при их реализации на основе соответствующих векторно-матричных соотношений, что повышает уровень усвоения учебного материала.

Использование данного учебного пособия на практических занятиях формирует умения:

- корректно осуществлять спецификацию эконометрических моделей;
- проверять адекватность построенных моделей и значимость их параметров;
- интерпретировать содержательный смысл параметров эконометрических моделей;
- грамотно использовать компьютерное программное обеспечение для

расчёта оценок параметров эконометрических моделей.

Также учебное пособие ориентировано на развитие владений:

- навыками использования современного эконометрического инструментария для исследований экономических и финансовых решений на уровне индивидов, домохозяйств, фирм, финансовых рынков, финансовых институтов, отраслей, регионов;

- навыками моделирования результатов и эффективности субъектов экономической деятельности.

Авторами рекомендована целесообразная форма проведения практических занятий. Студенты сначала сообща в малых группах (2-3 человека) выполняют задание с применением Gretl на компьютере, делая при этом необходимые выводы. Затем представляют свой доклад по задаче перед аудиторией и интерпретируют результаты решения. Наличие расчетных формул и пошагового описания выполнения эконометрических расчетов позволяют применять учебное пособие для самостоятельной работы студентов заочного обучения.

Программный пакет Gretl доступен по ссылке http://gretl.sourceforge.net/ru.html. Вся информация о том, как установить Gretl, приводится на сайте, программа имеет версию как под ОС Windows, так и под Mac OS. В Gretl имеется значительные библиотеки данных, которые доступны по ссылке: http://gretl.sourceforge.net/gretl_data.html.

1. Модели бинарного выбора: логит-модель, пробит-модель

Расчетные формулы

В моделях бинарного выбора зависимые переменные являются двоичными.

Модель для двоичной переменной имеет вид:

$$y_i^* = \beta_0 + \sum_{j=1}^{k} \beta_i x_{ij} + u_i,$$

y_i^* – скрытая (латентная) переменная,

$$yi = \begin{cases} 1, y_i^* > 0, \\ 0, y_i^* \le 0. \end{cases}$$

Представленная модель называется моделью вероятности. Часто на практике используется логитовая модель вида:

$$y_i^* = \ln \frac{P_i}{1 - P_i} = \beta_0 + \sum_{j=1}^{k} \beta_i x_{ij} + u_i,$$

где y_i^* называется логитом, а P_i определяется вероятностью зависимой переменной Y_t, рассчитываемой на основе логистического распределения:

$$\frac{P_i}{1 - P_i} = e^{y_t^*} = e^{\beta_0 + \sum_{j=1}^{k} \beta_i x_{ij} + u_i},$$

$$\hat{P}_i = \frac{1}{1 + e^{-y_t^*}} = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{j=1}^{k} \beta_i x_{ij}\right)}},$$

В пробитовой модели, по аналогии с логитовой моделью, ненаблюдаемая величина P_i рассчитывается исходя из функции стандартного нормального распределения как:

$$Pi = F(\beta_0 + \sum_{j=1}^{k} \beta_i x_{ij}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0 + \sum_{j=1}^{k} \beta_i x_{ij}} e^{-\frac{u^2}{2}} du,$$

Параметры β логитовой и пробитовой моделей связаны соотношением: $\beta_{\log it} = 1,68 \beta_{probit}$.

На практике может использоваться любой из этих методов, поскольку все меры соответствия моделей эмпирическим данным оказываются для них идентичными.

Логитовый анализ используется в экономических исследованиях применительно к срезам индивидуальных данных тогда, когда эндогенная переменная имеет двоичный характер. Чаще всего эта переменная представляет результаты принятия рациональных экономических решений, например, приобретения автомобиля или квартиры, предоставления банковского кредита, слияния или поглощения фирм.

Задание для выполнения в аудитории

Задание 1.1. Фирма «Бэст Перспектива» заключила договор с ОАО «Центртелеком», предметом которого является оказание услуг связи, в том числе и доступ в Интернет. С целью увеличения интеллектуального потенциала фирмы руководство решило расширить круг своих сотрудников, обладающих правом бесплатного доступа. Выбор таких сотрудников решено было осуществлять с учетом ряда факторов, тем или иным образом характеризующих претендентов с точки зрения эффективного использования Интернет-ресурсов. В связи с эти возник вопрос: Кому из претендентов на бесплатный доступ предоставить такую возможность в первую очередь? Для того, чтобы получить обоснованный ответ, руководство фирмы поручило экономико-аналитическому отделу разработать модель, позволяющую по каждому претенденту рассчитать прогнозную оценку целесообразности предоставления ему права бесплатного доступа к ресурсам Интернета. В основу построения такой модели была положена идея применения бинарной переменной Y: $Y=1$ если сотрудник, обладающий правом бесплатного доступа к Интернет-ресурсам, по оценке экспертной группы, эффективно пользуется эти правом; $Y=0$, в противном случае. Построение модели руководство фирмы предложило провести по пяти факторам: возраст ($X1$, лет), стаж профессиональной деятельности ($X2$, лет), заработная плата ($X3$, тыс. ден. ед.), число случаев поступления полезной для фирмы информации от сотрудника ($X4$, ед.), результат тестирования на предмет оценки навыков работы в Интернет ($X5$, в баллах). Значения этих показателей, а также значения бинарной переменной для 100 сотрудников фирмы представлены в таблице 1.1.

Таблица 1.1

№	Y	$X1$	$X2$	$X3$	$X4$	$X5$	№	Y	$X1$	$X2$	$X3$	$X4$	$X5$
1	0	22	1	2,5	2	6	51	0	57	18	10,2	2	2
2	0	24	1	3	3	8	52	0	28	4	6,7	1	11
3	1	25	1	2,1	1	7	53	0	52	14	9,9	2	2
4	1	27	4	4,6	8	11	54	0	45	15	13,1	3	4
5	1	28	3	5,9	9	13	55	0	53	16	9,1	1	4
6	0	21	1	3,3	5	14	56	0	54	20	9,7	1	5
7	1	22	1	3,2	9	15	57	0	30	4	6,8	3	12
8	0	29	5	6	1	13	58	0	46	12	12,2	2	1
9	1	26	4	2,7	8	11	59	0	28	4	3,9	0	2
10	0	27	6	3,8	0	3	60	0	44	14	12,5	1	2
11	1	28	4	4,2	9	11	61	0	22	1	2,8	3	7

12	0	29	3	6,7	5	13	62	0	29	4	3,4	0	2
13	1	29	9	6,5	10	14	63	0	23	1	4,9	4	15
14	1	30	4	3,4	9	13	64	1	26	8	6,7	12	15
15	1	41	7	7,2	7	15	65	1	27	9	6,6	11	13
16	0	32	8	7,9	4	7	66	0	39	4	8,9	2	8
17	1	46	9	6,4	9	12	67	1	27	7	5,5	11	11
18	1	34	3	6,3	8	13	68	0	26	6	5,7	2	11
19	0	33	3	6,1	1	9	69	1	28	9	6,1	12	12
20	1	47	9	9,7	5	13	70	1	28	8	7,3	10	13
21	0	37	2	6,4	3	7	71	0	40	4	9,2	2	7
22	0	38	5	10,5	2	8	72	1	28	10	7,9	10	11
23	0	49	13	12,3	3	3	73	1	22	1	2,5	8	12
24	1	46	7	10,1	5	14	74	1	26	9	6,8	11	5
25	1	48	8	7,9	8	15	75	0	40	5	6	2	9
26	0	52	17	9,7	2	1	76	1	26	8	5,4	11	15
27	0	54	12	8,9	3	4	77	0	39	5	7,6	2	7
28	1	18	1	3,4	4	11	78	0	32	9	5,5	5	8
29	0	19	1	3,7	5	12	79	1	33	4	5,4	7	11
30	1	20	1	2,9	6	13	80	1	23	1	2,8	8	13
31	0	19	1	3,8	5	15	81	0	36	3	6,5	2	9
32	0	20	1	4,3	1	8	82	0	35	9	6,9	5	8
33	1	28	2	6,9	9	14	83	1	23	1	3,2	9	11
34	0	27	4	6,5	5	12	84	1	34	3	6,1	8	12
35	0	43	8	9,2	2	9	85	0	35	8	5,9	5	9
36	1	44	13	10,3	11	13	86	0	38	2	7,8	3	10
37	0	56	12	10,8	2	4	87	0	32	9	7,3	5	8
38	1	54	14	9,4	12	13	88	0	33	2	6,4	3	10
39	0	44	11	9,8	2	6	89	1	34	3	6,5	8	11
40	1	28	4	6	8	12	90	0	35	10	7,9	5	10
41	0	19	1	3,4	2	7	91	0	33	2	7,8	1	9
42	0	20	1	4,3	5	11	92	1	24	1	3	8	11
43	0	19	1	4,9	5	12	93	1	34	4	6	9	13
44	1	21	1	3,2	6	13	94	0	33	3	5,4	2	10
45	0	45	12	12,4	2	4	95	1	22	1	2,9	7	13
46	0	19	1	3,1	4	13	96	1	32	5	6,4	8	15
47	0	48	13	12,7	2	5	97	0	33	6	5,7	2	7
48	1	20	1	3,5	4	13	98	1	23	1	3,6	9	15
49	0	58	18	9,9	2	2	99	1	31	4	6	8	13
50	0	49	12	11,8	3	3	100	0	33	4	6,1	3	10

Имеются следующие претенденты на право бесплатного доступа:

1) возраст – 27 лет, стаж – 3 года, заработная плата – 3200 руб., количество случаев нахождения полезной для фирмы информации – 9 раз, тест – 15 баллов;

2) возраст – 44 года, стаж – 12 лет, заработная плата – 5600 руб., количество случаев нахождения полезной для фирмы информации – 2 раза, тест – 5 баллов;

3) возраст – 35 лет, стаж – 10 лет, заработная плата – 4100 руб., количество случаев нахождения полезной для фирмы информации – 4 раза, тест – 7 баллов;

4) возраст – 39 лет, стаж – 13 лет, заработная плата – 7500 руб., количество случаев нахождения полезной для фирмы информации – 11 раз, тест – 15 баллов.

Используя построенную прогнозную модель, определить среди имеющихся претендентов тех, кому в первую очередь следует предоставить право бесплатного доступа к ресурсам Интернета.

Методические указания для выполнения задания

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «Занятие_Бинарные модели.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1.

2. Построение логитовой модели: Модель/Нелинейные модели/Логит/Бинарный… (рис. 1.1).

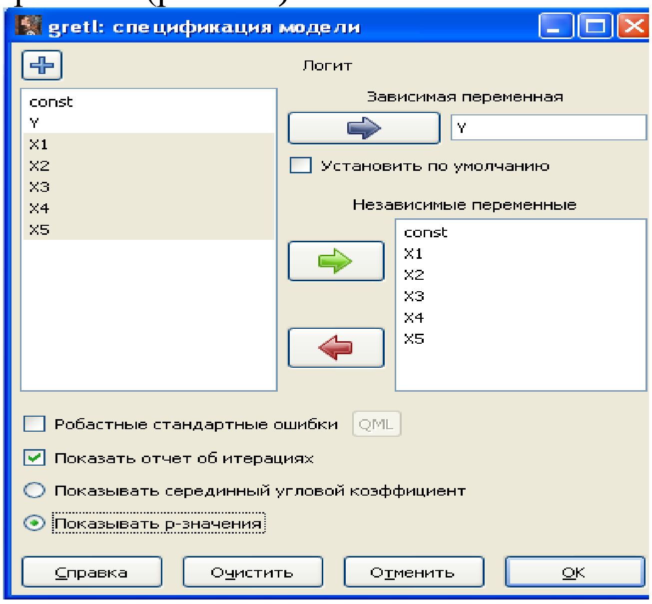


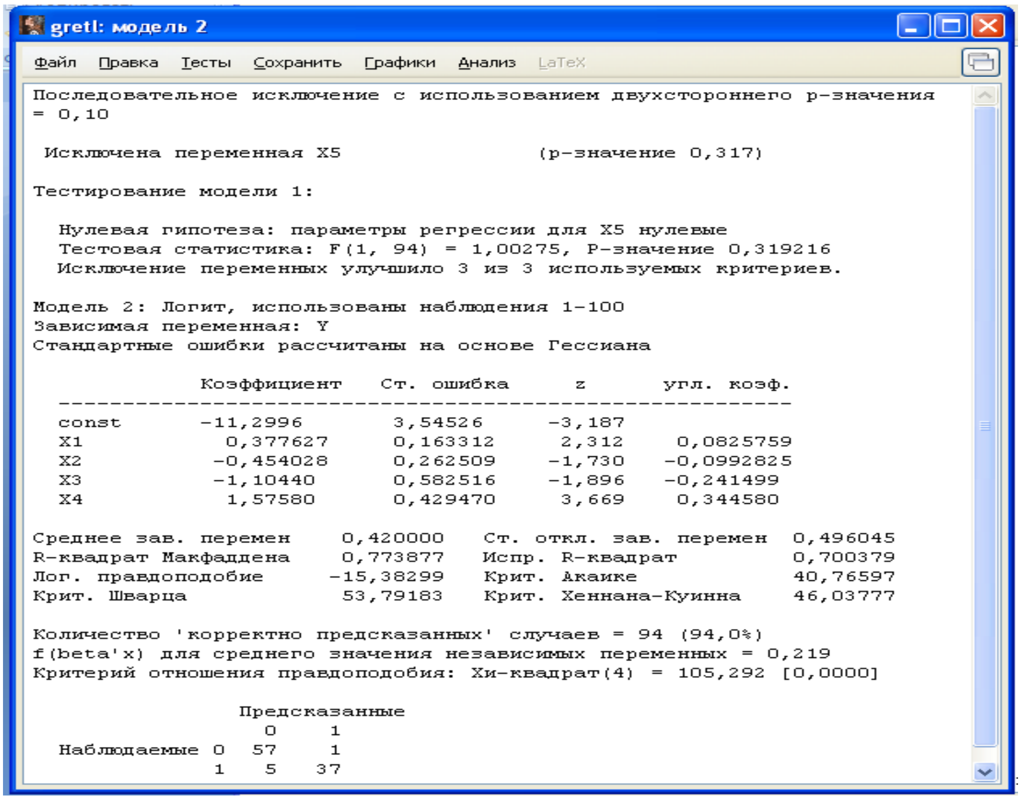
Рис. 1.1. Окно построения логитовой модели

```
--- ОКОНЧАТЕЛЬНЫЕ ЗНАЧЕНИЯ:
лог. правдоподобие = -14,8947017586 (шаг = 1)
Параметры:       -13,202      0,40584     -0,31806      -1,3387
1,4422     0,22418
Градиенты:   3,4490e-011 6,4696e-010-1,0746e-011 9,5627e-011 4,0911e-010
5,5924e-010 (norm 1,61e-005)

Gradient within tolerance (1e-007)

Модель 1: Логит, использованы наблюдения 1-100
Зависимая переменная: Y
Стандартные ошибки рассчитаны на основе Гессиана

             Коэффициент   Ст. ошибка      z       P-значение
  -----------------------------------------------------------
  const       -13,2023      4,20814     -3,137      0,0017   ***
  X1            0,405838    0,173098     2,345      0,0190   **
  X2           -0,318064    0,307407    -1,035      0,3008
  X3           -1,33870     0,665340    -2,012      0,0442   **
  X4            1,44218     0,435792     3,309      0,0009   ***
  X5            0,224176    0,223868     1,001      0,3166

Среднее зав. перемен    0,420000   Ст. откл. зав. перемен  0,496045
R-квадрат Макфаддена    0,781054   Испр. R-квадрат         0,692857
Лог. правдоподобие     -14,89470   Крит. Акаике            41,78940
Крит. Шварца            57,42042   Крит. Хеннана-Куинна    48,11556

Количество 'корректно предсказанных' случаев = 96 (96,0%)
f(beta'x) для среднего значения независимых переменных = 0,202
Критерий отношения правдоподобия: Хи-квадрат(5) = 106,269 [0,0000]

               Предсказанные
                  0     1
  Наблюдаемые 0   57    1
              1   3    39

Исключая константу, наибольшее p-значение получено для переменной 6 (X5)
```

Рис. 1.1. Окно построения логитовой модели (окончание)

В предпоследнем столбце окна результатов приведено отношение оценок к среднему значению, которое может интерпретироваться как направление изменения вероятностей для всех переменных. Параметры направления (z) свидетельствуют, что при увеличении стажа профессиональной деятельности ($X2$) и заработной платы ($X3$) – отрицательный знак параметра, вероятность принятия решения о целесообразности предоставления сотруднику права бесплатного доступа к ресурсам Интернета снижается. При увеличении таких факторов как возраст, результаты тестирования, количество случаев нахождения полезной информации – вероятность увеличивается. Тест Стьюдента свидетельствует, что факторы $X2$, $X5$ являются несущественными.

3. Выполним последовательное исключение несущественных переменных: В окне модели: Тесты/Избыточные переменные/Последовательное исключение переменных (рис. 1.2).




```
Последовательное исключение с использованием двухстороннего p-значения
= 0,10

 Исключена переменная X5              (p-значение 0,317)

Тестирование модели 1:

  Нулевая гипотеза: параметры регрессии для X5 нулевые
  Тестовая статистика: F(1, 94) = 1,00275, P-значение 0,319216
  Исключение переменных улучшило 3 из 3 используемых критериев.

Модель 2: Логит, использованы наблюдения 1-100
Зависимая переменная: Y
Стандартные ошибки рассчитаны на основе Гессиана

             Коэффициент   Ст. ошибка      z       угл. коэф.
  ---------------------------------------------------------
  const       -11,2996       3,54526     -3,187
  X1            0,377627     0,163312      2,312    0,0825759
  X2           -0,454028     0,262509     -1,730   -0,0992825
  X3           -1,10440      0,582516     -1,896   -0,241499
  X4            1,57580      0,429470      3,669    0,344580

Среднее зав. перемен     0,420000   Ст. откл. зав. перемен  0,496045
R-квадрат Макфаддена     0,773877   Испр. R-квадрат         0,700379
Лог. правдоподобие     -15,38299    Крит. Акаике           40,76597
Крит. Шварца            53,79183    Крит. Хеннана-Куинна   46,03777

Количество 'корректно предсказанных' случаев = 94 (94,0%)
f(beta'x) для среднего значения независимых переменных = 0,219
Критерий отношения правдоподобия: Хи-квадрат(4) = 105,292 [0,0000]

                Предсказанные
                   0     1
  Наблюдаемые 0   57     1
              1    5    37
```

Рис. 1.2. Окно исключения несущественных переменных

Обратим внимание на изменение знаков коэффициентов на противоположные при записи модели:

$$\hat{P}_i = \frac{1}{1+e^{-y_t^*}} = \frac{1}{1+e^{-\left(\beta_0+\sum\limits_{j=1}^{k}\beta_i x_{ij}\right)}} = \frac{1}{1+e^{11{,}3-0{,}378X1+0{,}454X2+1{,}104X3-1{,}576X4}}.$$

О значимости совокупной регрессии свидетельствует Хи-квадрат (105,292) и его *p*-значение (0,0000<0,05).

Для оценки прогнозных возможностей построенной модели в качестве меры соответствия модели эмпирическим данным можно использовать коэффициент корреляции между переменной *Yi* и теоретическими значениями модели *Pi*.

4. Сохраним теоретические значения модели в базе данных: В окне модели: Сохранить/Расчетные значения/Название переменной *YR*.

5. Рассчитаем коэффициент корреляции между переменной *Yi* и теоретическими значениями модели *Pi*: Вид/Корреляционная матрица (рис. 1.3).

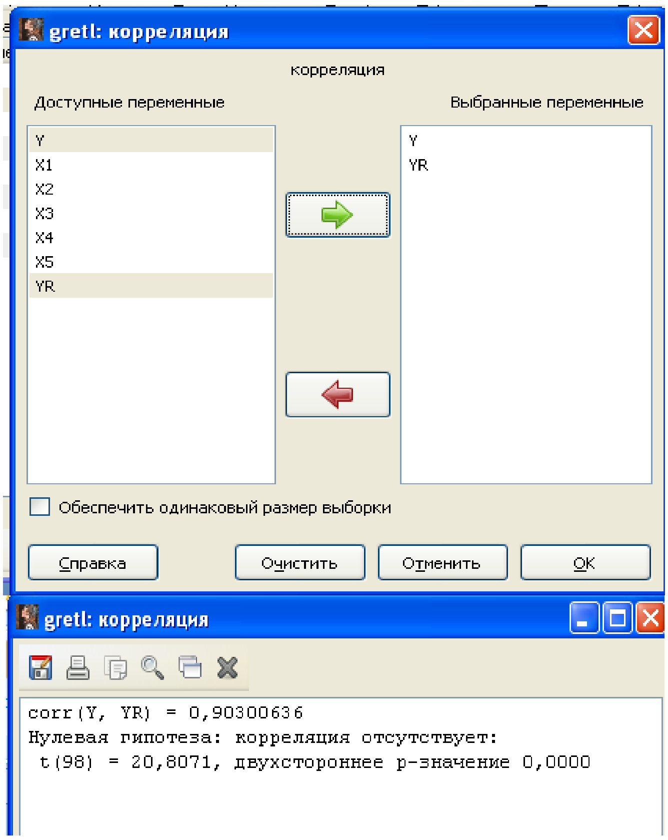


Рис. 1.3. Окно корреляционной матрицы

Коэффициент корреляции составил 0,903, значит, подтверждено соответствие модели эмпирическим данным. Статистика Стьюдента (20,81) подтверждает статистическую значимость коэффициента корреляции. Чаще всего логитовые модели имеют низкий уровень объяснения изменчивости. В нашем случае, наоборот. *R*-квадрат Маккфадена и исправленный *R*-квадрат, коэффициент корреляции имеют высокие значения.

6. Сопоставим фактические значения бинарной переменной *Y* с расчетными значениями *Pi*. Сохраним остатки модели в базе данных: В окне модели: Сохранить/Остатки. Затем создадим матрицу *YP*: Добавить/Добавить матрицу (рис. 1.4).

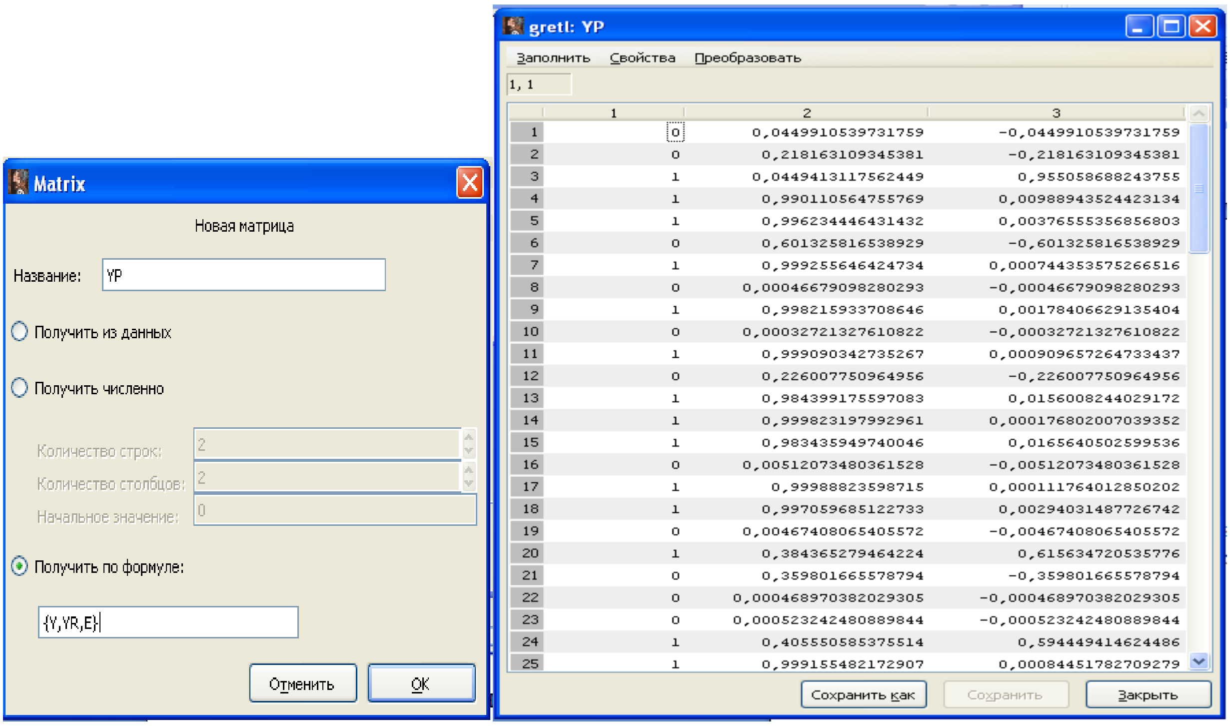


Рис. 1.4. Окно добавления матрицы *YR*

Попарное сравнение фактических значений бинарной переменной *Y* с расчетными значениями *Pi* позволило сделать следующий вывод: В 94 случаях из 100 удалось точно предсказать стратегию поведения фирмы в отношении сотрудников (предоставлять или нет право бесплатного доступа к ресурсам Интернет).

7. Используем построенную логит-модель для выбора среди имеющихся претендентов тех, кому в первую очередь следует предоставить право бесплатного доступа к ресурсам Интернета.

$$1)\ \hat{P}_i = \frac{1}{1+e^{-y_t^*}} = \frac{1}{1+e^{-\left(\beta_0+\sum\limits_{j=1}^{k}\beta_i x_{ij}\right)}} = \frac{1}{1+e^{11{,}3-0{,}378\cdot 27+0{,}454\cdot 3+1{,}104\cdot 3200-1{,}576\cdot 9}} =$$

$$=0{,}99973;$$

$$2)\ \hat{P}_i = \frac{1}{1+e^{-y_t^*}} = \frac{1}{1+e^{-\left(\beta_0+\sum\limits_{j=1}^{k}\beta_i x_{ij}\right)}} = \frac{1}{1+e^{11{,}3-0{,}378\cdot 44+0{,}454\cdot 12+1{,}104\cdot 5600-1{,}576\cdot 2}} =$$

$$=0{,}04049;$$

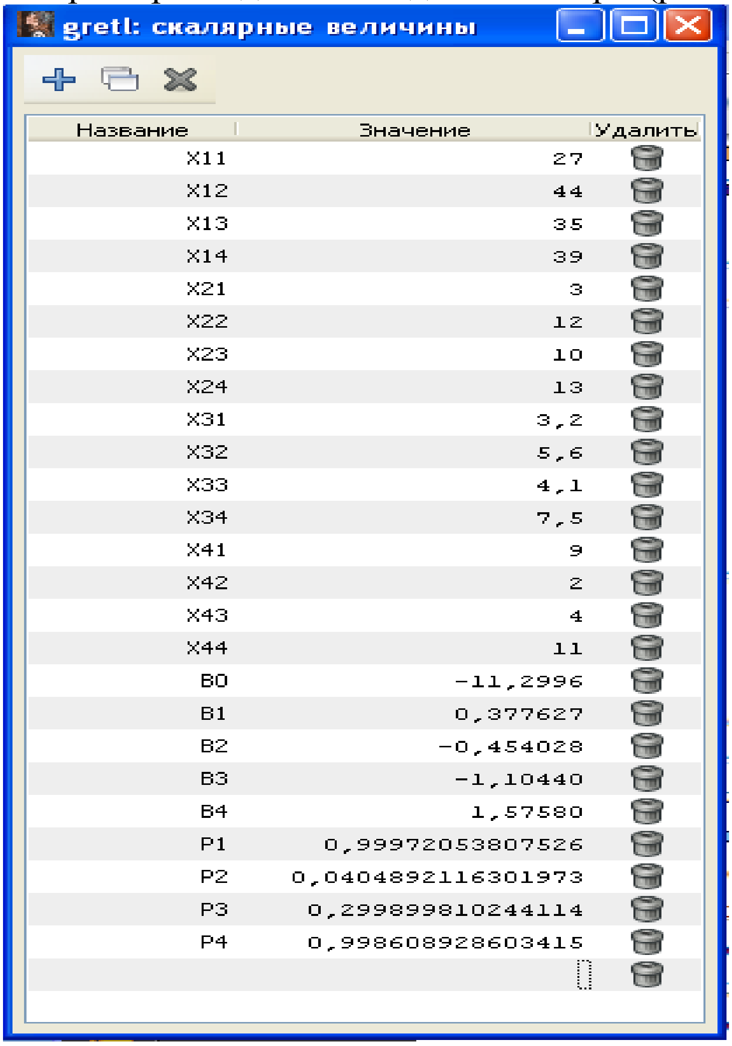
$$3)\ \hat{P}_i = \frac{1}{1+e^{-y_t^*}} = \frac{1}{1+e^{-\left(\beta_0+\sum\limits_{j=1}^{k}\beta_i x_{ij}\right)}} = \frac{1}{1+e^{11{,}3-0{,}378\cdot 35+0{,}454\cdot 10+1{,}104\cdot 4100-1{,}576\cdot 4}} =$$

$$=0{,}29990;$$

$$4)\ \hat{P}_i = \frac{1}{1+e^{-y_t^*}} = \frac{1}{1+e^{-\left(\beta_0+\sum\limits_{j=1}^{k}\beta_i x_{ij}\right)}} = \frac{1}{1+e^{11{,}3-0{,}378\cdot 39+0{,}454\cdot 13+1{,}104\cdot 7500-1{,}576\cdot 11}} =$$

$$=0{,}99861.$$

Оценки коэффициентов модели и значения факторов $X1$, $X2$, $X3$, $X4$ для каждого из четырех претендентов введем в скаляры (рис. 1.5).



gretl: скалярные величины

Название	Значение	Удалить
X11	27	
X12	44	
X13	35	
X14	39	
X21	3	
X22	12	
X23	10	
X24	13	
X31	3,2	
X32	5,6	
X33	4,1	
X34	7,5	
X41	9	
X42	2	
X43	4	
X44	11	
B0	-11,2996	
B1	0,377627	
B2	-0,454028	
B3	-1,10440	
B4	1,57580	
P1	0,99972053807526	
P2	0,0404892116301973	
P3	0,29899810244114	
P4	0,998608928603415	

Рис. 1.5. Окно ввода скалярных величин

$$P1 = 1/\left(1 + exp\left(-B0 - B1\cdot X11 - B2\cdot X21 - B3\cdot X31 - B4\cdot X41\right)\right);$$

$$P2 = 1/\left(1 + exp\left(-B0 - B1\cdot X12 - B2\cdot X22 - B3\cdot X32 - B4\cdot X42\right)\right);$$

$$P3 = 1/\left(1 + exp\left(-B0 - B1\cdot X13 - B2\cdot X23 - B3\cdot X33 - B4\cdot X43\right)\right);$$

$$P4 = 1/\left(1 + exp\left(-B0 - B1\cdot X14 - B2\cdot X24 - B3\cdot X34 - B4\cdot X44\right)\right).$$

Следовательно, первому и четвертому претендентам целесообразно предоставить право бесплатного доступа, а второму и третьему нет.

8. Построение пробитовой модели: Модель/Нелинейные модели/Пробит/Бинарный… (рис. 1.6).

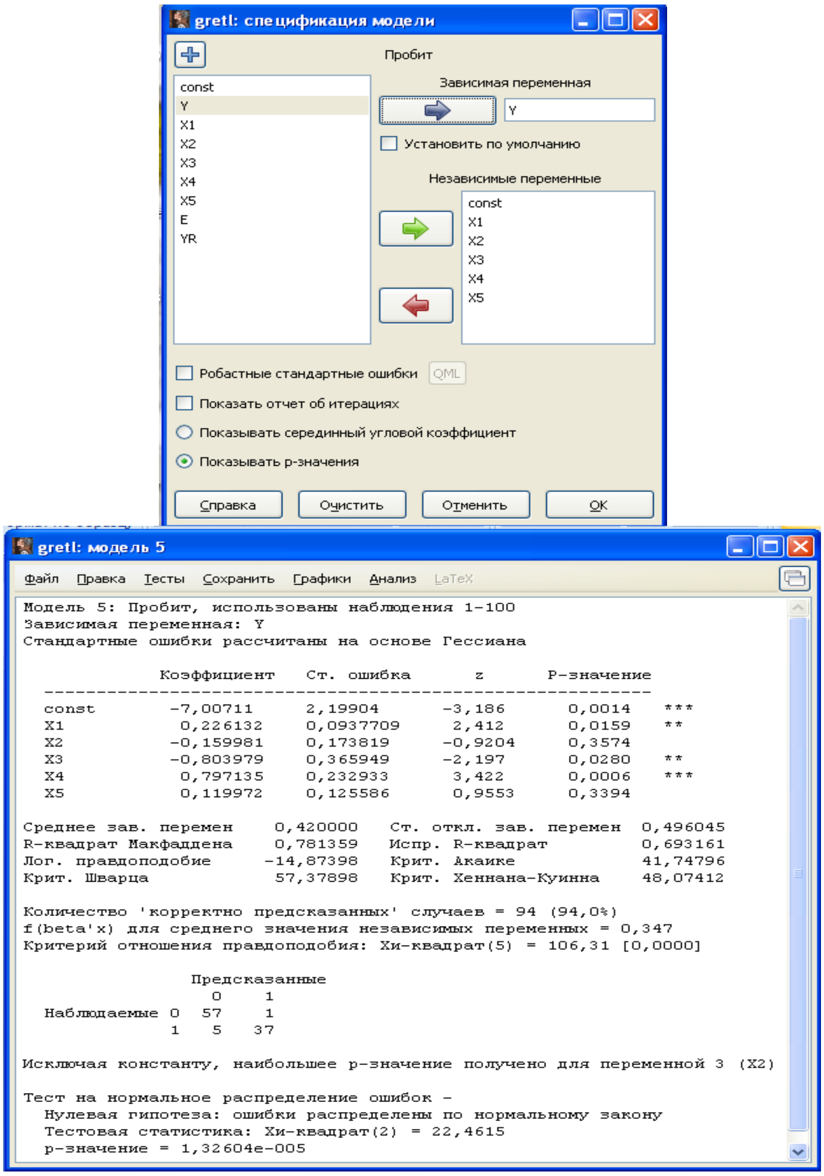
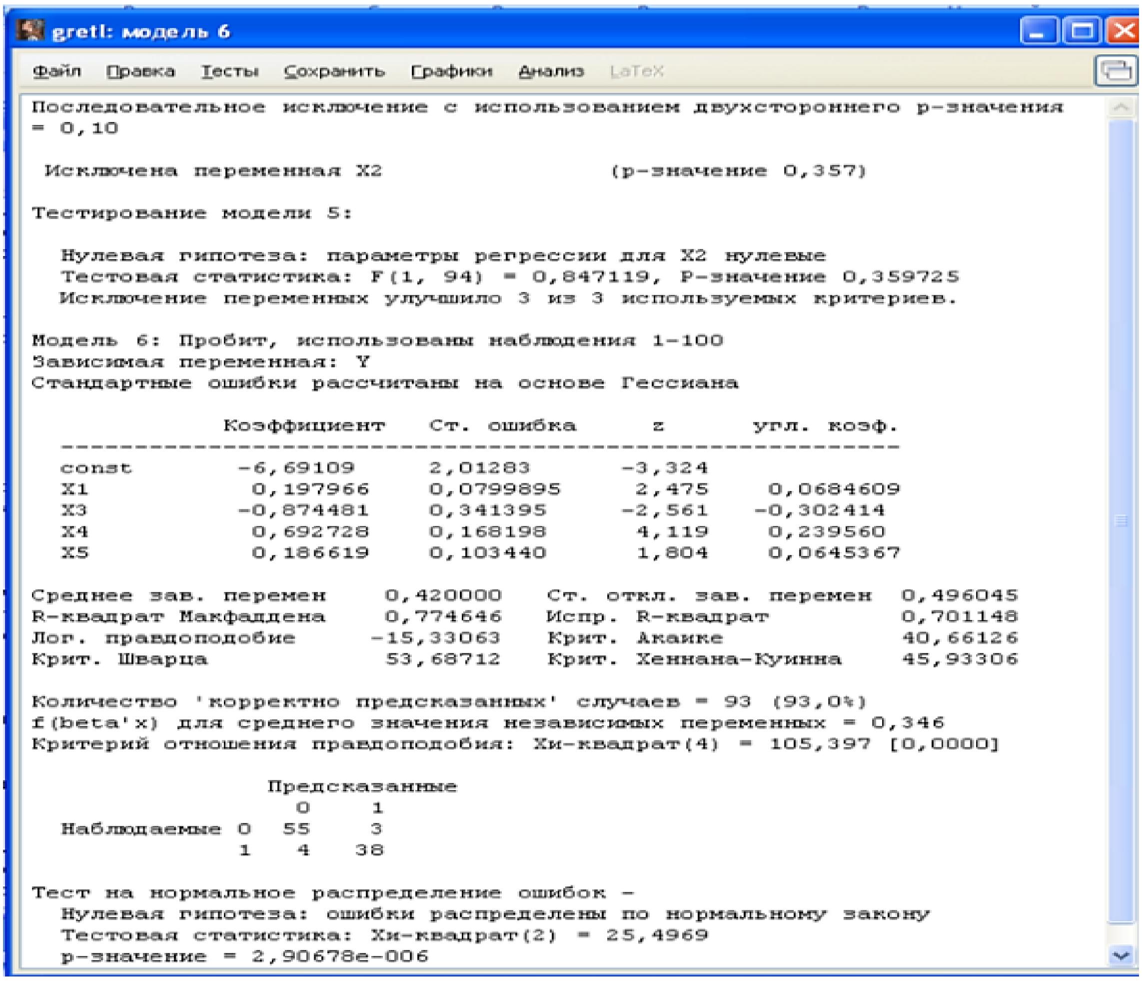


Рис. 1.6. Окно построения пробит-модели

В предпоследнем столбце окна результатов приведено отношение оценок к среднему значению, которое может интерпретироваться как направление изменения вероятностей для всех переменных. Параметры направления (z) свидетельствуют, что при увеличении стажа профессиональной деятельности ($X2$) и заработной платы ($X3$) – отрицательный знак параметра, вероятность принятия решения о целесообразности предоставления сотруднику права бесплатного доступа к ресурсам Интернета снижается. При увеличении таких факторов как возраст, результаты тестирования, количество случаев нахождения полезной информации – вероятность увеличивается. Тест Стьюдента свидетельствует, что факторы $X2$, $X5$ являются несущественными.

9. Выполним последовательное исключение несущественных переменных: В окне модели: Тесты/Избыточные переменные/Последовательное исключение переменных (рис. 1.7).



```
gretl: модель 6
Файл  Правка  Тесты  Сохранить  Графики  Анализ  LaTeX
Последовательное исключение с использованием двухстороннего p-значения
= 0,10

 Исключена переменная X2               (p-значение 0,357)

Тестирование модели 5:

  Нулевая гипотеза: параметры регрессии для X2 нулевые
  Тестовая статистика: F(1, 94) = 0,847119, P-значение 0,359725
  Исключение переменных улучшило 3 из 3 используемых критериев.

Модель 6: Пробит, использованы наблюдения 1-100
Зависимая переменная: Y
Стандартные ошибки рассчитаны на основе Гессиана

             Коэффициент   Ст. ошибка      z       угл. коэф.
  ----------------------------------------------------------
  const        -6,69109     2,01283      -3,324
  X1            0,197966    0,0799895     2,475    0,0684609
  X3           -0,874481    0,341395     -2,561   -0,302414
  X4            0,692728    0,168198      4,119    0,239560
  X5            0,186619    0,103440      1,804    0,0645367

Среднее зав. перемен    0,420000   Ст. откл. зав. перемен  0,496045
R-квадрат Макфаддена    0,774646   Испр. R-квадрат         0,701148
Лог. правдоподобие    -15,33063    Крит. Акаике            40,66126
Крит. Шварца           53,68712    Крит. Хеннана-Куинна    45,93306

Количество 'корректно предсказанных' случаев = 93 (93,0%)
f(beta'x) для среднего значения независимых переменных = 0,346
Критерий отношения правдоподобия: Хи-квадрат(4) = 105,397 [0,0000]

                Предсказанные
                   0     1
  Наблюдаемые 0   55     3
              1    4    38

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 25,4969
  p-значение = 2,90678e-006
```

Рис. 1.7. Окно исключения несущественных переменных

Запишем модель:

$$Pi = F\left(\beta_0 + \sum_{j=1}^{k} \beta_i x_{ij}\right) = F\left(-6,691 + 0,198X_1 - 0,874X_3 + 0,693X_4 + 0,187X_5\right).$$

О значимости совокупной регрессии свидетельствует Хи-квадрат (105,397) и его *p*-значение (0,0000<0,05).

10. Используем построенную модель для выбора среди имеющихся претендентов тех, кому в первую очередь следует предоставить право бесплатного доступа к ресурсам Интернета. Значения коэффициентов модели (*A*0, *A*1, *A*3, *A*4, *A*5), фактора *X*51, *X*52, *X*53, *X*54 введем в скаляры и выполним расчет:

1) $F1 = A0 + A1 \cdot X11 + A3 \cdot X31 + A4 \cdot X41 + A5 \cdot X51$;
2) $F2 = A0 + A1 \cdot X12 + A3 \cdot X32 + A4 \cdot X42 + A5 \cdot X52$;
3) $F3 = A0 + A1 \cdot X13 + A3 \cdot X33 + A4 \cdot X43 + A5 \cdot X53$;
4) $F4 = A0 + A1 \cdot X14 + A3 \cdot X34 + A4 \cdot X44 + A5 \cdot X54$.

Затем, Инструменты/Поиск *p*-значения (рис. 1.8).

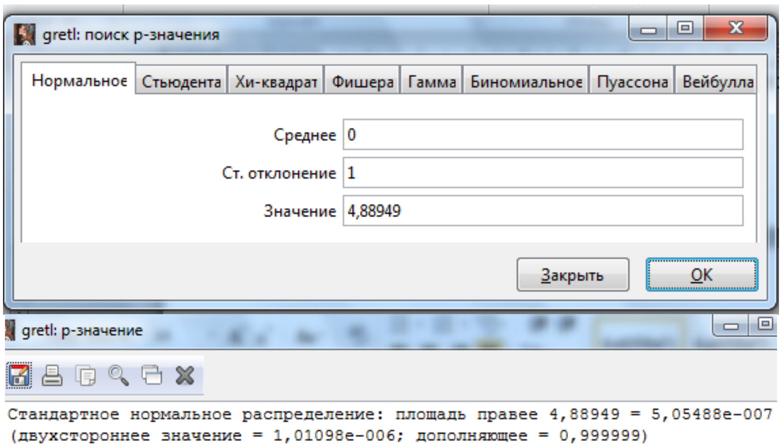


Рис. 1.8. Окно функции стандартизованного нормального распределения

$$Pi = F(-6{,}691 + 0{,}198 \cdot 27 - 0{,}874 \cdot 3{,}2 + 0{,}693 \cdot 9 + 0{,}187 \cdot 15) = F(4{,}88949) =$$
$$= 0{,}999999,$$
$$Pi = F(-6{,}691 + 0{,}198 \cdot 44 - 0{,}874 \cdot 5{,}6 + 0{,}693 \cdot 2 + 0{,}187 \cdot 5) = F(-0{,}5591286) =$$
$$= 0{,}423926,$$
$$Pi = F(-6{,}691 + 0{,}198 \cdot 35 - 0{,}874 \cdot 4{,}1 + 0{,}693 \cdot 4 + 0{,}187 \cdot 7) = F(0{,}729593) =$$
$$= 0{,}534361,$$
$$Pi = F(-6{,}691 + 0{,}198 \cdot 39 - 0{,}874 \cdot 7{,}5 + 0{,}693 \cdot 11 + 0{,}187 \cdot 15) = F(4{,}89027) =$$
$$= 0{,}999999.$$

Таким образом, право доступа к ресурсам Интернет можно предоставить первому и четвертому претендентам. Итоговые результаты моделирования приведены в таблице 3.2.

Таблица 1.2

Сводная таблица результатов моделирования

Меры соответствия эмпирическим данным	Логит-модель	Пробит-модель
Хи-квадрат	105,292	105,397
R-квадрат	0,700	0,701
Количество корректно предсказанных случаев	94	93

Задания для самостоятельного выполнения

Задание 1.2. ОАО «Нанотехнологии», занимающееся, в частности, проблемами макро- и микромоделирования, обеспечивает своим сотрудникам, стажировки в англоязычные страны. При отборе претендентов на поездку учитываются такие факторы, как: тест на знание английского языка ($X1$, баллов), стаж профессиональной деятельности ($X2$, лет), ученая степень ($X3$: 00 –без степени, 01 – кандидат наук, 10 – доктор наук), количество публикаций по указанной проблематике. В настоящее время руководству ОАО снова предстоит решить, кого из сотрудников отправить на стажировку. Для того, чтобы выбор претендента осуществлялся исключительно на научной основе, аналитику ОАО было поручено построить модель, позволяющую реализовать эти цели. В качестве основы построения модели было решено использовать бинарную переменную Y: $Y = 1$, если стажер вернулся с дипломом, $Y = 0$, если стажер вернулся с сертификатом. Данные для построения модели представлены в таблице 3.3.

Таблица 1.3

№	Y	$X1$	$X2$	$X3$	$X4$	№	Y	$X1$	$X2$	$X3$	$X4$
1	1	39	7	канд.	15	51	1	27	11	нет ст.	4
2	0	7	2	канд.	7	52	0	16	7	нет ст.	7
3	0	27	4	нет ст.	2	53	0	21	7	нет ст.	2
4	0	29	5	нет ст.	1	54	1	36	14	нет ст.	6
5	0	34	4	нет ст.	3	55	0	21	6	нет ст.	9
6	1	18	11	канд.	18	56	1	41	17	канд.	31
7	1	21	13	канд.	17	57	0	25	7	нет ст.	8
8	1	40	9	канд.	13	58	1	33	12	нет ст.	5
9	1	28	11	нет ст.	6	59	0	19	9	нет ст.	8
10	1	24	11	канд.	16	60	0	13	1	нет ст.	8
11	0	12	5	нет ст.	9	61	0	14	4	нет ст.	7
12	1	25	13	канд.	20	62	1	28	13	канд.	21
13	1	20	16	доктор	43	63	1	29	13	нет ст.	6
14	1	30	15	нет ст.	5	64	1	27	21	канд.	13
15	0	18	7	нет ст.	3	65	1	29	23	канд.	14
16	1	18	11	канд.	16	66	1	31	21	канд.	15
17	1	39	18	канд.	30	67	0	27	5	нет ст.	3
18	1	24	15	канд.	19	68	0	29	4	нет ст.	2
19	0	4	16	нет ст.	5	69	0	27	3	нет ст.	2
20	1	46	15	доктор	45	70	0	33	5	нет ст.	3
21	0	9	5	канд.	8	71	1	31	15	канд.	28
22	1	16	11	канд.	17	72	1	36	11	канд.	21
23	1	24	7	канд.	14	73	0	36	2	нет ст.	1
24	0	20	9	нет ст.	7	74	0	31	4	нет ст.	2
25	0	3	18	нет ст.	4	75	0	34	3	нет ст.	3
26	0	12	4	канд.	9	76	0	3	16	канд.	13
27	1	21	14	канд.	18	77	1	38	16	канд.	30
28	0	31	3	нет ст.	3	78	0	2	20	канд.	15
29	0	4	20	канд.	13	79	0	3	16	нет ст.	4
30	0	34	3	нет ст.	3	80	0	4	19	канд.	14

31	0	12	8	канд.	6	81	0	9	7	канд.	4
32	0	8	4	канд.	7	82	0	9	8	канд.	4
33	0	3	17	канд.	13	83	0	11	6	канд.	6
34	0	10	7	канд.	5	84	0	5	18	канд.	14
35	1	26	11	канд.	16	85	0	4	17	нет ст.	5
36	1	30	21	канд.	13	86	0	4	16	нет ст.	6
37	0	36	2	нет ст.	1	87	0	2	17	канд.	13
38	0	19	7	нет ст.	7	88	0	2	16	нет ст.	4
39	0	36	5	нет ст.	3	89	1	16	9	канд.	13
40	1	46	19	канд.	37	90	0	3	19	канд.	15
41	0	13	1	канд.	7	91	0	12	7	нет ст.	3
42	1	36	14	канд.	26	92	0	13	7	нет ст.	3
43	0	21	10	нет ст.	9	93	0	3	16	канд.	14
44	1	20	11	канд.	19	94	1	19	8	канд.	14
45	0	36	4	нет ст.	3	95	1	18	8	канд.	15
46	0	14	8	нет ст.	3	96	0	5	16	канд.	13
47	0	31	4	нет ст.	2	97	1	16	6	канд.	13
48	0	7	5	нет ст.	6	98	0	4	20	канд.	15
49	0	33	3	нет ст.	1	99	1	17	7	канд.	13
50	0	14	1	канд.	8	100	0	3	18	канд.	15

Потенциальные стажеры обладают следующими характеристиками:

Результат тестирования – 27 баллов, стаж профессиональной деятельности – 3 года, кандидат наук, количество публикаций – 21;

Результат тестирования – 48 баллов, стаж профессиональной деятельности – 12 лет, без степени, количество публикаций – 10;

Результат тестирования – 32 балла, стаж профессиональной деятельности – 25 лет, доктор наук, количество публикаций – 98.

Используя построенную прогнозную модель, определить среди имеющихся претендентов тех, кому в первую очередь следует предоставить право пройти зарубежную стажировку.

Задание 1.3. Банк исследует вероятность невозврата кредита предприятиями торговли ($Y=1$ – заемщик кредит возвращает, $Y=0$ – не возвращает), используя пять факторов: $X1$ – коэффициент финансовой зависимости, $X2$ – коэффициент рентабельности, $X3$ – коэффициент финансовой устойчивости, $X4$ – коэффициент оборачиваемости оборотных активов, $X5$ – коэффициент покрытия активов. Для того, чтобы обосновать выбор заемщиков, постройте логит- и пробит- модели. Данные по 50 предприятиям торговли для построения модели представлены в таблице 3.4.

Таблица 1.4

Y	$X1$	$X2$	$X3$	$X4$	$X5$	Y	$X1$	$X2$	$X3$	$X4$	$X5$
1	0,74	0,055	0,811	2,17	2,29	0	0,84	0,125	0,74	5,61	1,65
0	0,55	0,055	0,761	1,63	1,29	0	0,91	0,125	0,791	5,12	1,15
1	0,59	0,065	0,786	14,26	2,08	1	0,63	0,135	0,817	3,24	2,43
1	0,79	0,065	0,841	1,63	2,29	0	0,79	0,135	0,783	9,02	1,13
1	0,57	0,065	0,758	9,47	2,19	1	0,63	0,145	0,753	14,37	1,39
1	0,64	0,065	0,844	6,53	2,77	1	0,68	0,145	0,75	9,78	2,36
1	0,75	0,065	0,808	6,39	1,98	1	0,57	0,155	0,762	11,16	1,88
1	0,57	0,075	0,85	1,27	2,11	1	0,61	0,155	0,768	7,61	2,48
1	0,67	0,075	0,81	14,27	2,78	1	0,55	0,175	0,783	14,64	2,26
1	0,58	0,075	0,816	11,93	2,54	1	0,62	0,175	0,784	5,98	2,6
1	0,66	0,075	0,787	10,63	2,26	1	0,62	0,175	0,811	3,57	2,57
1	0,75	0,075	0,776	10,64	2	1	0,63	0,175	0,827	3,63	2,56
1	0,67	0,075	0,779	2,63	1,98	1	0,63	0,185	0,824	7,64	1,74
1	0,58	0,085	0,747	5,37	2,58	1	0,79	0,185	0,842	7,17	2,72
1	0,74	0,085	0,753	0,63	2,04	1	0,61	0,185	0,843	12,27	2,69
1	0,61	0,085	0,768	3,27	2,09	1	0,53	0,195	0,78	6,74	2,78
0	0,89	0,085	0,778	4,36	2,05	1	0,63	0,195	0,808	8,3	2,68
1	0,71	0,095	0,795	12,73	1,99	1	0,56	0,195	0,756	2,16	1,65
0	0,54	0,095	0,751	2,63	1,43	1	0,61	0,195	0,762	6,47	2,62
0	0,54	0,095	0,746	4,03	1,58	1	0,55	0,215	0,757	2,63	2,85
1	0,69	0,105	0,818	14,19	2	1	0,89	0,225	0,816	7,48	2,71
0	0,62	0,105	0,784	2,56	1,25	1	0,58	0,225	0,846	10,55	2,29
0	0,78	0,105	0,753	3,03	2,27	1	0,58	0,235	0,818	1,62	2,28
0	0,72	0,105	0,782	4,14	1,11	1	0,69	0,255	0,837	15,47	2,44
0	0,87	0,115	0,765	5,37	1,44	1	0,69	0,265	0,821	15,37	2,05

Потенциальные заемщики обладают следующими финансовыми коэффициентами, характеризующими их кредитоспособность:

1) $X1 = 0{,}534$; $X2 = 0{,}068$; $X3 = 0{,}762$; $X4 = 3{,}234$; $X5 = 2{,}106$;

2) $X1 = 0{,}670$; $X2 = 0{,}098$; $X3 = 0{,}773$; $X4 = 1{,}480$; $X5 = 1{,}890$;

3) $X1 = 0{,}375$; $X2 = 0{,}102$; $X3 = 0{,}810$; $X4 = 6{,}653$; $X5 = 1{,}172$.

Задание 1.4. Банк исследует вероятность невозвращения потребительского кредита (Y=1 – заемщик кредит возвращает, Y=0 – не возвращает), используя два фактора: $X1$ – сумма займа, $X2$ – среднемесячный доход заемщика. По логит-модели:

$$P(Y=1)=\frac{e^{(5-0,6(X1/X2))}}{1+e^{(5-0,6(X1/X2))}}$$

оцените вероятность невозвращения кредита при покупке на сумму 40 тыс. руб. и доходе 10 тыс. руб. Повторите расчет при стоимости покупки в 50 тыс. руб. и доходе 5 тыс. руб. Дайте рекомендацию банку о пороговом соотношении суммы займа и среднемесячного дохода, чтобы предсказанная по модели доля просроченных кредитов не превышала 5%.

Контрольные вопросы:

1. В каких ситуациях фиктивная переменная используется в качестве зависимой переменной?

2. Какие законы распределения чаще всего используются в моделях бинарного выбора?

3. В чем суть логит-модели и пробит –модели?

4. Как осуществляется проверка значимости коэффициентов в модели бинарного выбора?

5. Как получить прогноз вероятности по логит-модели и пробит-модели?

6. Можно ли рассчитать по логит-модели коэффициент детерминации?

7. Какие существуют варианты постановки моделей множественного выбора?

8. В чем отличие моделей упорядоченного и неупорядоченного выбора?

2. Модели анализа панельных данных

Расчетные формулы

Объединенная модель:

$$Y_{it}=\alpha+X_{it}\beta+\varepsilon_{it}.$$

Модель с фиксированными эффектами (fixed effects model):

$$Y_{it}=\alpha_i+X_{it}\beta+\varepsilon_{it},\alpha_i=z_i\alpha.$$

Модель со случайными эффектами (random effects model):

$$Y_{it}=X_{it}\beta+\alpha+m_i+\varepsilon_{it}.$$

Проверка на наличие фиксированных эффектов выполняется с помощью распределения Фишера.

$$F = \frac{RSS_{pool} - RSS_{FE}}{RSS_{FE}} \cdot \frac{nT - n - d}{n-1} = \frac{R^2{}_{FE} - R^2{}_{pool}}{1 - R^2{}_{FE}} \cdot \frac{nT - n - d}{n-1},$$

$$F > F(\alpha, n-1, nT - n - d) \Rightarrow H_1 : \alpha_i \neq 0.$$

МНК-оценки в модели со случайными эффектами неэффективны из-за присутствия автокорреляции в слагаемом ошибки m_i. Применяется двухшаговая процедура обобщенного метода наименьших квадратов – ВОМНК – выполнимый обобщенный метод наименьших квадратов.

Вводится вспомогательная переменная:

$$\theta = 1 - \frac{\sigma^2{}_V}{\sqrt{\sigma_V^2 + T\sigma_M^2}} = 1 - \left(\frac{1}{1 + T\left(\dfrac{\sigma_M^2}{\sigma_V^2} \right)} \right)^{\frac{1}{2}}.$$

Поскольку на практике дисперсии не известны, заменяем их на состоятельные оценки:

$$\theta = 1 - \frac{\hat{\sigma}_V^2}{\sqrt{\hat{\sigma}_V^2 + T\hat{\sigma}_M^2}}.$$

ВОМНК-оценка модели со случайным эффектом:

$$(y_{it} - \theta \overline{y}_i) = \mu(1-\theta) + (x_{it} - \theta \overline{x}_i)' \beta + (u_{it} - \theta \overline{u}_i),$$

$$\hat{\sigma}_A^2 = \frac{1}{n - d - 1} \sum_{i=1}^{N} \hat{u}_i{}^2,$$

$$\hat{\sigma}_M^2 = \hat{\sigma}_A^2 - \frac{1}{T} \hat{\sigma}_V^2.$$

Проверка на наличие случайных эффектов выполняется с помощью тестовой статистики Хаусмана. Тестируется нулевая гипотеза об отсутствии корреляции между индивидуальными эффектами и регрессорами (наличие случайных эффектов). Для проверки нулевой гипотезы используется тест Хаусмана. Определяется наблюдаемое значение статистики $Q_{\text{Н}}$:

$$Q_H = \left(\hat{\beta}_{внутр} - \hat{\beta}_{FGLS} \right)' \left[\hat{V}(\hat{\beta}_{внутр}) - V(\hat{\beta}_{FGLS}) \right]^{-1} \left(\hat{\beta}_{внутр} - \hat{\beta}_{FGLS} \right),$$

где $\hat{\beta}_{внутр}$ – внутригрупповая оценка;

$\hat{\beta}_{FGLS}$ – оценка доступного обобщенного метода наименьших квадратов.

В случае если Q_H-статистика больше, чем критическое значение χ2-распределения с k_w степенями свободы, где k_w – число регрессоров во внутригрупповой модели, то можно отклонить нулевую гипотезу и сделать выбор в пользу модели с фиксированными эффектами.

Задание для выполнения в аудитории

Задание 2.1. Учредитель сети супермаркетов «Пятерочка» с целью изыскания путей увеличения годового товарооборота (*Y*, млн. руб.), поручил специалистам компании изучить факторы, влияющие на этот показатель, в четырех регионах России. В ходе исследования было выявлено, что такими факторами являются: торговая площадь (*X*1, тыс. кВ. м), среднее число посетителей в день (*X*2, тыс. чел.) и сформирована таблица. В таблице 2.1 представлены панельные данные для моделирования. Необходимо построить модели анализа панельных данных, отражающие гетерогенность товарооборота в разрезе территорий.

Таблица 2.1

№	*Y*	*X*1	*X*2	№	*Y*	*X*1	*X*2
Ульяновская область				Саратовская область			
1	20,76	0,24		15	65,01	0,94	10,36
2	28,09	0,31		16	69,05	1,21	11,36
3	32,95	0,55		17	73,13	1,29	8,89
4	38,15	0,67		18	81,18	1,49	7,55
5	46,78	0,83		19	89,24	1,67	7,81
6	55,31	0,98		20	97,30	1,84	8,08
7	60,92	1,14		21	115,36	2,02	11,84
Оренбургская область				Самарская область			
8	41,08	0,45	1,45	22	91,26	1,12	10,72
9	56,29	0,78	2,02	23	99,84	1,29	11,27
10	68,51	0,98	3,77	24	108,55	1,49	13,02
11	82,72	1,24	5,52	25	117,17	1,67	13,41
12	96,43	1,49	7,51	26	125,81	1,85	13,62
13	110,15	1,74	9,04	27	134,46	2,04	14,34
14	123,86	1,99	12,01	28	143,10	2,22	14,85

Методические указания для выполнения задания

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «Занятие_Панели.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1.

Интерпретировать данные как панельные (рис. 2.1).

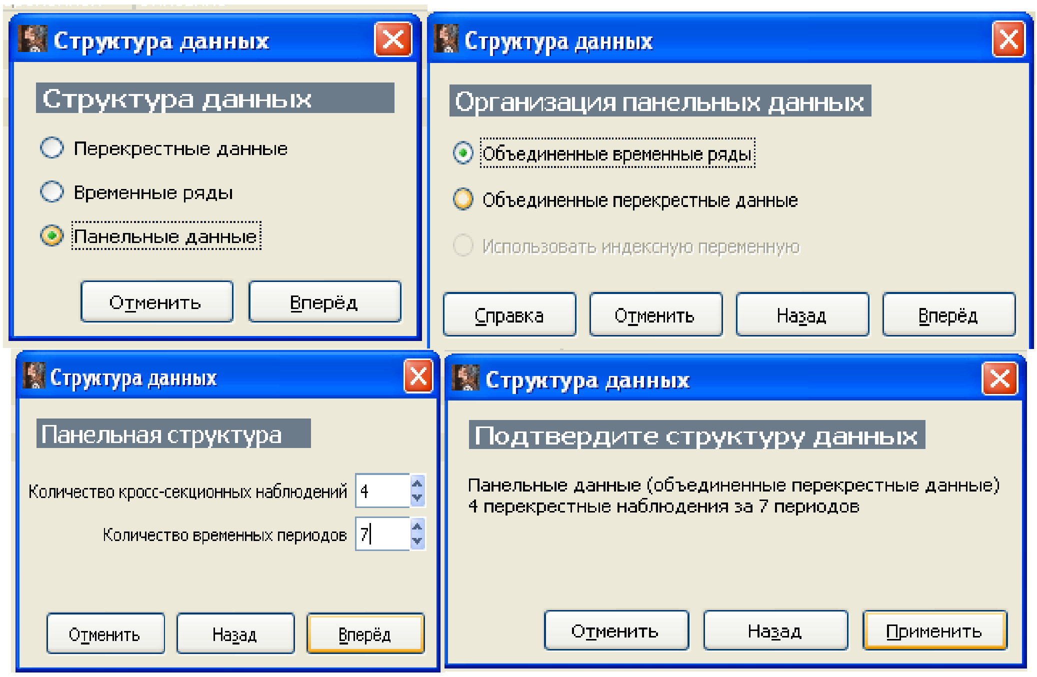


Рис. 2.1. Окно импорта данных

2. Построение регрессионной модели со свободным коэффициентом: Модель/Метод наименьших квадратов (рис. 2.2).

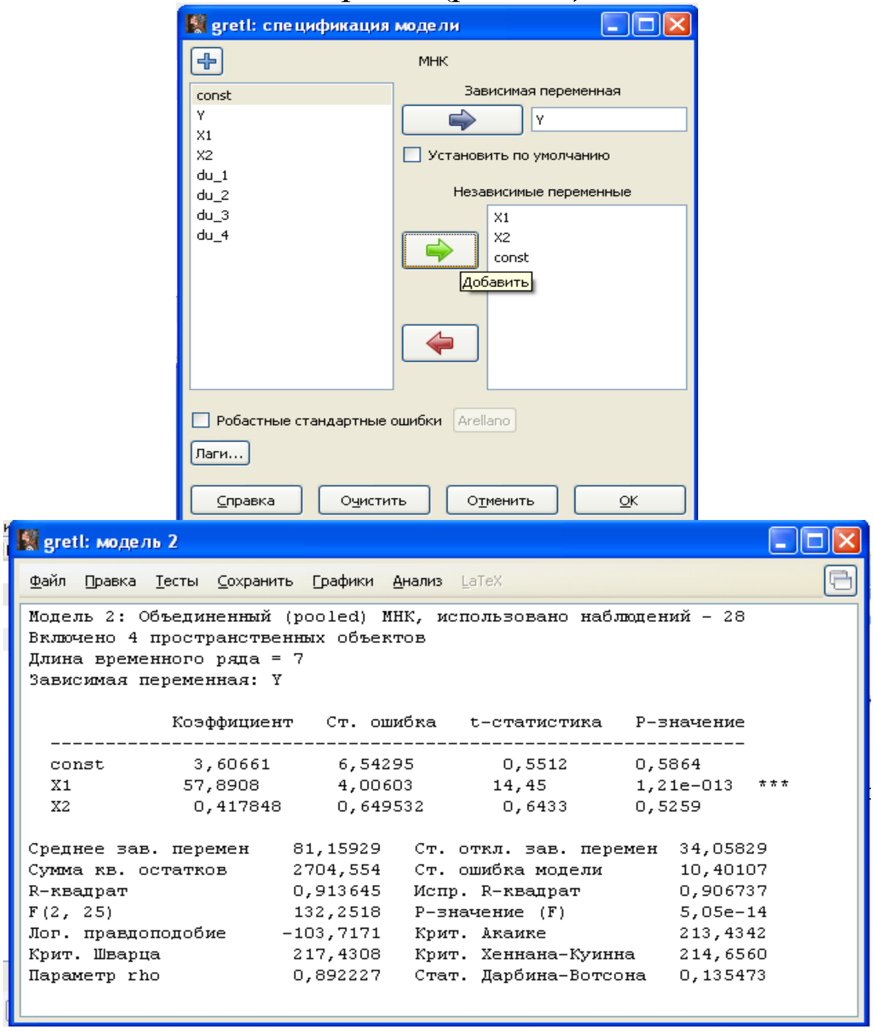
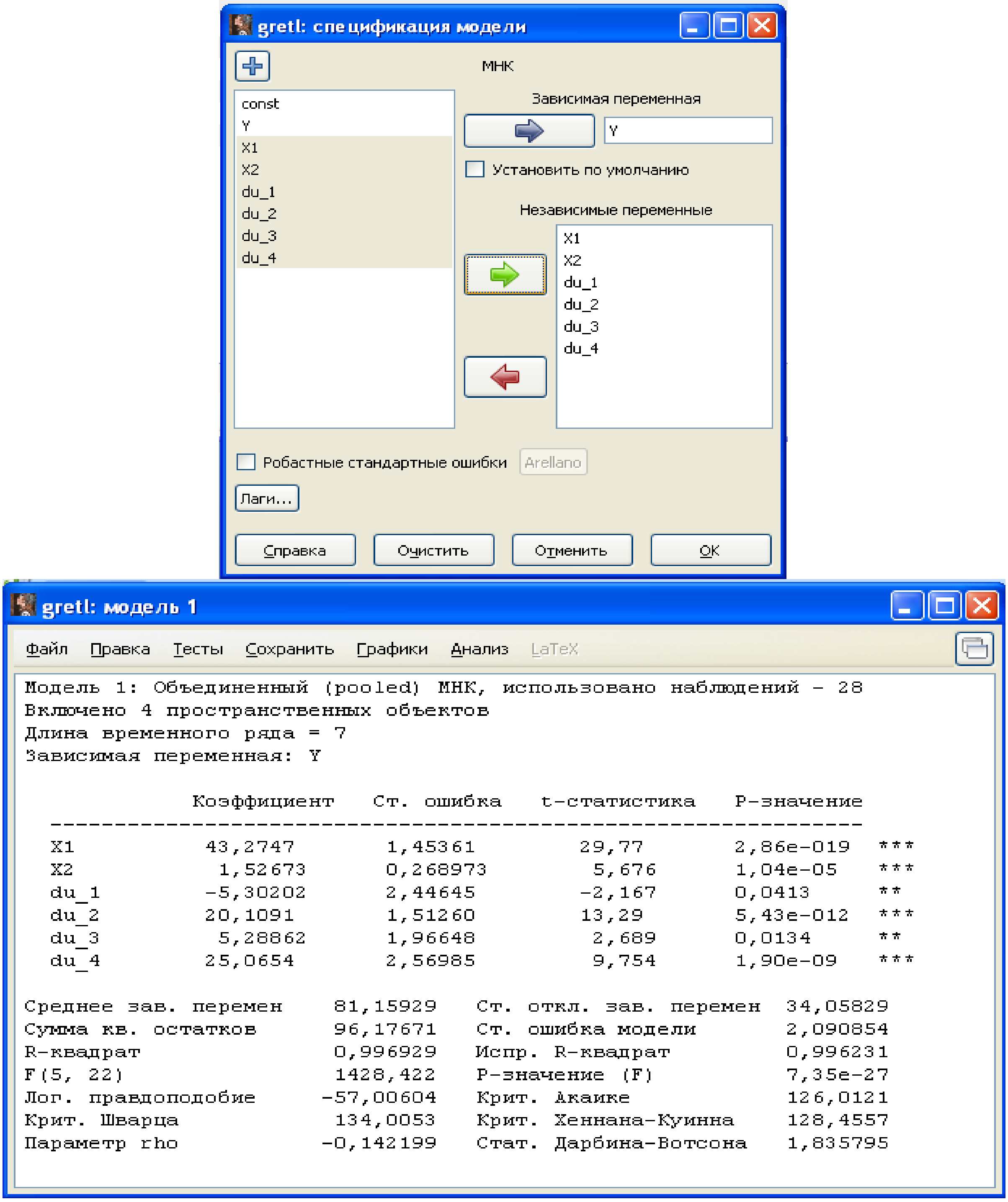


Рис. 2.2. Окно регрессионной модели со свободным коэффициентом

$$Y = 3{,}606 + 57{,}891X1 + 0{,}418X2,\ R^2 = 0{,}913645.$$

3. Добавление фиктивных переменных-фильтров du_1, du_2, du_3, du_4: Добавить/Единичную фиктивную переменную (панельные данные).

4. Построение регрессионной модели с фиксированными эффектами, без свободного коэффициента: Модель/Метод наименьших квадратов (рис. 2.3).



Модель 1: Объединенный (pooled) МНК, использовано наблюдений - 28
Включено 4 пространственных объектов
Длина временного ряда = 7
Зависимая переменная: Y

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
X1	43,2747	1,45361	29,77	2,86e-019	***
X2	1,52673	0,268973	5,676	1,04e-05	***
du_1	-5,30202	2,44645	-2,167	0,0413	**
du_2	20,1091	1,51260	13,29	5,43e-012	***
du_3	5,28862	1,96648	2,689	0,0134	**
du_4	25,0654	2,56985	9,754	1,90e-09	***

Среднее зав. перемен	81,15929	Ст. откл. зав. перемен	34,05829
Сумма кв. остатков	96,17671	Ст. ошибка модели	2,090854
R-квадрат	0,996929	Испр. R-квадрат	0,996231
F(5, 22)	1428,422	Р-значение (F)	7,35e-27
Лог. правдоподобие	-57,00604	Крит. Акаике	126,0121
Крит. Шварца	134,0053	Крит. Хеннана-Куинна	128,4557
Параметр rho	-0,142199	Стат. Дарбина-Вотсона	1,835795

Рис. 2.3. Окно регрессионной модели с фиксированными эффектами

$$Yx = -5{,}302i1 + 20{,}109i2 + 5{,}29i3 + 25{,}065i4 + 43{,}275X1 + 1{,}527X2,$$
$$R^2 = 0{,}9969.$$

5. Проверка гипотезы об отсутствии фиксированных групповых эффектов. Пусть $v1 = 4-1$, $v2 = 4\cdot 7 - 4 - 2$. Введем скаляры: $v1 = 3$, $v2 = 22$, $R1 = 0,9969$, $R2 = 0,913645$. Скаляр: $F = (R1/v1)/(R2/v2)$. Затем определим критическое значение: Инструменты/Критические значения (рис. 2.4).

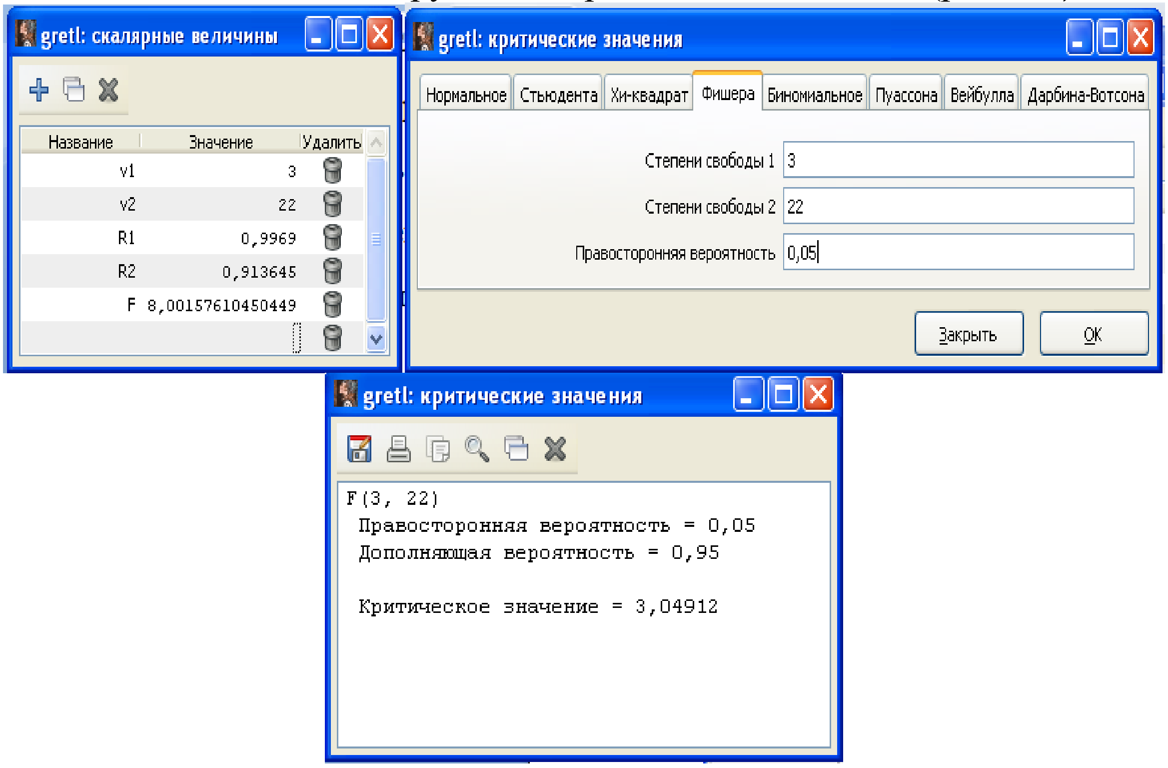


Рис. 2.4. Скаляры и критическое значение распределения Фишера

$F = 8,001 > F(0,05;3;22) = 3,049$. Значит, нулевую гипотезу об отсутствии фиксированных групповых эффектов следует отвергнуть. Следовательно, уравнение:

$$Yx = -5,302i1 + 20,109i2 + 5,29i3 + 25,065i4 + 43,275X1 + 1,527X2,$$

учитывающее групповые фиксированные эффекты, правомерно. Одной из главных причин этого, скорее всего является то, что на годовой товарооборот сети магазинов «Пятерочка» влияет различие в доходах населения в разных регионах.

6. Построение регрессионной модели со случайными эффектами.

6.1. Вычисляем средние значения *Y*, *X*1, *X*2 для каждой панели данных:

Находим частные подвыборки для каждой панели: $Y1 = Y\cdot du\_1$, $Y2 = Y\cdot du\_2$, $Y3 = Y\cdot du\_3$, $Y4 = Y\cdot du\_4$: Добавить/Добавить новую переменную. То же для *X*1, *X*2: $X11 = X1\cdot du\_1$, $X12 = X1\cdot du\_2$ и т.д.

Затем скаляры: $SY1 = sum(Y1)/7$ и т. д.

Используя обычные МНК-оценки (пункт 2), находим расчетное значение *Y* по средним значениям *X*1, *X*2. Вводим скаляры: $a = 3,60661$, $b1 = 57,89079$, $b2 = 0,417848$.

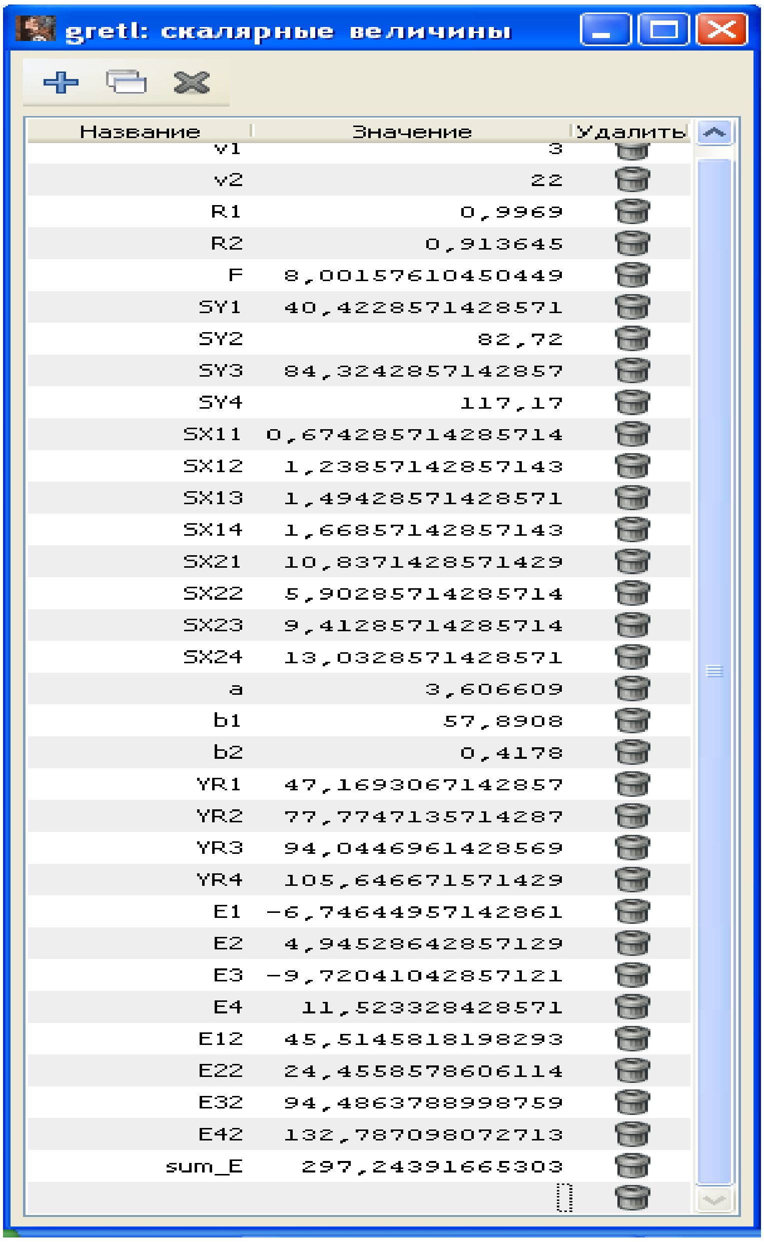
$YR1 = a + b1 \cdot SX11 + b2 \cdot SX21$ и т.д.

Находим остатки: $E1 = SY1 - YR1$, $E2 = SY2 - YR2$ и т. д.

Находим квадраты остатков: $E12 = E1 \cdot E1$ и т.д.

Находим сумму квадратов остатков:

$sum\_sq\_E = E12 + E22 + E32 + E42 \ = 297{,}243$ (рис. 2.5).


gretl: скалярные величины

Название	Значение	Удалить
v1	3	
v2	22	
R1	0,9969	
R2	0,913645	
F	8,00157610450449	
SY1	40,4228571428571	
SY2	82,72	
SY3	84,3242857142857	
SY4	117,17	
SX11	0,674285714285714	
SX12	1,23857142857143	
SX13	1,49428571428571	
SX14	1,66857142857143	
SX21	10,8371428571429	
SX22	5,90285714285714	
SX23	9,41285714285714	
SX24	13,0328571428571	
a	3,606609	
b1	57,8908	
b2	0,4178	
YR1	47,1693067142857	
YR2	77,7747135714287	
YR3	94,0446961428569	
YR4	105,646671571429	
E1	-6,74644957142861	
E2	4,94528642857129	
E3	-9,72041042857121	
E4	11,523328428571	
E12	45,5145818198293	
E22	24,455857860 6114	
E32	94,4863788998759	
E42	132,787098072713	
sum_E	297,24391665303	

Рис. 2.5. Окно скаляров

Вычисляем дисперсию σ^2_u.

Сначала находим остаточную дисперсию для модели с фиксированными эффектами: $96{,}176 / 22 = 4{,}3716$. Считаем дисперсию:

$297{,}243 - 4{,}3716 / 7 = 296{,}618$.

Выполним расчет параметра T:

$$T = 1 - \left(\frac{4{,}3716}{7 \cdot 296{,}6183 + 4{,}3716} \right)^{0{,}5} = 0{,}9541.$$

6.2. Преобразуем исходные данные.

Добавим скаляр: $T = 0{,}9541$, $S = 4{,}3716$.

Добавим новые переменные (рис. 2.6).

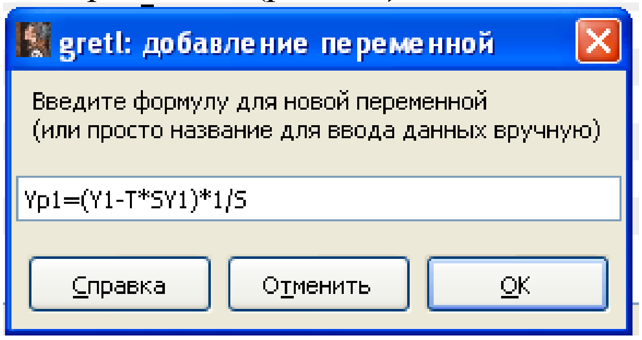


Рис. 2.6. Окно добавления новой переменной

Затем объединим частные подвыборки (рис. 2.7).

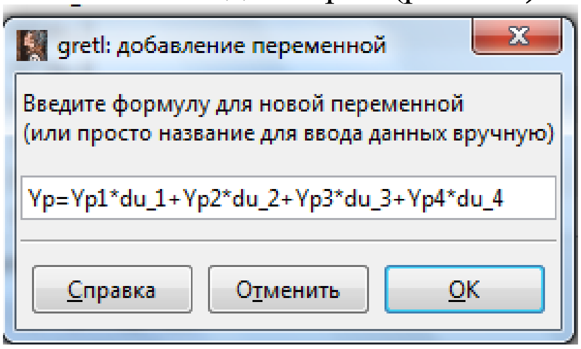


Рис. 2.7. Окно добавления переменной для объединения подвыборок

Выполним такое же объединение подвыборок для переменных $Xp1$, $Xp2$ (рис. 2.8).

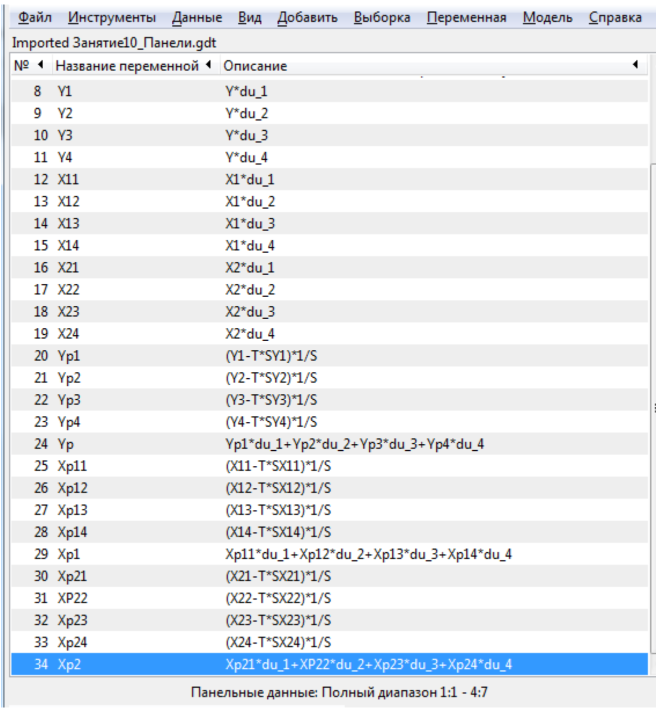
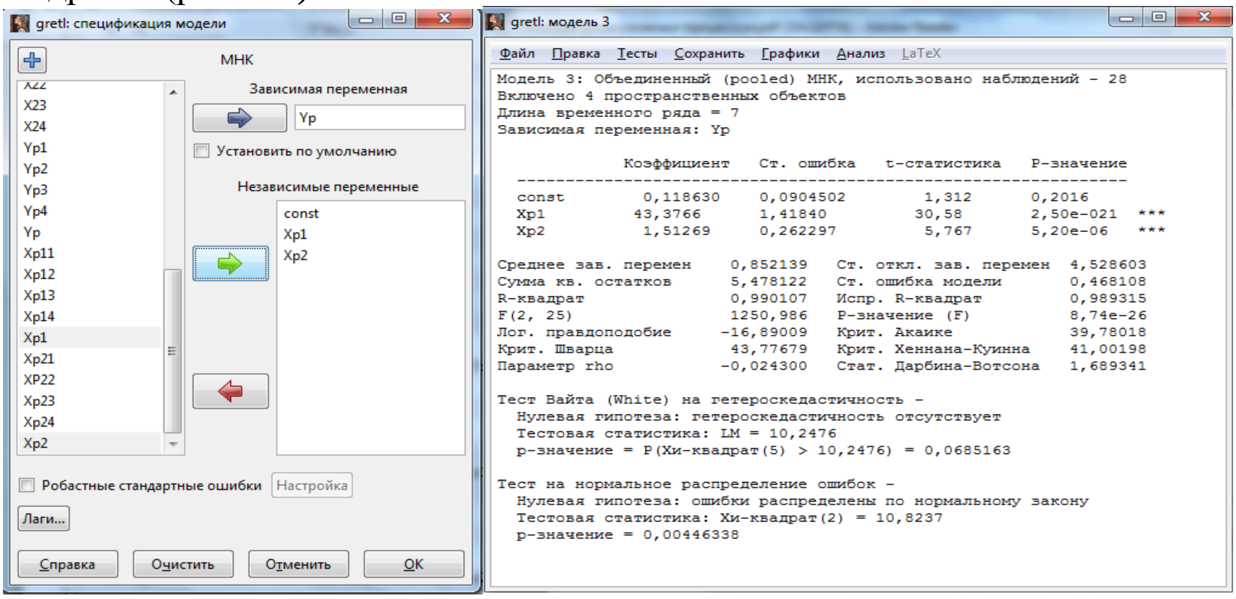


Рис. 2.8. Панель Gretl с набором переменных для модели со случайными эффектами

Построим регрессию *Yp* на *Xp*1, *Xp*2: Модель/Метод наименьших квадратов (рис. 2.9).



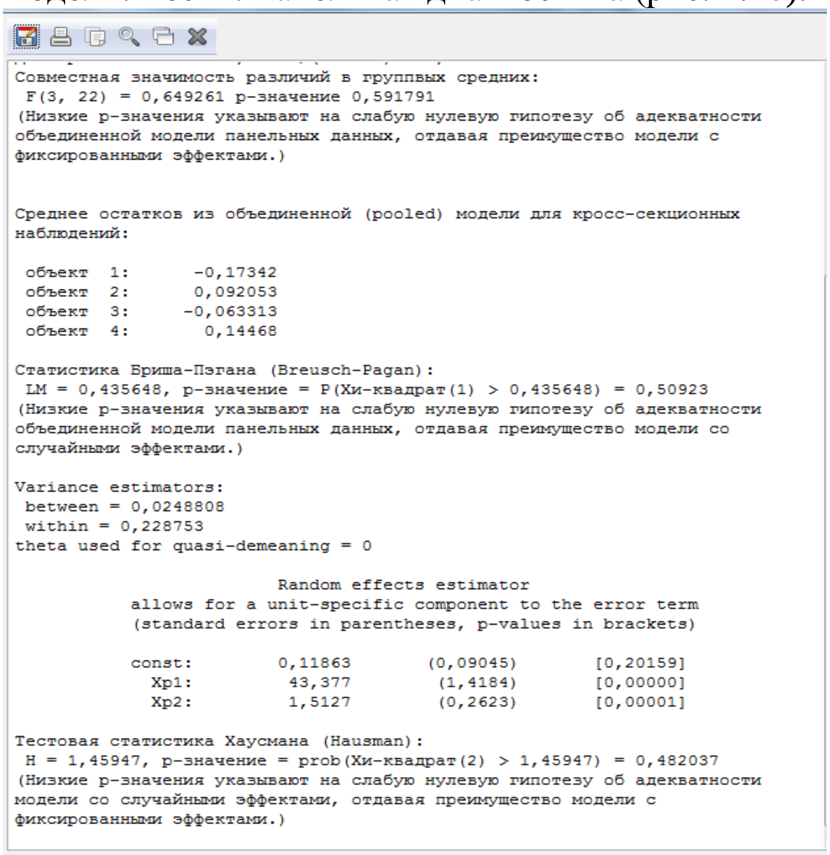

```
Модель 3: Объединенный (pooled) МНК, использовано наблюдений - 28
Включено 4 пространственных объектов
Длина временного ряда = 7
Зависимая переменная: Yp

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        0,118630    0,0904502        1,312      0,2016
  Xp1         43,3766      1,41840         30,58       2,50e-021  ***
  Xp2          1,51269     0,262297         5,767      5,20e-06   ***

Среднее зав. перемен    0,852139   Ст. откл. зав. перемен  4,528603
Сумма кв. остатков      5,478122   Ст. ошибка модели       0,468108
R-квадрат               0,990107   Испр. R-квадрат         0,989315
F(2, 25)                1250,986   Р-значение (F)          8,74e-26
Лог. правдоподобие    -16,89009   Крит. Акаике            39,78018
Крит. Шварца            43,77679   Крит. Хеннана-Куинна    41,00198
Параметр rho           -0,024300   Стат. Дарбина-Вотсона   1,689341

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 10,2476
  р-значение = P(Хи-квадрат(5) > 10,2476) = 0,0685163

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 10,8237
  р-значение = 0,00446338
```

Рис. 2.9. Модель регрессии со случайными эффектами

$Y = 0{,}118 + 43{,}377X1 + 1{,}513X2$.

$R^2 = 0{,}9901$, гетероскедастичность отсутствует, но нормальный закон распределения остатков нарушен.

В окне модели: Тесты/Панельная диагностика (рис. 2.10).



```
Совместная значимость различий в групповых средних:
 F(3, 22) = 0,649261 р-значение 0,591791
(Низкие р-значения указывают на слабую нулевую гипотезу об адекватности
объединенной модели панельных данных, отдавая преимущество модели с
фиксированными эффектами.)


Среднее остатков из объединенной (pooled) модели для кросс-секционных
наблюдений:

 объект  1:      -0,17342
 объект  2:       0,092053
 объект  3:     -0,063313
 объект  4:       0,14468

Статистика Бриша-Пэгана (Breusch-Pagan):
 LM = 0,435648, р-значение = P(Хи-квадрат(1) > 0,435648) = 0,50923
(Низкие р-значения указывают на слабую нулевую гипотезу об адекватности
объединенной модели панельных данных, отдавая преимущество модели со
случайными эффектами.)

Variance estimators:
 between = 0,0248808
 within = 0,228753
theta used for quasi-demeaning = 0

                     Random effects estimator
           allows for a unit-specific component to the error term
           (standard errors in parentheses, p-values in brackets)

           const:        0,11863       (0,09045)       [0,20159]
             Xp1:         43,377        (1,4184)       [0,00000]
             Xp2:         1,5127        (0,2623)       [0,00001]

Тестовая статистика Хаусмана (Hausman):
 H = 1,45947, р-значение = prob(Хи-квадрат(2) > 1,45947) = 0,482037
(Низкие р-значения указывают на слабую нулевую гипотезу об адекватности
модели со случайными эффектами, отдавая преимущество модели с
фиксированными эффектами.)
```

Рис. 2.10. Тестовая статистика Хаусмана

Тест Хаусмана показывает о преимуществе модели со случайными эффектами.

7. Построение регрессионной модели со случайными эффектами с помощью встроенных инструментов Gretl: Модель/Панельные модели/Модель фиксированных или случайных эффектов.

Итоговые результаты приведены в таблице 2.2.

Таблица 2.2

Сводная таблица моделей для панельных данных

Тип модели	Вид модели	R^2	Se	DW
Линейная модель множественной регрессии	$Y = 3,606 + 57,891X1 + 0,418X2$	0,9136	10,40	0,13
Модель с фиксированными эффектами	$Yx = -5,302i1 + 20,109i2 + 5,29i3 + $ $+25,065i4 + 43,275X1 + 1,527X2$	0,9969	2,09	1,04
Модель со случайными эффектами	$Y = 0,118 + 43,377X1 + 1,513X2$	0,9901	0,468	1,69

Задания для самостоятельного выполнения

Задание 2.2. Отделу труда и заработной платы ОАО «Сельхозмаш» было поручено провести исследование факторов, влияющих на среднемесячный размер оплаты труда, выплачиваемой рабочим (Y, руб.). Такими факторами являются: процент перевыполнения месячного плана ($X1$, %), разряд рабочего ($X2$). На основании этого результата по данным четырех цехов была сформирована таблица 4.3. Выполните следующие задания: постройте модель с фиксированными эффектами с помощью фиктивных переменных; постройте модель со случайными эффектами; выберите из построенных моделей наиболее подходящую для аналитических целей.

Таблица 2.3

№	Y	$X1$	$X2$	№	Y	$X1$	$X2$
Первый цех				Третий цех			
1	31700	26,4	6	21	45200	13,4	10
2	20000	17,3	3	22	47900	29,7	10
3	43500	23,8	8	23	31300	21,6	6
4	22000	17,6	5	24	34800	25,1	7
5	27600	26,2	5	25	17200	14,1	2
6	16100	21,1	1	26	27900	24,1	6
7	20000	17,5	3	27	35300	10,5	9
8	24100	22,9	5	28	20500	22,1	2

9	22900	22,9	4	29	21700	17,0	2
10	33200	14,9	7	30	30700	20,5	2
Второй цех				Четвертый цех			
11	21800	19,6	3	31	23300	14,2	4
12	21000	22,8	2	32	45900	18,0	10
13	33800	27,8	6	33	21900	29,9	2
14	26000	14,0	4	34	25500	14,1	5
15	21200	11,4	2	35	26700	18,4	6
16	44000	16,0	9	36	36100	20,1	8
17	17700	28,8	1	37	38800	27,6	9
18	26900	16,8	5	38	26600	27,4	5
19	31900	11,8	6	39	34000	28,5	8
20	23600	18,6	5	40	37100	28,6	9

Задание 2.3. На основе наблюдений за период с 2003 по 2013 г. для четырех стран постройте модели анализа панельных данных о ВВП, млрд. долларов (Y), в зависимости от импорта, млрд. долларов ($X1$); экспорта, млрд. долларов ($X2$); уровня инфляции, % ($X3$) (таблица 4.4.).

Таблица 2.4

№	Y	$X1$	$X2$	$X3$	№	Y	$X1$	$X2$	$X3$
США					Китай				
1	11142,2	1165	687	2,3	23	4229,1	292,1	383,8	-0,3
2	11853,3	1260	714,5	2,7	24	4605,9	346,6	447,1	0
3	12623	1476	795	3,4	25	4552,2	401,8	538,8	-0,3
4	13377,2	1727	927,5	3,2	26	4362,6	451,1	550,5	0,3
5	14028,7	1869	1024	2,9	27	4378	524,1	590,3	0
6	14291,6	2112	1291	3,8	28	4879,8	708,3	746,5	1,4
7	13938,9	1575	1069	-0,3	29	5033	501,6	545,3	-1,4
8	14526,6	1935	1289	1,6	30	5458,8	639,1	730,1	-0,7
9	15060	2236	1497	3,1	31	5855	807,6	787	-0,3
10	15650	2303	1561	2,1	32	5984	830,1	776,6	0
11	16720	2273	1575	1,5	33	5007	766,6	697	0,2
Россия					Япония				
12	430,3	74,8	134,4	13,7	34	1641	295,3	325,6	1,2
13	591,2	92,9	162,5	10,9	35	1931,6	397,4	436,1	3,9
14	763,7	125	245	12,7	36	2256,9	552,4	583,1	1,8
15	989,9	171,5	317,6	9,7	37	2712,9	631,8	752,2	1,5
16	1299,7	223,4	355,5	9	38	3494,2	777,9	974	4,8
17	1660,8	302	471,6	14,1	39	4520	1074	1435	5,9
18	1222	191,8	303,4	11,7	40	4990,5	954,3	1204	-0,7

19	1479,8	248,7	400,4	6,9	41	5878,3	1327	1578	3,2
20	1885	323,8	522	8,4	42	6989	1741	1899	5,4
21	1954	335,7	528	5,1	43	8250	1653	1818	2,6
22	2113	341	515	6,8	44	8939	1772	2210	2,6

Контрольные вопросы:

1. Назовите преимущества использования панельных данных.

2. В чем отличия моделей с фиксированными и случайными эффектами для панельных данных?

3. Можно ли модель с фиксированными эффектами для панельных данных рассматривать как частный случай использования фиктивных переменных?

4. Для проверки какой гипотезы применяется тест Хаусмана?

5. Как проверить значимость фиксированных эффектов и случайных эффектов?

6. Каковы достоинства и недостатки моделей фиксированных и случайных эффектов?

3. Модели с фиктивными независимыми переменными

Расчетные формулы

В общем случае модель с фиктивными переменными имеет вид:

$$y = f(x_1, ... x_p, d_{11}, d_{12}, ..., d_{21}, d_{22}, ..., d_{j1}, d_{j2}, ... \varepsilon),$$

где y – зависимая переменная; $x_1, x_2, \ldots, x_p$ – количественные независимые переменные, d_{11}, d_{12} – фиктивные переменные, соответствующие категориям первого неколичественного показателя; d_{21}, d_{22} – фиктивные переменные, соответствующие категориям второго неколичественного показателя; d_{j1}, d_{j2} – фиктивные переменные, соответствующие категориям j-го неколичественного показателя; ε – случайный остаток.

Регрессионные модели, в которых объясняющие переменные носят как количественный, так и качественный характер, называются ANCOVA-моделями (моделями ковариационного анализа). ANCOVA-модель при наличии у фиктивной переменной сдвига двух альтернатив:

$y = a + b \cdot x + \gamma \cdot D + \varepsilon$, $D = 1$ – лица мужского пола, $D = 0$ – лица женского пола. Ожидаемое потребление кофе при цене *x* будет:

$y = a + b \cdot x + \varepsilon$ для женщины;

$y = a + b \cdot x + \gamma \cdot D + \varepsilon = (a + \gamma) + b \cdot x + \varepsilon$ – для мужчины.

Если γ будет статистически значим по *t*-статистике, то пол влияет на потребление кофе. При $\gamma > 0$ – в пользу мужчин, при $\gamma < 0$ – в пользу женщин.

Для учета структурных изменений в уравнении регрессии фиктивную переменную вводят как сомножитель при количественной переменной и тогда зависимость может быть выражена так:

$$y = a + bx + g_1 D + g_2 Dx + e,$$

где

$$D = \begin{cases} 0, & \textit{до изменения условий}, \\ 1, & \textit{после изменения условий}. \end{cases}$$

В этой ситуации ожидаемое значение зависимой переменной определяется следующим образом:

$$\hat{y} = a + bx, \qquad D = 0,$$

$$\hat{y} = (a + g_1) + (b + g_2)x \qquad D = 1.$$

Тест Чоу*:* Выборка объёма *n* разбивается на две подвыборки объёмами n_1 и n_2, $(n_1 + n_2 = n)$ и для каждой строится уравнение регрессии: s_1 и s_2 остаточные суммы квадратов отклонений для каждой из регрессий, s_3 – для общей регрессии. *F*-статистика имеет распределение Фишера с $(p+1,\ n-2p-2)$ степенями свободы:

$$F = \frac{s_3 - (s_1 + s_2)}{s_1 + s_2} \cdot \frac{n - 2p - 2}{p + 1},$$

где p – число факторов. Если на заданном уровне значимости α $F_{набл} < F(\alpha; p+1; n-2p-2)$, то нет смысла разбивать уравнение регрессии на части. В противном случае разбиение на подвыборки целесообразно с точки зрения улучшения качества модели.

Задания для выполнения в аудитории

Задание 3.1. Провести регрессионный анализ данных о продаже 95 подержанных автомобилей четырех моделей (таблица 3.1). Для измерения влияния модели автомобиля на его цену оцените дифференциальные коэффициенты свободного члена и дифференциальные угловые коэффициенты. Примените фиктивную переменную Zi – модель автомобиля: $Z1 = 1$, если модель VAZ2110, $Z1 = 0$ в других случаях, $Z2 = 1$, если модель

VAZ2115, $Z2 = 0$ в других случаях, $Z3 = 1$, если модель VAZ1119, $Z3 = 0$ в других случаях. Сравните прогнозные качества полученных моделей.

Таблица 3.1

Модель	Цена продажи, руб.	Возраст, лет	Модель	Цена продажи, руб.	Возраст, лет	Модель	Цена продажи, лет	Возраст, лет
VAZ2115	100 000	10	VAZ2110	66 000	11	VAZ2115	105 000	7
VAZ2110	85 000	12	VAZ2181	200 000	5	VAZ2115	85 000	9
VAZ1119	130 000	6	VAZ2115	45 000	15	VAZ2115	45 000	16
VAZ2181	230 000	3	VAZ2115	40 000	17	VAZ2115	30 000	17
VAZ2115	75 000	13	VAZ2110	35 000	23	VAZ2115	48 000	16
VAZ2115	75 000	13	VAZ1119	90 000	9	VAZ1119	160 000	6
VAZ2115	50 000	15	VAZ2110	52 000	14	VAZ2115	87 000	9
VAZ2115	65 000	13	VAZ1119	95 000	9	VAZ1119	135 000	8
VAZ2115	35 000	22	VAZ2110	56 000	13	VAZ1119	120 000	9
VAZ2115	30 000	24	VAZ2115	85 000	10	VAZ2115	40 000	18
VAZ1119	110 000	7	VAZ2110	70 000	8	VAZ2181	297 000	2
VAZ2181	220 000	4	VAZ2181	225 000	4	VAZ2115	45 000	18
VAZ2181	260 000	2	VAZ1119	90 000	9	VAZ2110	110 000	7
VAZ1119	116 000	8	VAZ2181	290 000	2	VAZ2181	325 000	2
VAZ2110	65 000	11	VAZ2110	72 000	12	VAZ2110	95 000	9
VAZ2181	190 000	5	VAZ2181	215 000	4	VAZ2181	195 000	5
VAZ2181	210 000	4	VAZ2181	240 000	3	VAZ2181	295 000	2
VAZ2181	190 000	5	VAZ2181	295 000	2	VAZ2181	275 000	3
VAZ2115	100 000	9	VAZ2181	270 000	3	VAZ2181	270 000	3
VAZ2181	220 000	4	VAZ1119	130 000	7	VAZ1119	150 000	6
VAZ2181	240 000	3	VAZ1119	135 000	7	VAZ1119	130 000	7
VAZ2181	230 000	4	VAZ1119	165 000	6	VAZ2110	73 000	13
VAZ2181	300 000	2	VAZ2115	75 000	11	VAZ2115	98 000	7
VAZ2181	230 000	3	VAZ1119	160 000	6	VAZ2115	98 000	7
VAZ2181	220 000	4	VAZ2115	80 000	10	VAZ2115	98 000	7
VAZ2181	240 000	3	VAZ2115	90 000	9	VAZ2181	227 000	4
VAZ2181	290 000	2	VAZ2115	95 000	9	VAZ2115	80 000	10
VAZ2181	190 000	5	VAZ2115	95 000	9	VAZ1119	140 000	6
VAZ2115	65 000	13	VAZ1119	155 000	7	VAZ2181	245 000	3
VAZ2110	75 000	9	VAZ1119	135 000	8			
VAZ2181	245 000	3	VAZ2115	65 000	12			
VAZ2181	225 000	4	VAZ2115	67 000	12			
VAZ2181	305 000	2	VAZ2115	77 000	10			

Методические указания для выполнения задания

1. Построим регрессию цены от возраста автомобиля без учета модели автомобиля (без фиктивных переменных).

1.1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «Занятие_Фиктивные.xlsx». Импорт данных из таблицы Excel: Файл/Открыть/Пользовательские/лист 1.

1.2. Построение уравнения регрессии: Модель/Метод наименьших квадратов…

1.3. Проверка остатков на соблюдение предпосылок МНК: В окне модели: Тесты/Гетероскедастичность… Затем для визуального анализа остатков: В окне модели: Графики/График остатков/В зависимости от *Age* (рис. 3.1).

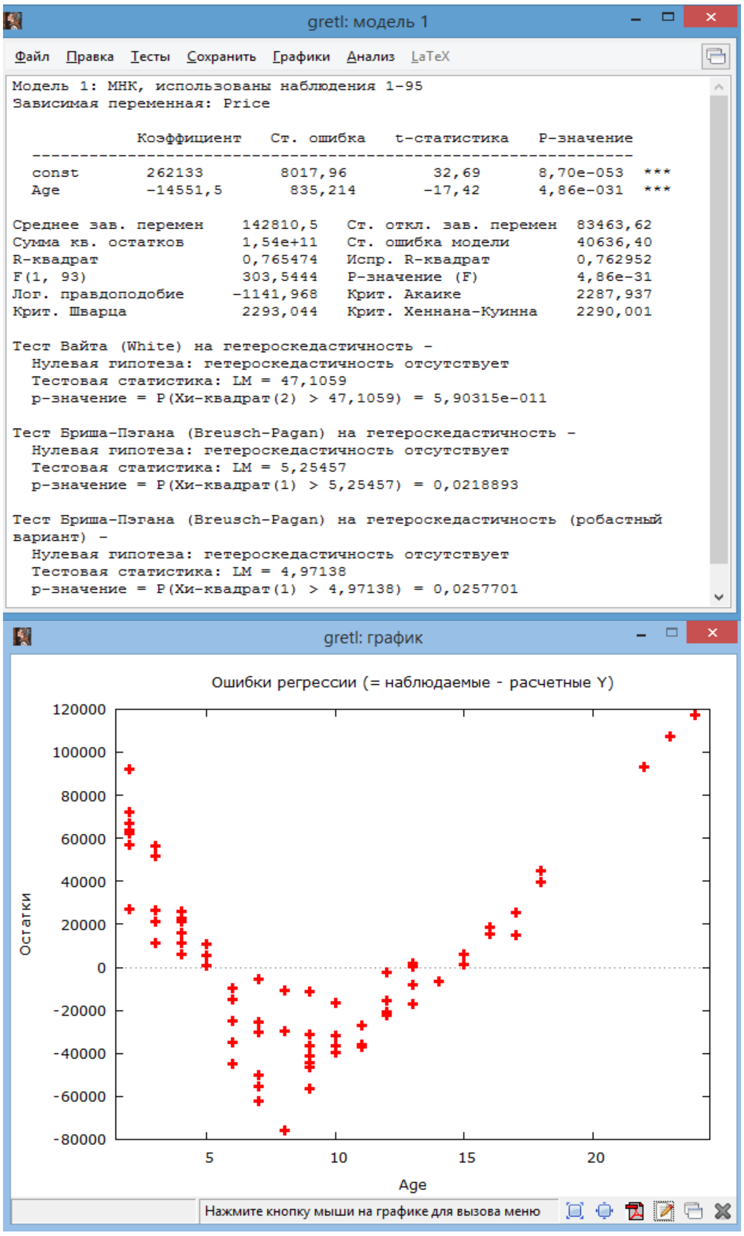
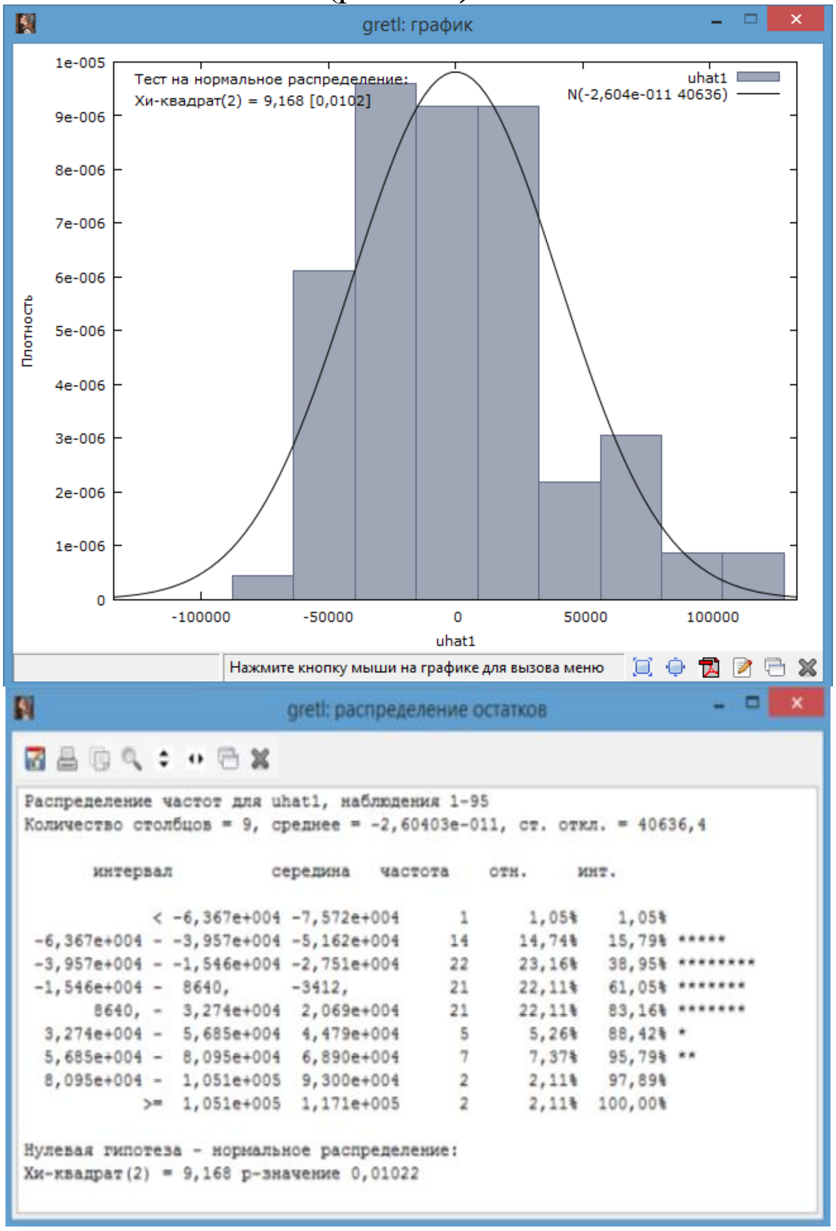


Рис. 3.1. Модель регрессии без включения фиктивных переменных

$$Y = 262133 - 14551{,}5 Age + e.$$

Интерпретация коэффициентов: $a = 262133$ – цена нового автомобиля, $b = -14551{,}5$ – снижение цены за каждый год эксплуатации. $R^2 = 0{,}765$. По статистике Фишера и Стьюдента модель и коэффициенты значимы с вероятностью 99%. Все тесты на гетероскедастичность и график остатков обнаружили ее присутствие в остатках ($p < 0{,}05$).

1.4. Проверка остатков на нормальность в окне модели: Тесты/Нормальность остатков… (рис. 3.2).




```
Распределение частот для uhat1, наблюдения 1-95
Количество столбцов = 9, среднее = -2,60403e-011, ст. откл. = 40636,4

       интервал           середина   частота    отн.     инт.

            < -6,367e+004 -7,572e+004      1      1,05%    1,05%
 -6,367e+004 - -3,957e+004 -5,162e+004     14     14,74%   15,79% *****
 -3,957e+004 - -1,546e+004 -2,751e+004     22     23,16%   38,95% ********
 -1,546e+004 -  8640,      -3412,          21     22,11%   61,05% *******
       8640, -  3,274e+004  2,069e+004     21     22,11%   83,16% *******
  3,274e+004 -  5,685e+004  4,479e+004      5      5,26%   88,42% *
  5,685e+004 -  8,095e+004  6,890e+004      7      7,37%   95,79% **
  8,095e+004 -  1,051e+005  9,300e+004      2      2,11%   97,89%
            >=  1,051e+005  1,171e+005      2      2,11%  100,00%

Нулевая гипотеза - нормальное распределение:
Хи-квадрат(2) = 9,168 p-значение 0,01022
```

Рис. 3.2. Тест на нормальное распределение остатков в модели без фиктивных переменных

Нормальный закон распределения остатков нарушен ($p < 0{,}05$).

1.5. На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.3).

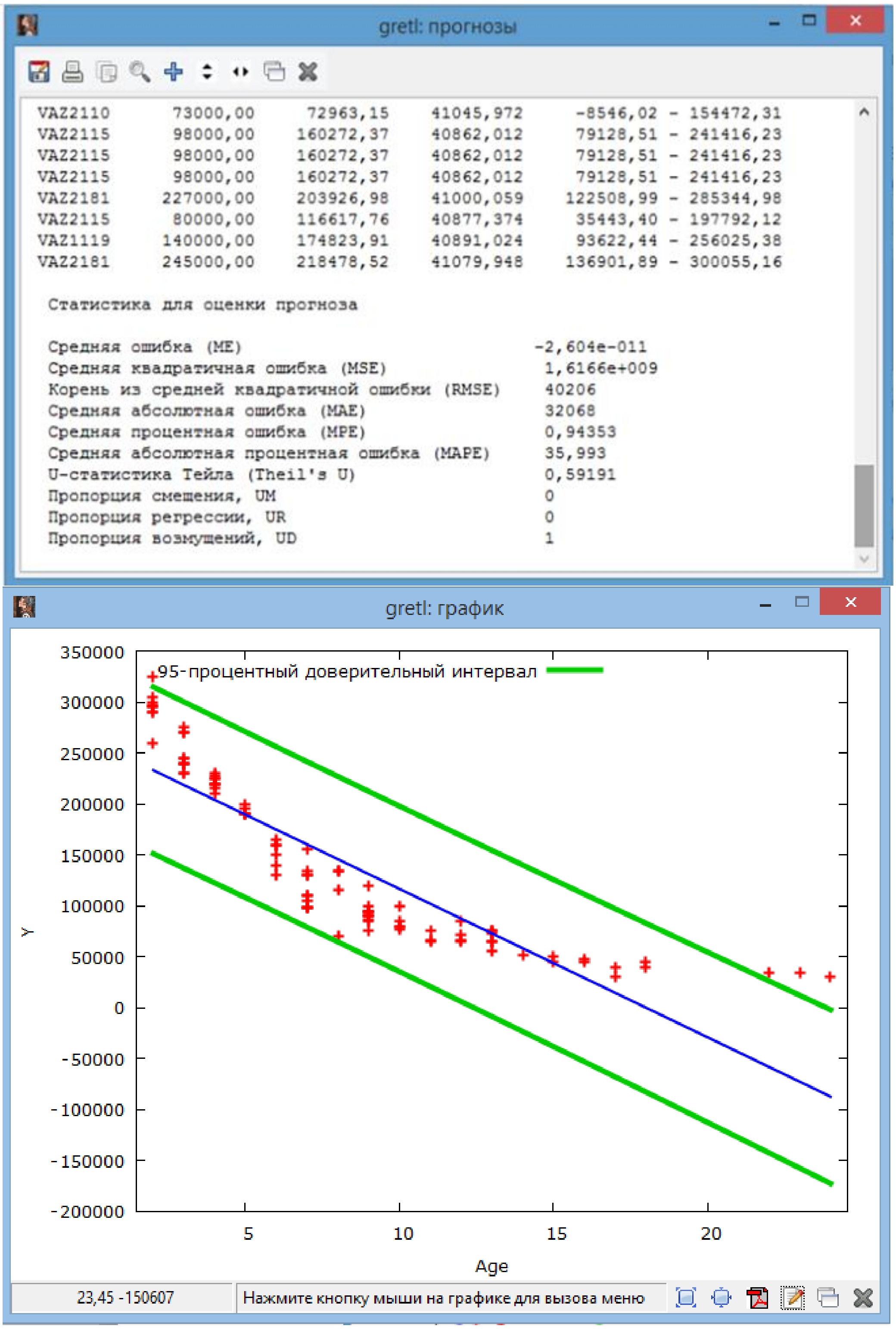
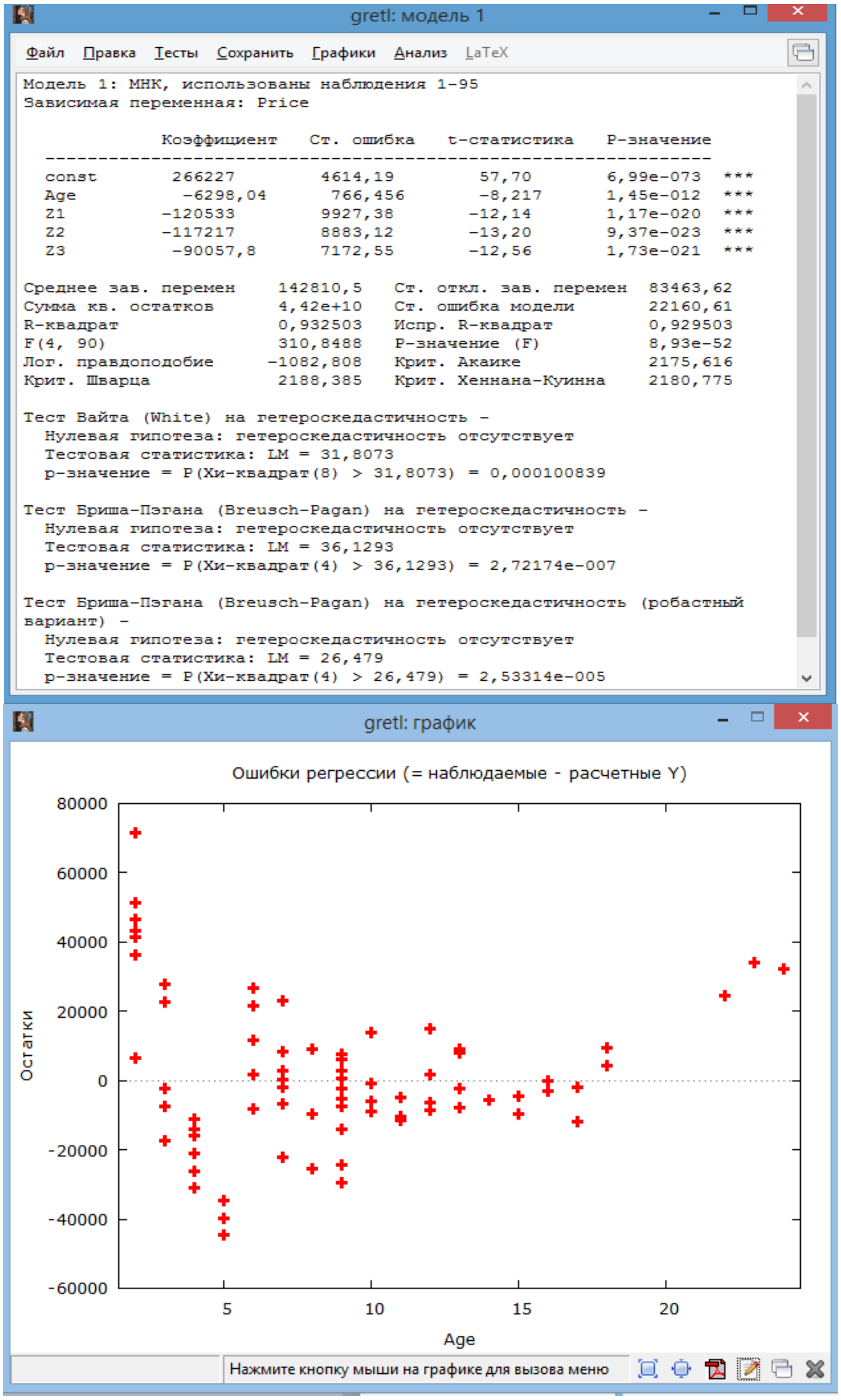


Рис. 3.3. Анализ прогнозных качеств модели без фиктивных переменных

Средняя процентная ошибка: $MPE = 0{,}94$.

Средняя абсолютная процентная ошибка: $MAPE = 35{,}99$.

2. Перестроим модель, включив в нее фиктивные переменные: Модель/Метод наименьших квадратов… (рис. 3.4).




```
Модель 1: МНК, использованы наблюдения 1-95
Зависимая переменная: Price

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        266227       4614,19        57,70       6,99e-073  ***
  Age           -6298,04     766,456       -8,217      1,45e-012  ***
  Z1          -120533       9927,38       -12,14       1,17e-020  ***
  Z2          -117217       8883,12       -13,20       9,37e-023  ***
  Z3           -90057,8     7172,55       -12,56       1,73e-021  ***

Среднее зав. перемен   142810,5   Ст. откл. зав. перемен   83463,62
Сумма кв. остатков     4,42e+10   Ст. ошибка модели        22160,61
R-квадрат              0,932503   Испр. R-квадрат          0,929503
F(4, 90)               310,8488   Р-значение (F)           8,93e-52
Лог. правдоподобие    -1082,808   Крит. Акаике             2175,616
Крит. Шварца           2188,385   Крит. Хеннана-Куинна     2180,775

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 31,8073
  p-значение = P(Хи-квадрат(8) > 31,8073) = 0,000100839

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 36,1293
  p-значение = P(Хи-квадрат(4) > 36,1293) = 2,72174e-007

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 26,479
  p-значение = P(Хи-квадрат(4) > 26,479) = 2,53314e-005
```


Рис. 3.4. Модель регрессии с фиктивными переменными

$Y = 266227 - 6298{,}04 Age - 120533 Z1 - 117217 Z2 - 90057{,}8 Z3 + \varepsilon$.

Для модели VAZ2110: $Y = (266227 - 120533) - 6298{,}04 Age + \varepsilon$.

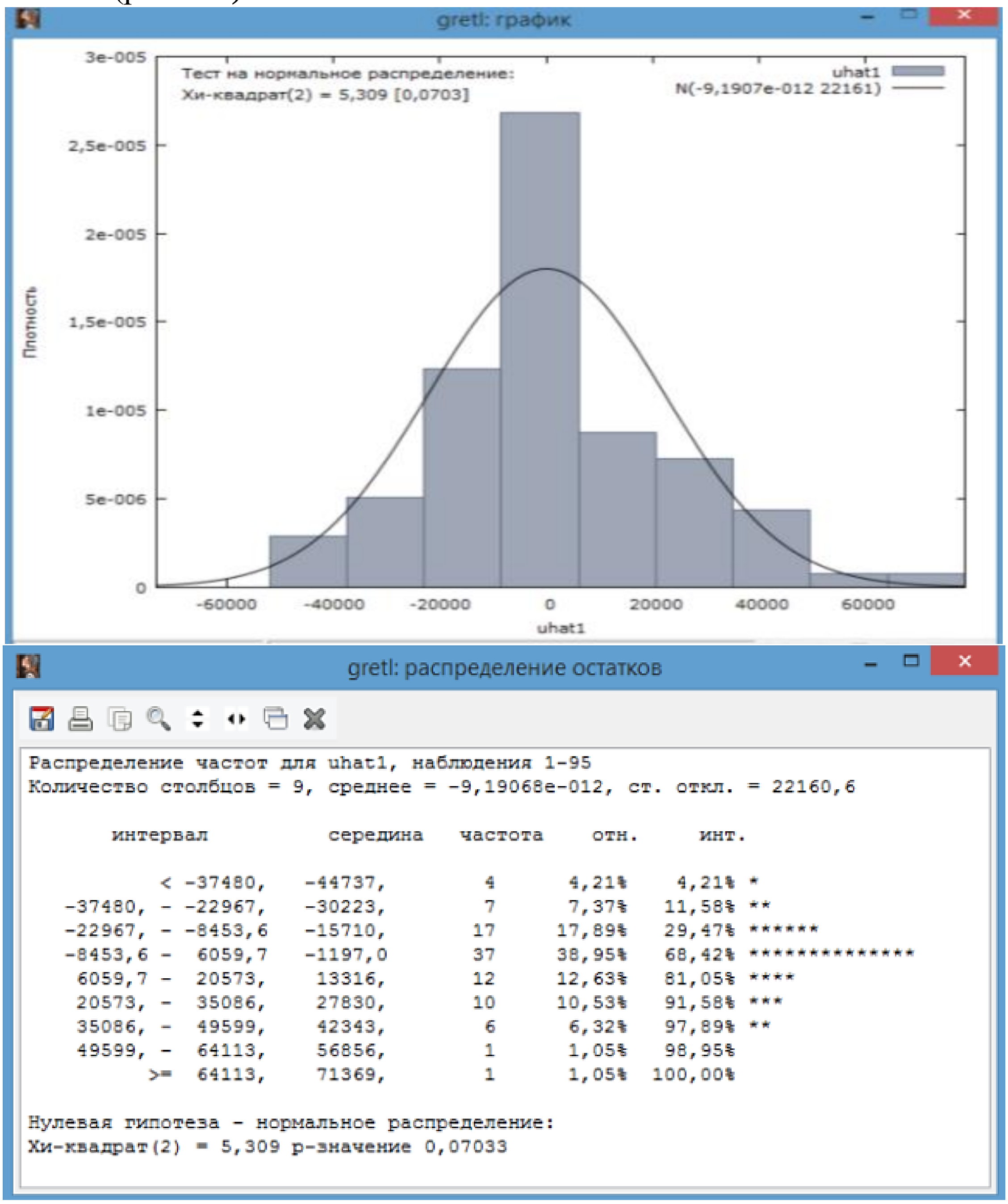
Для модели VAZ2115: $Y = (266227 - 117217) - 6298{,}04 Age + \varepsilon$.

Для модели VAZ1119: $Y = (266227 - 90057{,}8) - 6298{,}04 Age + \varepsilon$.

Для модели VAZ2181: $Y = 266227 - 6298{,}04 Age + \varepsilon$.

Значит, самая высокая начальная цена у модели VAZ2181. Статистическая значимость дифференциальных коэффициентов свободного члена при фиктивных переменных подтверждает влияние модели автомобиля на его цену. $R^2 = 0{,}93$. По статистике Фишера и Стьюдента модель и коэффициенты значимы с вероятностью 99%. Тесты на гетероскедастичность обнаружили ее присутствие в остатках ($p < 0{,}05$). Визуальный анализ остатков: В окне модели: Графики/График остатков/В зависимости от *Age*. График остатков также показывает наличие гетероскедастичности.

2.1. Проверка остатков на нормальность: Тесты/Нормальность остатков… (рис. 3.5).

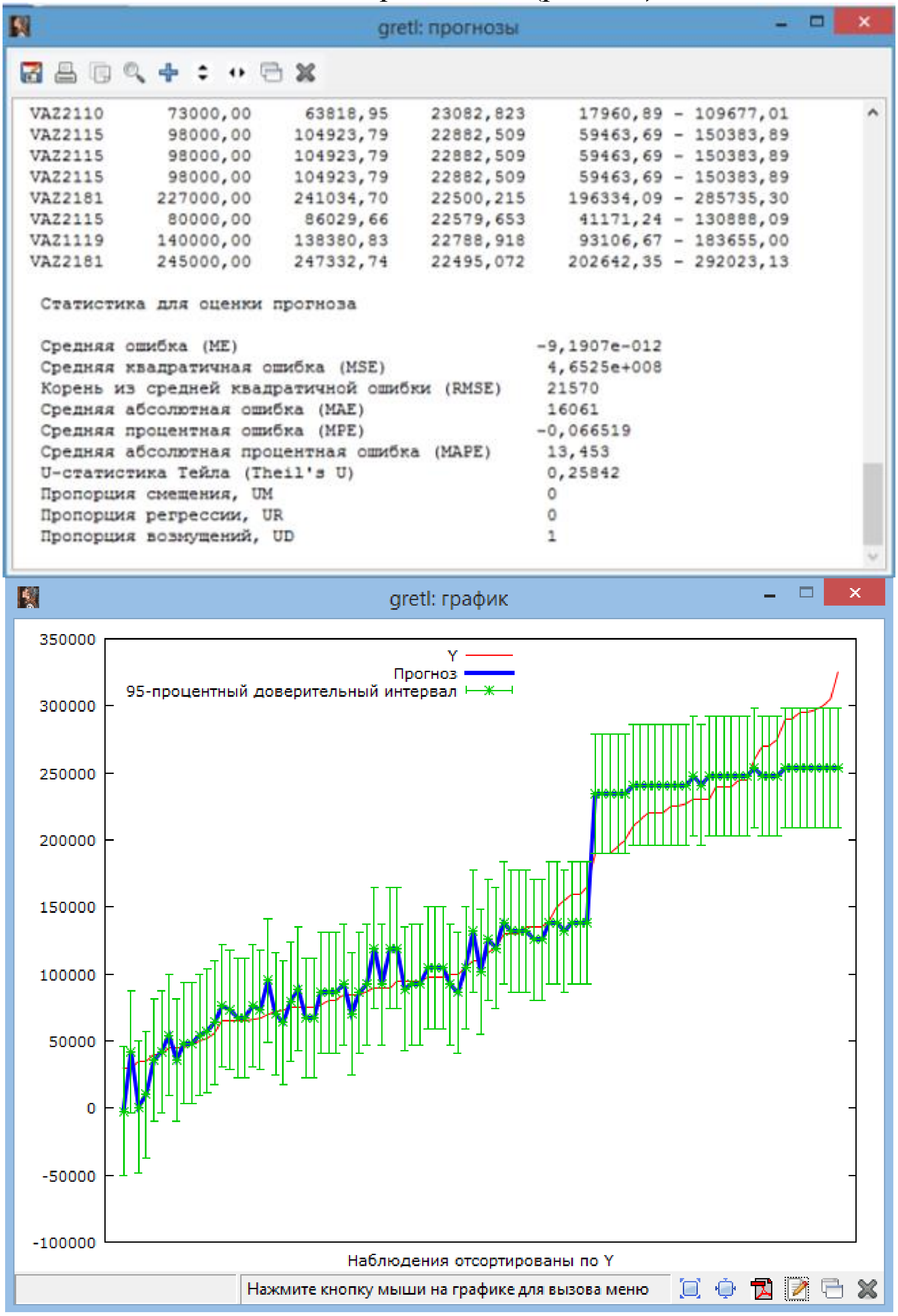



```
Распределение частот для uhat1, наблюдения 1-95
Количество столбцов = 9, среднее = -9,19068e-012, ст. откл. = 22160,6

       интервал          середина   частота    отн.     инт.

           < -37480,   -44737,        4      4,21%    4,21% *
   -37480, - -22967,   -30223,        7      7,37%   11,58% **
   -22967, - -8453,6   -15710,       17     17,89%   29,47% ******
   -8453,6 -  6059,7   -1197,0       37     38,95%   68,42% *************
    6059,7 -  20573,    13316,       12     12,63%   81,05% ****
    20573, -  35086,    27830,       10     10,53%   91,58% ***
    35086, -  49599,    42343,        6      6,32%   97,89% **
    49599, -  64113,    56856,        1      1,05%   98,95%
          >=  64113,    71369,        1      1,05%  100,00%

Нулевая гипотеза - нормальное распределение:
Хи-квадрат(2) = 5,309 p-значение 0,07033
```

Рис. 3.5. Тест на нормальное распределение остатков в модели без фиктивных переменных

Нормальный закон распределения остатков подтвержден ($p > 0{,}05$).

2.2. На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.6).



VAZ2110	73000,00	63818,95	23082,823	17960,89 - 109677,01
VAZ2115	98000,00	104923,79	22882,509	59463,69 - 150383,89
VAZ2115	98000,00	104923,79	22882,509	59463,69 - 150383,89
VAZ2115	98000,00	104923,79	22882,509	59463,69 - 150383,89
VAZ2181	227000,00	241034,70	22500,215	196334,09 - 285735,30
VAZ2115	80000,00	86029,66	22579,653	41171,24 - 130888,09
VAZ1119	140000,00	138380,83	22788,918	93106,67 - 183655,00
VAZ2181	245000,00	247332,74	22495,072	202642,35 - 292023,13

Статистика для оценки прогноза

Средняя ошибка (ME)	-9,1907e-012
Средняя квадратичная ошибка (MSE)	4,6525e+008
Корень из средней квадратичной ошибки (RMSE)	21570
Средняя абсолютная ошибка (MAE)	16061
Средняя процентная ошибка (MPE)	-0,066519
Средняя абсолютная процентная ошибка (MAPE)	13,453
U-статистика Тейла (Theil's U)	0,25842
Пропорция смещения, UM	0
Пропорция регрессии, UR	0
Пропорция возмущений, UD	1

Рис. 3.6. Анализ прогнозных качеств модели с фиктивными переменными

Средняя процентная ошибка: $MPE = -0{,}07$. Средняя абсолютная процентная ошибка: $MAPE = 13{,}45$.

2.3. Выполним коррекцию на гетероскедастичность: Модель/Другие линейные модели/С коррекцией на гетероскедастичность… (рис. 3.7).

```
gretl: модель 2
Файл  Правка  Тесты  Сохранить  Графики  Анализ  LaTeX

Модель 2: С поправкой на гетероскедастичность, использованы наблюдения
1-95
Зависимая переменная: Price

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        260740       6057,02         43,05      7,81e-062  ***
  Age           -6190,99     388,434       -15,94      6,05e-028  ***
  Z1          -119388       7819,39        -15,27      1,03e-026  ***
  Z2          -114785       6672,06        -17,20      3,36e-030  ***
  Z3           -85994,7     6983,50        -12,31      5,26e-021  ***

Статистика, полученная по взвешенным данным:

Сумма кв. остатков     241,1019   Ст. ошибка модели      1,636737
R-квадрат              0,925743   Испр. R-квадрат        0,922442
F(4, 90)               280,4997   Р-значение (F)         6,51e-50
Лог. правдоподобие    -179,0379   Крит. Акаике           368,0759
Крит. Шварца           380,8453   Крит. Хеннана-Куинна   373,2357

Статистика, полученная по исходным данным:

Среднее зав. перемен     142810,5   Ст. откл. зав. перемен  83463,62
Сумма кв. остатков       4,53e+10   Ст. ошибка модели       22433,61
```

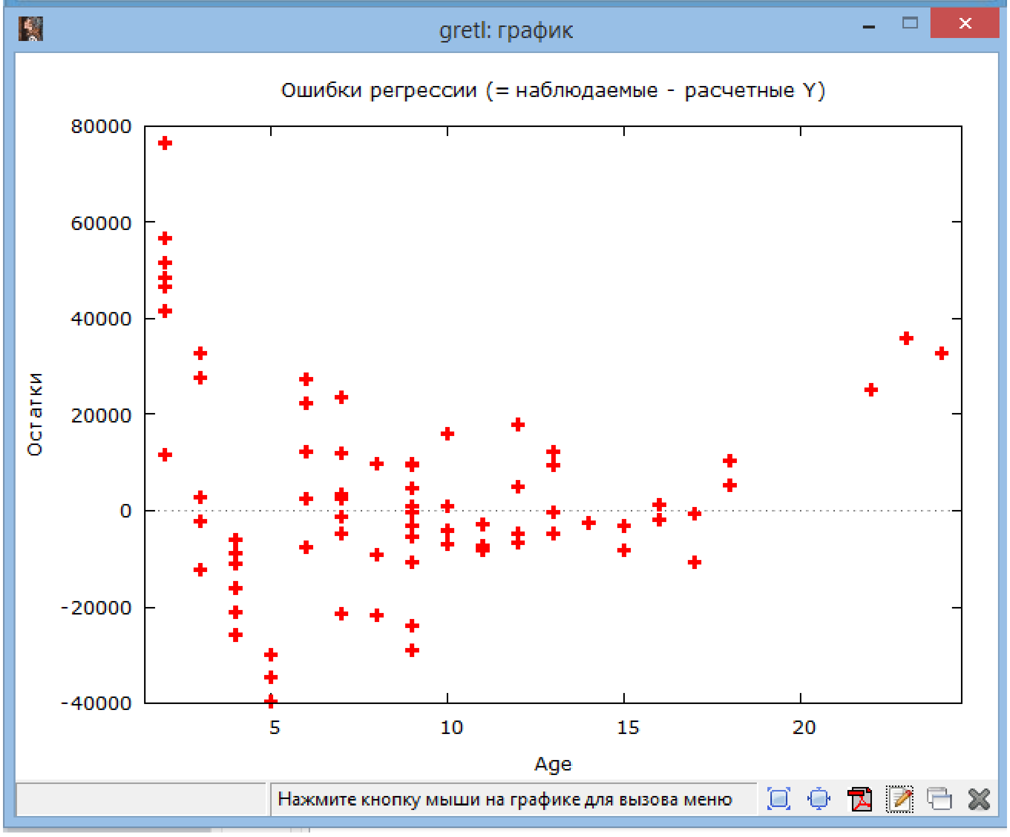


Рис. 3.7. Модель регрессии с фиктивными переменными с поправкой на гетероскедастичность

$$Y = 260740 - 6190{,}99 Age - 119388 Z1 - 114785 Z2 - 85994{,}7 Z3 + \varepsilon .$$

Для модели VAZ2110: $Y = \left(260740 - 119388\right) - 6190{,}99 Age + \varepsilon$.

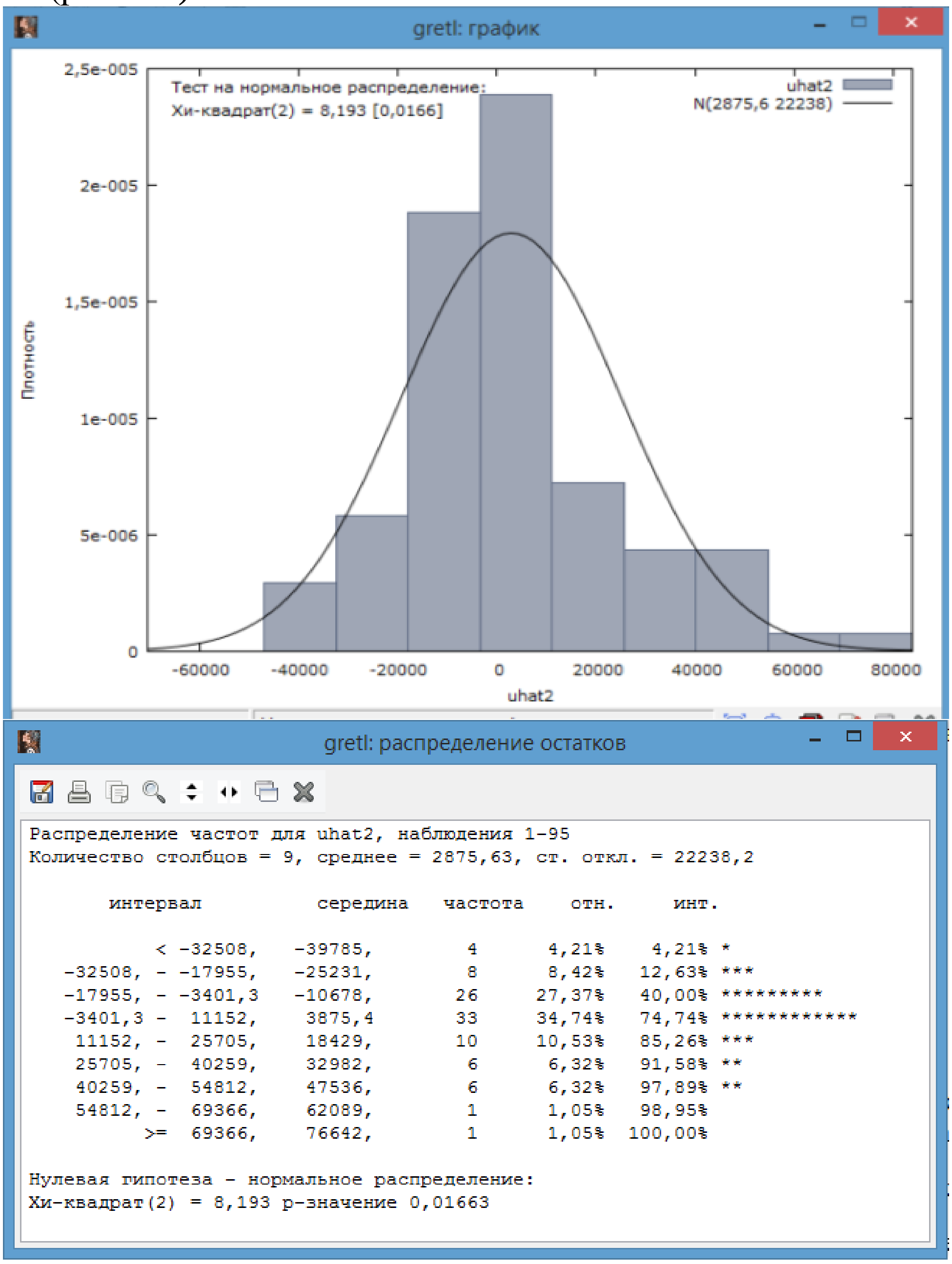
Для модели VAZ2115: $Y = \left(260740 - 114785\right) - 6190{,}99 Age + \varepsilon$.

Для модели VAZ1119: $Y = (260740 - 85994{,}7) - 6190{,}99 Age + \varepsilon$.

Для модели VAZ2181: $Y = 260740 - 6190{,}99 Age + \varepsilon$.

$R^2 = 0{,}93$. По статистике Фишера и Стьюдента модель и коэффициенты значимы с вероятностью 99%. Тесты на гетероскедастичность не активны.

2.4. Проверка остатков на нормальность: Тесты/Нормальность остатков… (рис. 3.8).



Распределение частот для uhat2, наблюдения 1-95
Количество столбцов = 9, среднее = 2875,63, ст. откл. = 22238,2

интервал	середина	частота	отн.	инт.	
< -32508,	-39785,	4	4,21%	4,21%	*
-32508, - -17955,	-25231,	8	8,42%	12,63%	***
-17955, - -3401,3	-10678,	26	27,37%	40,00%	*********
-3401,3 - 11152,	3875,4	33	34,74%	74,74%	************
11152, - 25705,	18429,	10	10,53%	85,26%	***
25705, - 40259,	32982,	6	6,32%	91,58%	**
40259, - 54812,	47536,	6	6,32%	97,89%	**
54812, - 69366,	62089,	1	1,05%	98,95%	
>= 69366,	76642,	1	1,05%	100,00%	

Нулевая гипотеза - нормальное распределение:
Хи-квадрат(2) = 8,193 p-значение 0,01663

Рис. 3.8. Тест на нормальное распределение остатков в модели без фиктивных переменных с поправкой на гетероскедастичность

Нормальный закон распределения остатков не подтвержден ($p < 0{,}05$).

2.5. На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.9).

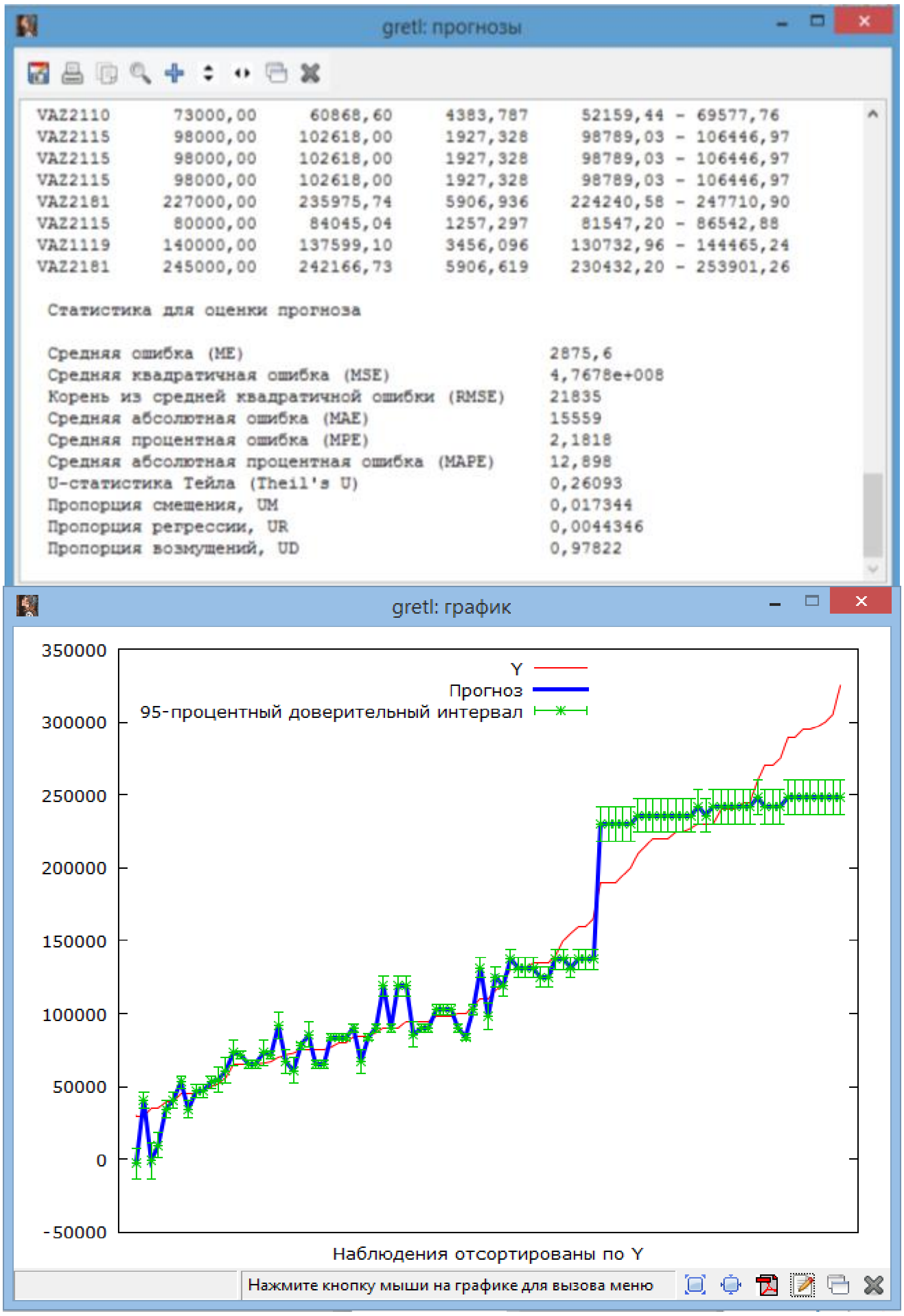


Рис. 3.9. Анализ прогнозных качеств модели с фиктивными переменными

Средняя процентная ошибка: $MPE = 2{,}18$. Средняя абсолютная процентная ошибка: $MAPE = 12{,}9$.

3. Построим модель с дифференциальными угловыми коэффициентами, учитывая взаимодействие возраста автомобиля (Age) и модели автомобиля (Zi). Создадим новые переменные: $Z1Age$, $Z2Age$, $Z3Age$: Добавить/Добавить новую переменную. Ввести вручную формулу для новой переменной: $Z1Age = Z1 \cdot Age$ и т. д. для $Z2$, $Z3$. Построим модель: Модель/Метод наименьших квадратов… (рис. 3.10).

```
gretl: модель 3
Файл  Правка  Тесты  Сохранить  Графики  Анализ  LaTeX
Модель 3: МНК, использованы наблюдения 1-95
Зависимая переменная: Price

                 Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const            357753        7146,81        50,06       6,43e-066  ***
  Age              -34007,8      2065,72       -16,46       1,83e-028  ***
  Z1              -241775       13223,3        -18,28       1,51e-031  ***
  Z2              -224094        9730,71       -23,03       8,79e-039  ***
  Z3              -110085       19852,3         -5,545      3,10e-07   ***
  Z1Age             30220,9      2249,96        13,43       6,32e-023  ***
  Z2Age             28947,1      2126,08        13,62       2,83e-023  ***
  Z3Age             17885,4      3253,79         5,497      3,81e-07   ***

Среднее зав. перемен    142810,5   Ст. откл. зав. перемен   83463,62
Сумма кв. остатков      1,30e+10   Ст. ошибка модели        12215,65
R-квадрат               0,980174   Испр. R-квадрат          0,978579
F(7, 87)                614,4611   Р-значение (F)           3,35e-71
Лог. правдоподобие     -1024,616   Крит. Акаике             2065,231
Крит. Шварца            2085,662   Крит. Хеннана-Куинна     2073,487

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 20,9539
  р-значение = P(Хи-квадрат(11) > 20,9539) = 0,0338518

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 20,914
  р-значение = P(Хи-квадрат(7) > 20,914) = 0,00389962

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 19,4562
  р-значение = P(Хи-квадрат(7) > 19,4562) = 0,00687207
```

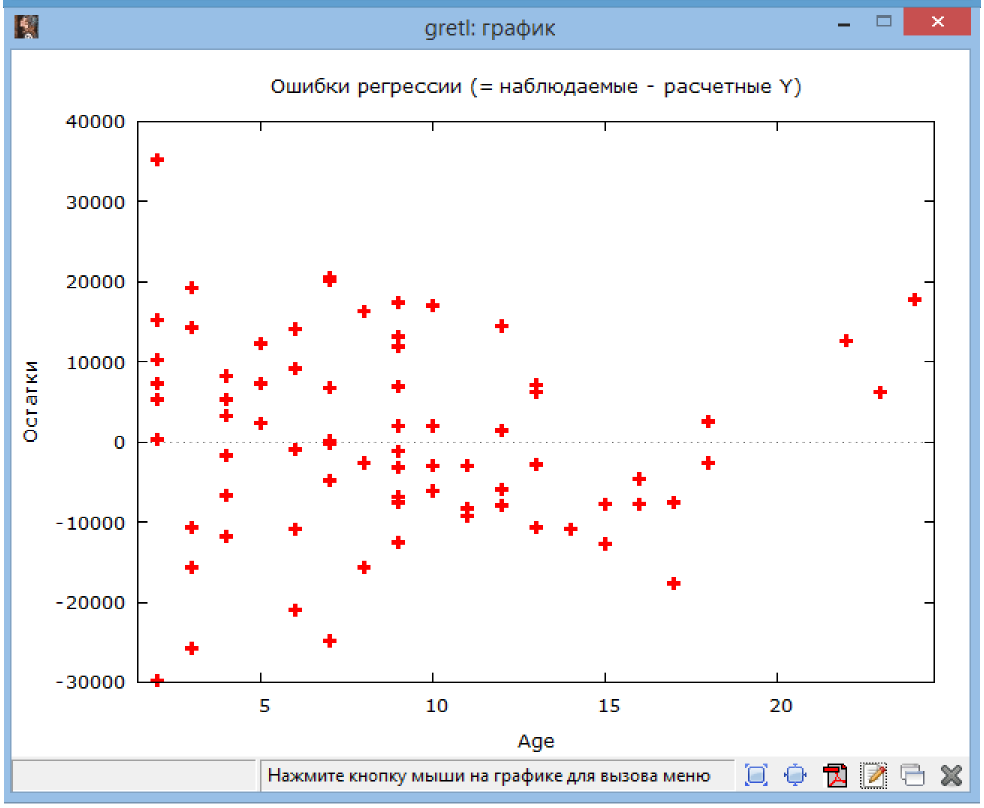


Рис. 3.10. Модель регрессии с дифференциальными угловыми коэффициентами

$$Y = 357753 - 34007{,}8Age - 241775Z1 - 224094Z2 - 110085Z3 + 30220{,}9 \times$$
$$\times Z1Age + 28947{,}1Z2Age + 17885{,}4Z3Age + \varepsilon.$$

Для модели VAZ2110:

$$Y = (357753 - 241775) + (\ -34007{,}8 + 30220{,}9)\,Age\ + \varepsilon.$$

Для модели VAZ2115:

$$Y = (357753 - 224094) + (-34007,8 + 28947,1)Age + \varepsilon .$$

Для модели VAZ1119:

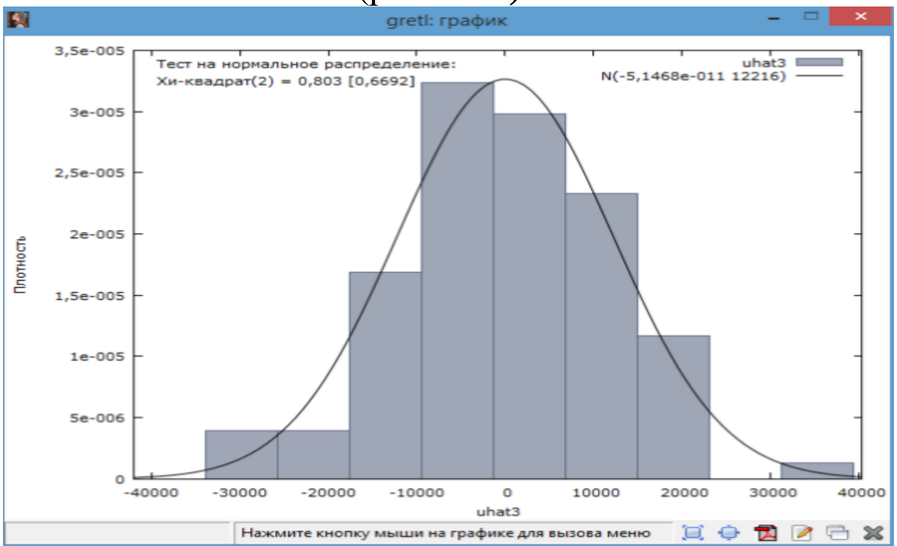
$$Y = (357753 - 110085) + (-34007,8 + 17885,4)Age + \varepsilon .$$

Для модели VAZ2181:

$$Y = 357753 - 34007,8Age + \varepsilon .$$

$R^2 = 0,98$. По статистике Фишера и Стьюдента модель и коэффициенты значимы с вероятностью 99%. Тест Вайта, тесты Бриша-Пэгана и Коэнкера обнаружили гетероскедастичность в остатках ($p < 0,05$). Статистическая значимость дифференциальных коэффициентов свободного члена и дифференциальных угловых коэффициентов при фиктивных переменных подтверждает влияние модели автомобиля на его цену. По графику остатков можно предположить гетероскедастичность.

3.1. Проверка остатков на нормальность в окне модели: Тесты/Нормальность остатков… (рис. 3.11).



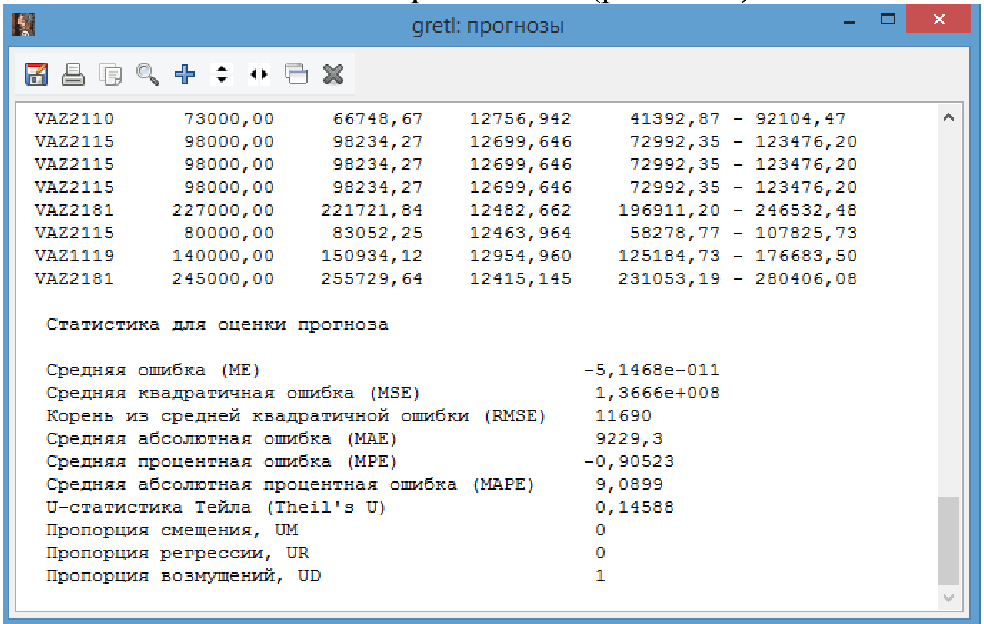
gretl: распределение остатков

```
Распределение частот для uhat3, наблюдения 1-95
Количество столбцов = 9, среднее = -5,14678e-011, ст. откл. = 12215,6

       интервал           середина   частота    отн.     инт.

          < -25675,      -29737,        3       3,16%    3,16% *
   -25675, - -17550,     -21612,        3       3,16%    6,32% *
   -17550, - -9424,9     -13487,       13      13,68%   20,00% ****
   -9424,9 - -1299,9     -5362,4       25      26,32%   46,32% *********
   -1299,9 -  6825,1      2762,6       23      24,21%   70,53% ********
    6825,1 -  14950,      10888,       18      18,95%   89,47% ******
    14950, -  23075,      19013,        9       9,47%   98,95% ***
    23075, -  31200,      27138,        0       0,00%   98,95%
          >=  31200,      35263,        1       1,05%  100,00%

Нулевая гипотеза - нормальное распределение:
Хи-квадрат(2) = 0,803 p-значение 0,66917
```

Рис. 3.11. Тест на нормальное распределение остатков в модели
с дифференциальными угловыми коэффициентами

Подтверждается соблюдение нормального закона распределения остатков ($p > 0{,}05$).

3.2. На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.12).



gretl: прогнозы

VAZ2110	73000,00	66748,67	12756,942	41392,87 - 92104,47
VAZ2115	98000,00	98234,27	12699,646	72992,35 - 123476,20
VAZ2115	98000,00	98234,27	12699,646	72992,35 - 123476,20
VAZ2115	98000,00	98234,27	12699,646	72992,35 - 123476,20
VAZ2181	227000,00	221721,84	12482,662	196911,20 - 246532,48
VAZ2115	80000,00	83052,25	12463,964	58278,77 - 107825,73
VAZ1119	140000,00	150934,12	12954,960	125184,73 - 176683,50
VAZ2181	245000,00	255729,64	12415,145	231053,19 - 280406,08

Статистика для оценки прогноза

Средняя ошибка (ME)	-5,1468e-011
Средняя квадратичная ошибка (MSE)	1,3666e+008
Корень из средней квадратичной ошибки (RMSE)	11690
Средняя абсолютная ошибка (MAE)	9229,3
Средняя процентная ошибка (MPE)	-0,90523
Средняя абсолютная процентная ошибка (MAPE)	9,0899
U-статистика Тейла (Theil's U)	0,14588
Пропорция смещения, UM	0
Пропорция регрессии, UR	0
Пропорция возмущений, UD	1

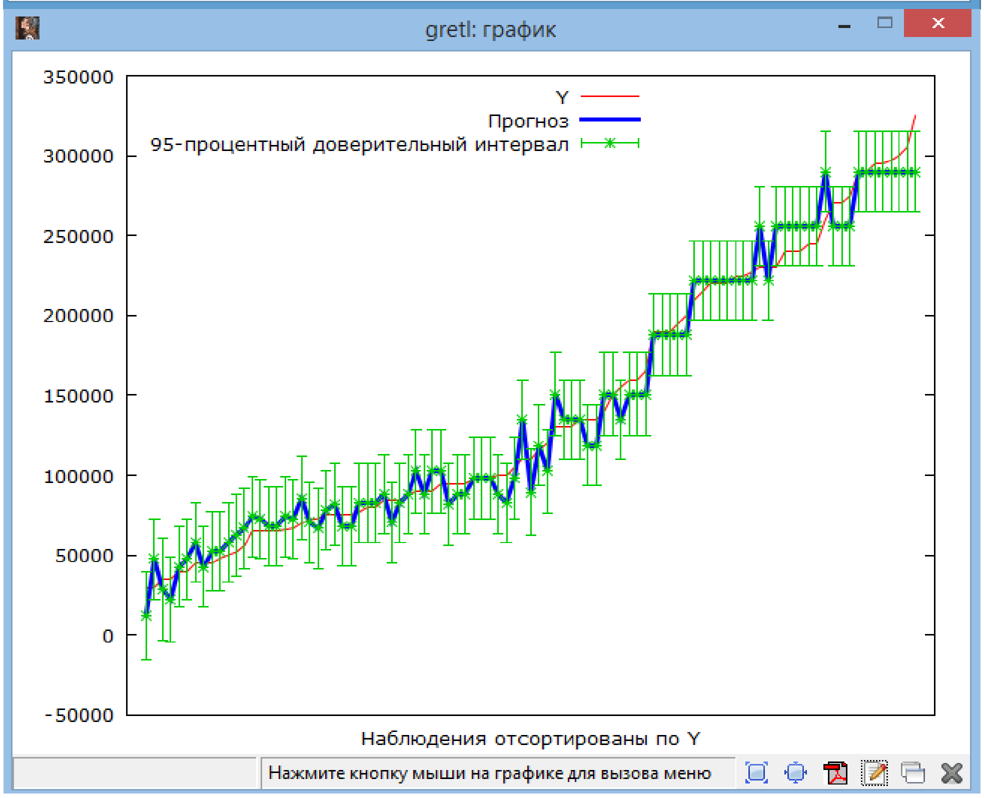


Рис. 3.12. Анализ прогнозных качеств модели с дифференциальными угловыми коэффициентами

Средняя процентная ошибка: $MPE = -0{,}9$. Средняя абсолютная процентная ошибка: $MAPE = 9{,}09$.

4. Построим модель с дифференциальными угловыми коэффициентами, учитывая взаимодействие возраста автомобиля (*Age*) и

модели автомобиля (Zi), дополнив ее квадратом переменной *Age*: Добавить/Квадраты выделенных переменных (предварительно выделить мышкой переменную *Age*) (рис. 3.13).

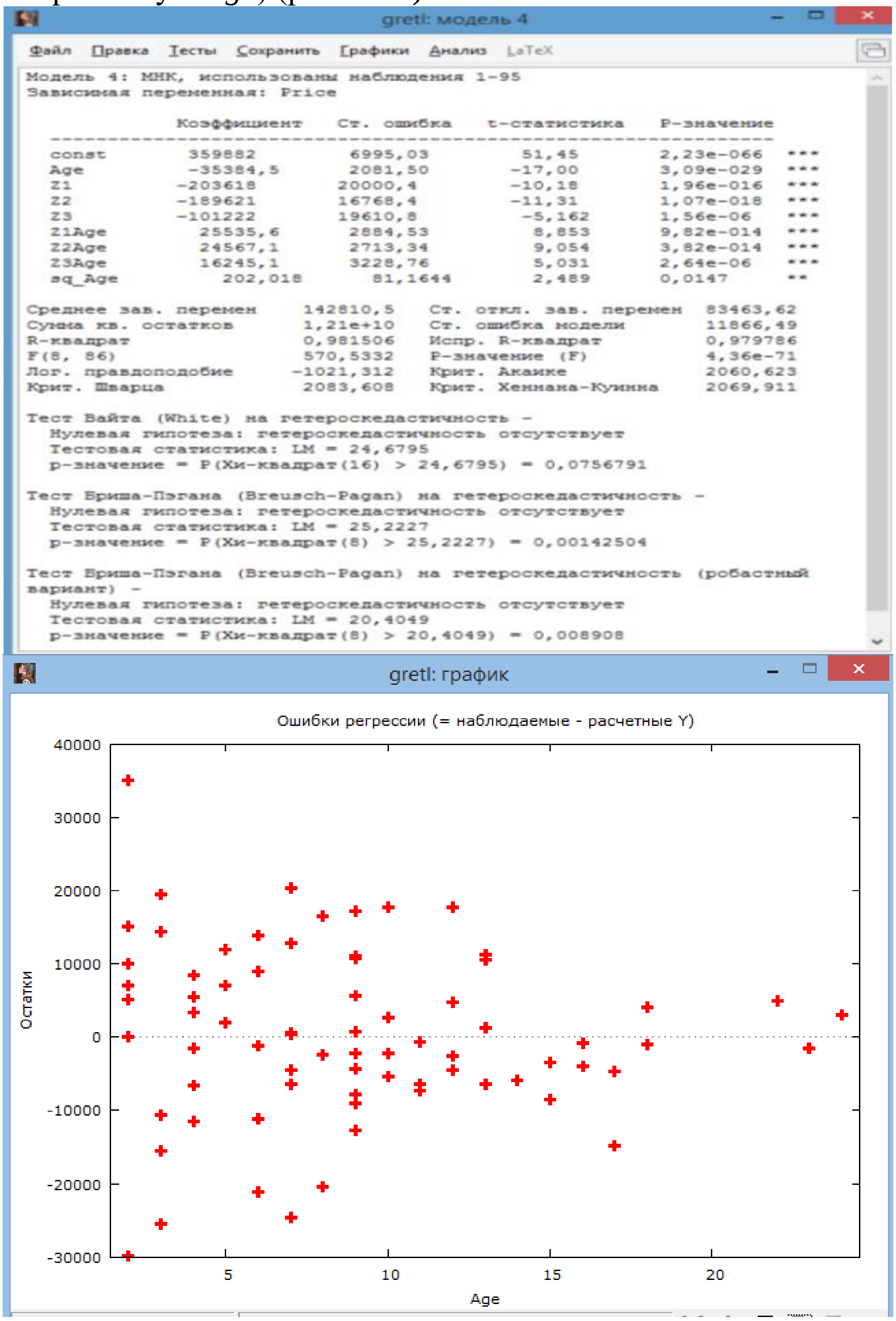


Рис. 3.13. Модель регрессии с квадратом количественной переменной

$$Y = 359882 - 35384{,}5 Age + 202{,}018 Age^2 - 203618 Z1 - 189621 Z2 -$$
$$-101222 Z3 + 25535{,}6 Z1 Age + 24567{,}1 Z2 Age + 16245{,}1 Z3 Age + \varepsilon .$$

МодельVAZ2110:

$$Y = (359882 - 203618) + (-35384{,}5 + 25535{,}6) Age + 202{,}018 Age^2 + \varepsilon .$$

МодельVAZ2115:

$$Y = (359882 - 189621) + (-35384{,}5 + 24567{,}1)Age + 202{,}018Age^2 + \varepsilon .$$

МодельVAZ1119:

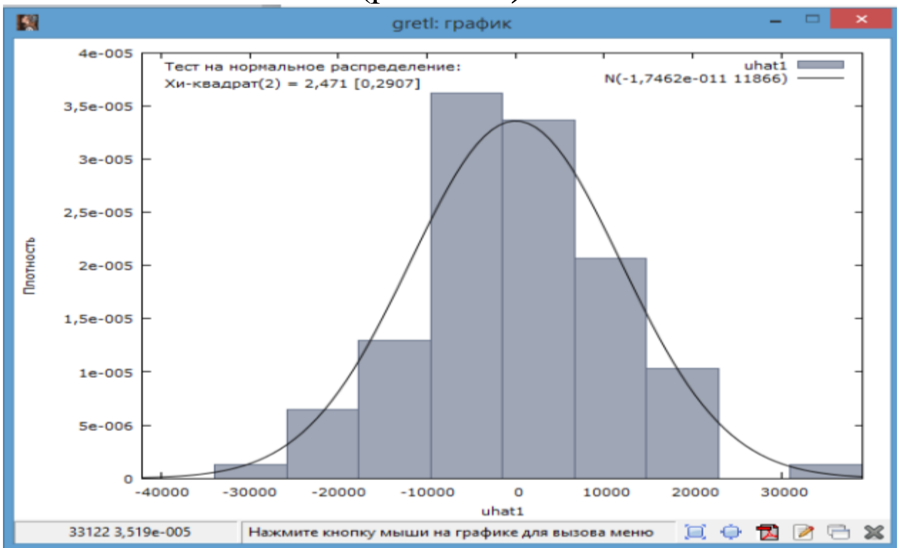
$$Y = (359882 - 101222) + (-35384{,}5 + 16245{,}1)Age + 202{,}018Age^2 + \varepsilon .$$

МодельVAZ2181:

$$Y = 359882 - 35384{,}5Age + 202{,}018Age^2 + \varepsilon .$$

$R^2 = 0{,}98$. По статистике Фишера и Стьюдента модель и коэффициенты значимы с вероятностью 99%. Тест Вайта подтверждает гомоскедастичность в остатках ($p > 0{,}05$), а тесты Бриша-Пэгана указывают на наличие гетероскедастичности в остатках ($p < 0{,}05$).

4.1. Проверка остатков на нормальность в окне модели: Тесты/Нормальность остатков… (рис. 3.14).



gretl: распределение остатков

```
Распределение частот для uhat1, наблюдения 1-95
Количество столбцов = 9, среднее = -1,74623e-011, ст. откл. = 11866,5

       интервал          середина   частота    отн.     инт.

           < -25859,     -29921,        1      1,05%    1,05%
  -25859, - -17734,     -21796,        5      5,26%    6,32% *
  -17734, - -9608,7     -13671,       10     10,53%   16,84% ***
  -9608,7 - -1483,7     -5546,2       28     29,47%   46,32% **********
  -1483,7 -  6641,3     2578,8        26     27,37%   73,68% *********
   6641,3 -  14766,     10704,        16     16,84%   90,53% ******
   14766, -  22891,     18829,         8      8,42%   98,95% ***
   22891, -  31016,     26954,         0      0,00%   98,95%
         >=  31016,     35079,         1      1,05%  100,00%

Нулевая гипотеза - нормальное распределение:
Хи-квадрат(2) = 2,471 p-значение 0,29072
```

Рис. 3.14. Тест на нормальное распределение остатков в модели с квадратом количественной переменной

Подтверждается соблюдение нормального закона распределения остатков ($p > 0{,}05$).

4.5. На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.15).

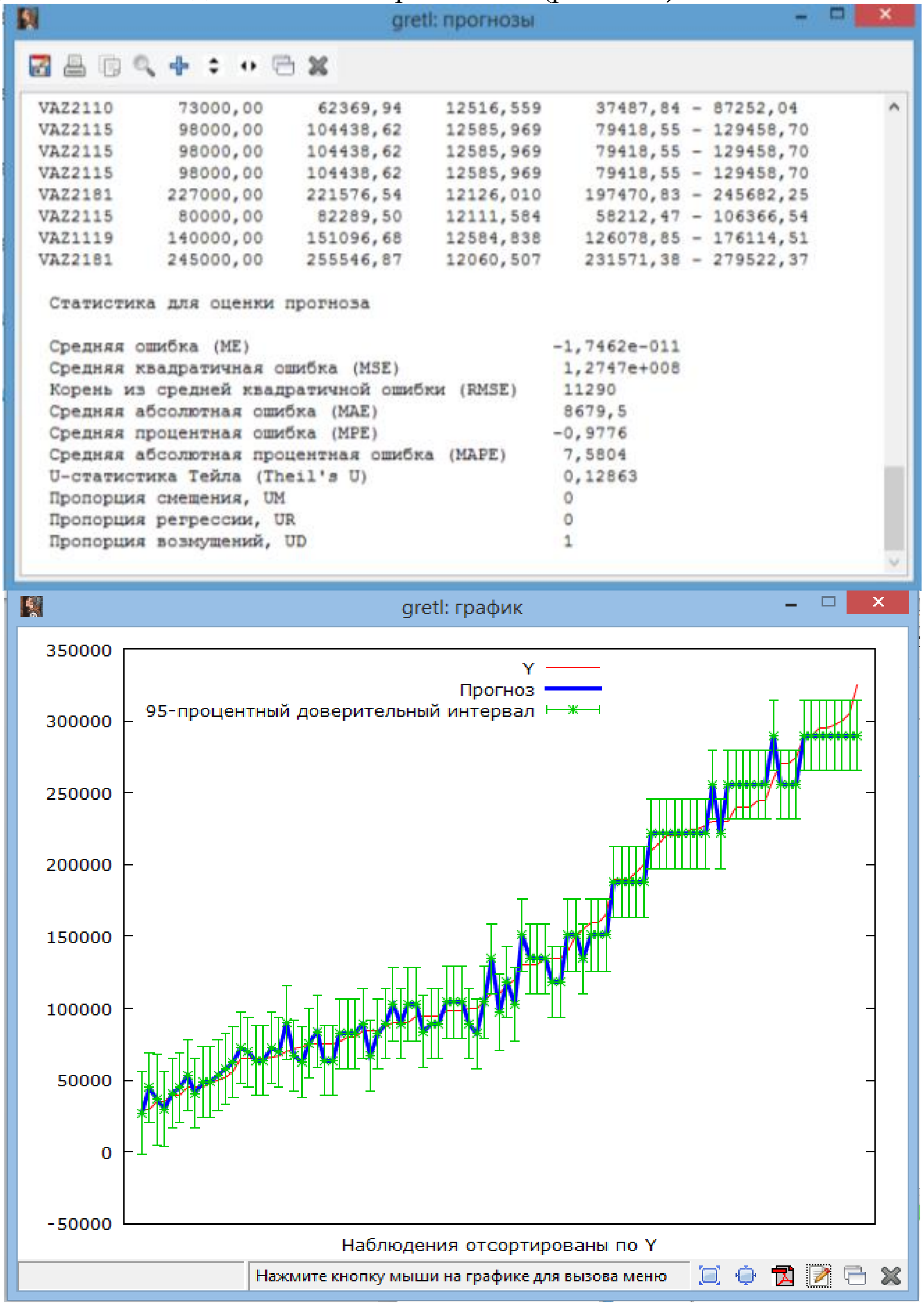


Рис. 3.15. Анализ прогнозных качеств модели с квадратом количественной переменной

Средняя процентная ошибка: $MPE = -0{,}98$. Средняя абсолютная процентная ошибка: $MAPE = 7{,}58$.

Итоговые результаты моделирования приведены в таблице 3.2.

Таблица 3.2

Сводная таблица результатов моделирования

Тип модели	Вид модели	Se	R^2	MPE	$MAPE$
Без фиктивных переменных	$Y = 262133 - 14551{,}5Age + \varepsilon$	40636,4	0,76	0,94	35,99
С фиктивными переменными	$Y = 266227 - 6298{,}04Age - -120533Z1 - 117217Z2 - -90057{,}8Z3 + \varepsilon$	22160,61	0,93	-0,07	13,45
С коррекцией на гетероскедастичность	$Y = 260740 - 6190{,}99Age - -119388Z1 - 114785Z2 - -85994{,}7Z3 + \varepsilon$	1,437	0,93	2,18	12,9
С угловым коэффициентом	$Y = 357753 - 34007{,}8Age - -241775Z1 - 224094Z2 - -110085Z3 + 30220{,}9Z1Age + +28947{,}1Z2Age + +17885{,}4Z3Age + \varepsilon$	12215,45	0,98	-0,9	9,09
С квадратом количественной переменной	$Y = 359882 - 35384{,}5Age + +202{,}018Age^2 - 203618Z1 - -189621Z2 - 101222Z3 + +25535{,}6Z1Age + 24567{,}1Z2Age +16245{,}1Z3Age + \varepsilon$	11866,49	0,98	-0,98	7,58

Задание 3.2. Провести регрессионный анализ динамики сбережений в зависимости от дохода за 20 лет (таблица 7.3), учитывая структурный сдвиг на 12-м году наблюдений, что стимулировало к большим сбережениям по сравнению с первым этапом рассматриваемого интервала. *Y* – сбережения, *X* – доход.

Таблица 3.3

Сбережения	4,7	6,1	6,5	6,8	5,2	6,5	7,5	8	9	9,1
Доход	100	105	108	111	115	122	128	135	143	142
Сбережения	8,7	12	16,2	18,5	18	17,6	20	23	22,5	24,3
Доход	147	155	167	177	188	195	210	226	238	255

Методические указания для выполнения задания

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «Занятие_Чоу.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1. Строим график исходных данных: Вид/График/Разброс *X*-*Y* (рис. 3.16).

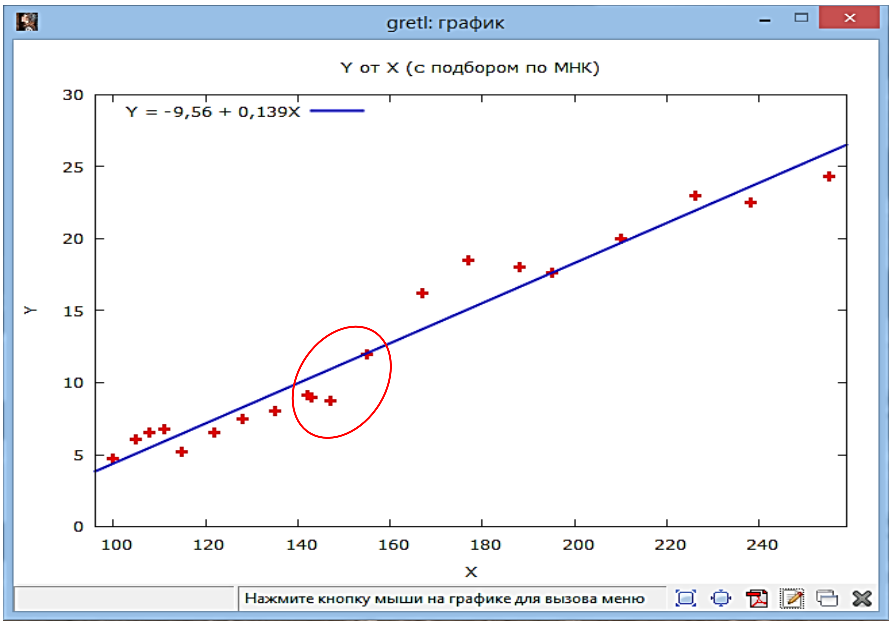


Рис. 3.16. Динамика сбережений в зависимости от дохода

2. Построение уравнения регрессии без учета структурного сдвига (без фиктивной переменной): Модель/Метод наименьших квадратов… Проверка остатков на соблюдение предпосылок МНК: В окне модели: Тесты/Гетероскедастичность… Затем для визуального анализа остатков: В окне модели: Графики/График остатков/В зависимости от *X* (рис. 3.17).

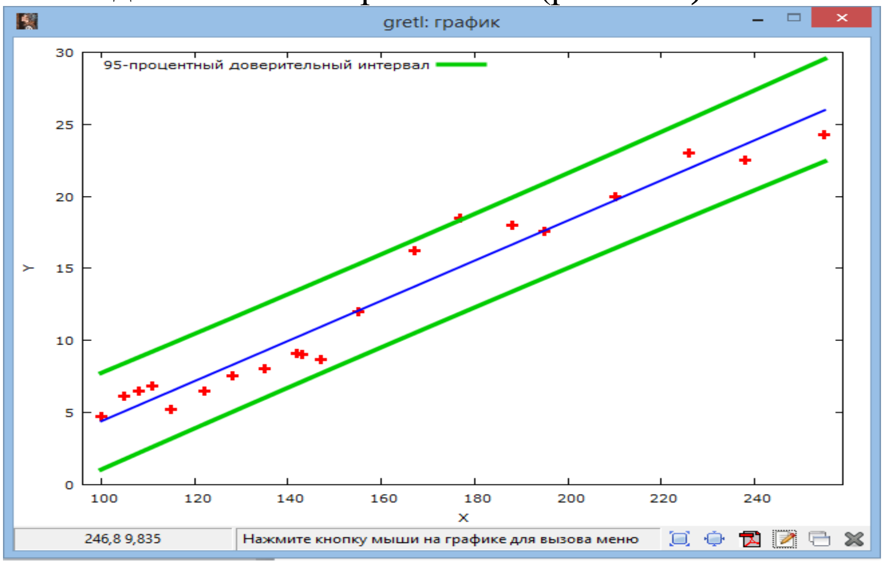
```
Файл  Правка  Тесты  Сохранить  Графики  Анализ  LaTeX

Модель 1: МНК, использованы наблюдения 1-20
Зависимая переменная: Y

             Коэффициент   Ст. ошибка    t-статистика   P-значение
  ---------------------------------------------------------------
  const       -9,56407     1,20850         -7,914       2,86e-07    ***
  X            0,139400    0,00733163      19,01        2,30e-013   ***

Среднее зав. перемен    12,51000   Ст. откл. зав. перемен   6,708275
Сумма кв. остатков      40,55252   Ст. ошибка модели        1,500972
R-квадрат               0,952571   Испр. R-квадрат          0,949936
F(1, 18)                361,5159   Р-значение (F)           2,30e-13
Лог. правдоподобие     -35,44743   Крит. Акаике             74,89485
Крит. Шварца            76,88632   Крит. Хеннана-Куинна     75,28361

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 2,69362
  p-значение = P(Хи-квадрат(2) > 2,69362) = 0,260069

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 0,377617
  p-значение = P(Хи-квадрат(1) > 0,377617) = 0,538881

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 0,440521
  p-значение = P(Хи-квадрат(1) > 0,440521) = 0,506871

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 1,73254
  p-значение = 0,420517
```

Рис. 3.17. Модель регрессии без фиктивной переменной

$$Y = -9{,}564 + 0{,}139 + \varepsilon,\ R^2 = 0{,}95.$$

По статистике Фишера и Стьюдента модель и коэффициенты значимы с вероятностью 99%. Остатки гомоскедастичны ($p > 0{,}05$).

2.1. На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.18).



Статистика для оценки прогноза

Средняя ошибка (ME)	3,1086e-016
Средняя квадратичная ошибка (MSE)	2,0276
Корень из средней квадратичной ошибки (RMSE)	1,4239
Средняя абсолютная ошибка (MAE)	1,1831
Средняя процентная ошибка (MPE)	-1,5471
Средняя абсолютная процентная ошибка (MAPE)	11,507
U-статистика Тейла (Theil's U)	0,83241
Пропорция смещения, UM	0
Пропорция регрессии, UR	0
Пропорция возмущений, UD	1

Рис. 3.18. Анализ прогнозных качеств модели без фиктивной переменной

Средняя процентная ошибка: $MPE = -1{,}54$. Средняя абсолютная процентная ошибка: $MAPE = 11{,}51$.

3. Построение уравнения регрессии с учетом структурного сдвига (с фиктивной переменной и дифференциальным угловым коэффициентом). Создадим новую переменную D. D=0 для наблюдений с 1 по 11, D=1 для наблюдений с 12 по 20 (ввод цифровых меток выполняем вручную). Создадим новую переменную *DX*: Добавить/Добавить новую переменную… Ввести вручную формулу $DX = D \cdot X$. Построим модель с фиктивной переменной: Модель/Метод наименьших квадратов… $Y = -3{,}001 + 0{,}08X + 0{,}43D + 0{,}03DX + \varepsilon$. После структурного сдвига: $Y = (-3{,}001 + 0{,}43) + (0{,}08 + 0{,}03)X + \varepsilon$. $R^2 = 0{,}98$. По статистике Фишера модель значима, по статистике Стьюдента значим с вероятностью 99% только коэффициент при количественной переменной. Тест Вайта подтверждает гомоскедастичность в остатках ($p > 0{,}05$), тест Бриша-Пагана и тест Коэнкера показывают гетероскедастичность ($p < 0{,}05$) (рис. 3.19).

```
Модель 2: МНК, использованы наблюдения 1-20
Зависимая переменная: Y

              Коэффициент    Ст. ошибка    t-статистика    P-значение
  ---------------------------------------------------------------
  const       -3,00135       2,33005        -1,288          0,2160
  X            0,0819431     0,0187456       4,371          0,0005    ***
  D            0,428249      3,15136         0,1359         0,8936
  DX           0,0258746     0,0214451       1,207          0,2451

Среднее зав. перемен    12,51000   Ст. откл. зав. перемен  6,708275
Сумма кв. остатков      15,69871   Ст. ошибка модели       0,990540
R-квадрат               0,981639   Испр. R-квадрат         0,978197
F(3, 16)                285,1425   Р-значение (F)          4,28e-14
Лог. правдоподобие     -25,95723   Крит. Акаике            59,91446
Крит. Шварца            63,89739   Крит. Хеннана-Куинна    60,69197

Исключая константу, наибольшее р-значение получено для переменной 3 (D)

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 10,96
  р-значение = P(Хи-квадрат(5) > 10,96) = 0,0521779

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 10,6517
  р-значение = P(Хи-квадрат(3) > 10,6517) = 0,0137661

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 8,62614
  р-значение = P(Хи-квадрат(3) > 8,62614) = 0,0346976

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 3,3161
  р-значение = 0,19051
```

Рис. 3.19. Модель регрессии с фиктивной переменной

На основе имеющихся наблюдений оценим прогнозные качества модели. В окне модели: Анализ/Прогнозы… (рис. 3.20).


```
Статистика для оценки прогноза

Средняя ошибка (ME)                                1,1546e-015
Средняя квадратичная ошибка (MSE)                  0,78494
Корень из средней квадратичной ошибки (RMSE)       0,88597
Средняя абсолютная ошибка (MAE)                    0,68826
Средняя процентная ошибка (MPE)                   -0,75478
Средняя абсолютная процентная ошибка (MAPE)        6,6955
U-статистика Тейла (Theil's U)                     0,53217
Пропорция смещения, UM                             0
Пропорция регрессии, UR                            0
Пропорция возмущений, UD                           1
```

Рис. 3.20. Анализ прогнозных качеств модели с фиктивной переменной

Средняя процентная ошибка: $MPE = -0,75$. Средняя абсолютная процентная ошибка: $MAPE = 6,69$.

4. Проведем тест Чоу на целесообразность учета структурного сдвига с помощью фиктивной переменной. В окне модели: Тесты/Тест Чоу (рис. 3.21).

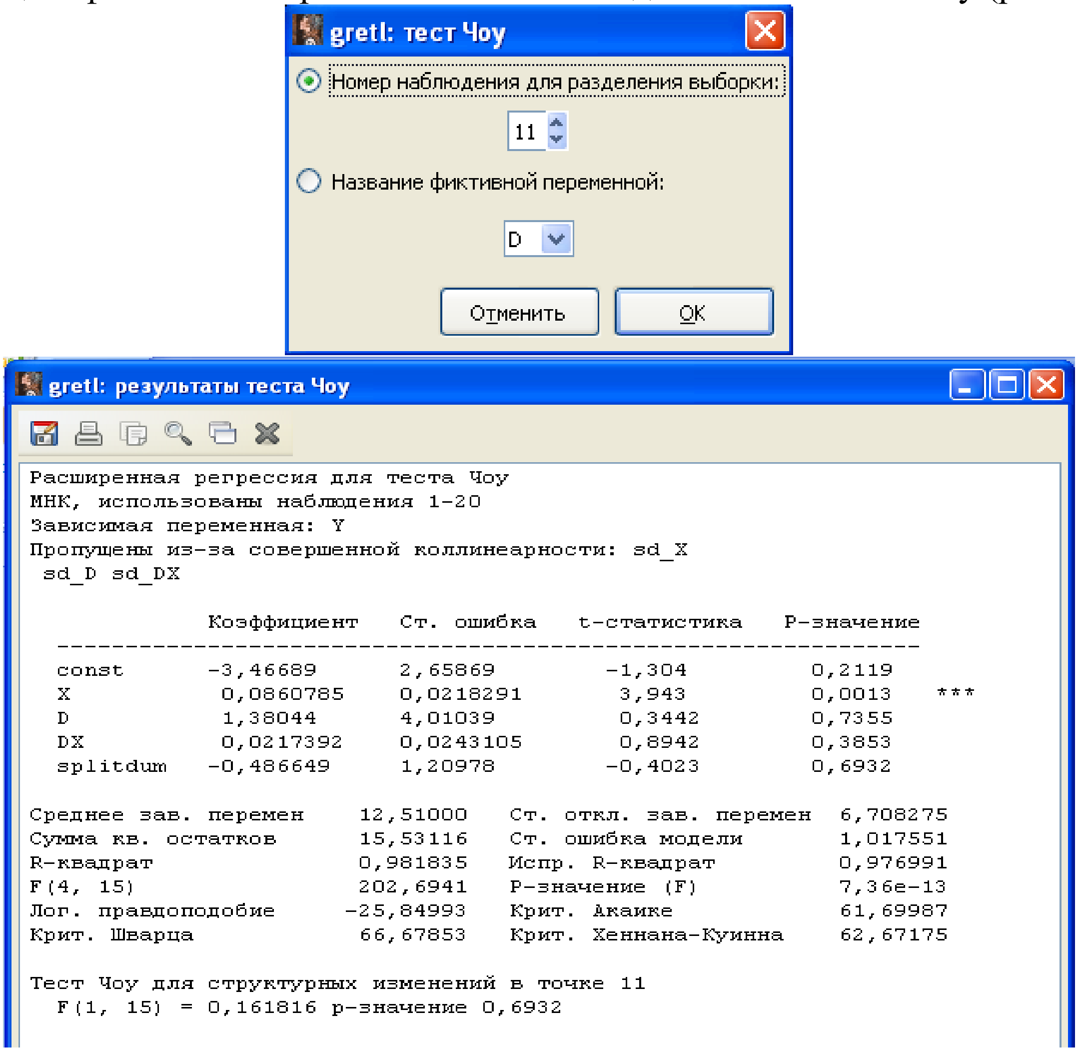


Рис. 3.21. Тест Чоу

Нулевая гипотеза теста Чоу: учет структурных изменений не требуется. В модели необходимость учета структурных сдвигов не подтвердилась ($p > 0,05$).

Задания для самостоятельной работы

Задание 3.3. На основе данных о динамике среднегодовых цен на нефть различных марок (долл./баррель) за период с 1994 по 2005 г. (таблица 3.4) проанализировать, зависит ли средний ежегодный прирост цены от сорта нефти и насколько существенен этот фактор.

Таблица 3.4

Годы	Дубай	Брент	Нигерийская форкадос	Западно-техасская нефть
1994	14,74	15,82	16,25	17,21
1995	16,10	17,02	17,26	18,42
1996	18,52	20,67	21,16	22,16
1997	18,23	19,09	19,33	20,61
1998	12,21	12,72	12,62	14,39
1999	17,25	17,97	18,00	19,31
2000	26,20	28,50	28,42	30,37
2001	22,81	24,44	24,23	25,93
2002	23,74	25,02	25,04	26,16
2003	26,78	28,83	28,66	31,07
2004	33,64	38,27	38,13	41,49
2005	57,65	63,48	62,20	66,49

Летом 2005 г. был катастрофический ураган «Катрина». Как сказалось данное явление на изменении цен? Как отразить это в модели?

Задание 3.4. На основе квартальных данных о реальном объеме произведенного валового внутреннего продукта в основных ценах (таблица 3.5) построить модель с фиктивными переменными: $D1 = 1$ – если первый квартал, $D2 = 1$ – если второй квартал, $D3 = 1$ – если третий квартал. Определить, влияют ли сезонные факторы на объем ВВП. Построить модель прогноза продукции сельского хозяйства в следующем году с учетом сезонных факторов.

Таблица 3.5

Год	Квартал	ВВП, млрд. руб.	в т.ч. сельское хозяйство, млрд. руб.
1	1	1468,4	48,9
	2	1542,5	62,3
	3	1810,1	238,5
	4	1709,4	70,5
2	1	1534,5	51,6
	2	1617,2	67,4
	3	1916,6	272,3
	4	1782,6	76,8
3	1	1589,1	56,6
	2	1686,4	72,2
	3	1967,7	263,1

	4	1895,3	89,8
4	1	1715,2	59,0
	2	1818,7	74,6
	3	2131,1	275,3
	4	2045,6	100,2
5	1	1837,3	58,6
	2	1960,7	74,6
	3	2277,0	288,7
	4	2186,4	102,0

Задание 7.5. На основе квартальных данных с 2009 по 2014 годы было получено следующее уравнение регрессии, описывающее зависимость цены на товар Yt от нескольких факторов:

$Yt = 3,5 + 0,4Xt + 1,1Wt$, $ESS = 70,4$, $RSS = 40,5$, $mb1 = 0,001$, $mb2 = 0,01$.

Когда в уравнение были добавлены фиктивные переменные, соответствующие первым трем кварталам года, величина ESS выросла до 86,4. Напишите спецификацию уравнения регрессии с учетом сезонности. Сформулируйте и проверьте гипотезу о наличии сезонности (уровень значимости 5%).

Контрольные вопросы:

1. В чем преимущества фиктивных переменных?
2. В чем суть ANOVA-моделей?
3. В чем суть ANCOVA-моделей?
4. В чем состоит правило применения фиктивных переменных?
5. Какой смысл имеет дифференциальный свободный член?
6. Какой смысл имеет дифференциальный угловой коэффициент?
7. Какова идея теста Чоу?
8. Как сезонные переменные применяются для устранения сезонного фактора?

3. Одномерные тренд-сезонные модели временных рядов

Интерактивная форма проведения занятия

Занятие проводится в форме кресельного кейс-метода. Условие задачи содержит реальный практический смысл. Поэтому преподаватель предлагает выполнить комплекс эконометрических расчетов, идентифицировать и верифицировать результаты с целью их экономической интерпретации.

На первом этапе преподаватель объявляет группы из двух-трех магистрантов для совместного выполнения практического задания. Роль преподавателя на данном этапе заключается в постоянном поддержании активного внутригруппового взаимодействия, снятии напряженности во взаимоотношениях между участниками, оперативном вмешательстве в случае возникновения непредвиденных трудностей, а также в целях пояснения новых положений учебной программы.

На втором этапе преподаватель назначает лидера для руководства ходом обсуждения результатов выполнения задания. Лидер, применяя уникальное сочетание компьютерных и традиционных методов организации учебной деятельности демонстрирует основные формулы, используемые в решении задания. Совместно с преподавателем лидер руководит групповым обсуждением области применения формул, полученных эконометрических оценок и их качества, экономической интерпретацией результатов. Преподаватель побуждает магистрантов к обоснованию в аудитории полученных результатов, конструктивной критике и поддержке выводов одногруппников с целью формирования навыков изучения реальных социально-экономических явлений и процессов, проблемных ситуаций, высказывания и отстаивания собственной точки зрения, концентрируя внимание на следующих вопросах:

1. Виды моделей регрессии для временных рядов.
2. Тренд-сезонные модели временных рядов. Автокорреляция уровней ряда.
3. Ложная регрессия в моделях временных рядов.

В конце занятия преподаватель подводит итоги и оценивает каждого студента в зависимости от его участия в выполнении заданий и обсуждении вопросов.

Расчетные формулы

Компоненты динамического ряда позволяют представить уровень динамического ряда в виде аддитивной или мультипликативной моделей:

$$Y_t = T + S + \xi;$$

$$Y_t = T \cdot S \cdot \xi.$$

Если амплитуда сезонных колебаний постоянна, то применяется аддитивная модель. Если амплитуда колебаний изменяется во времени, то рассматривается мультипликативная модель.

Построение тренд-сезонных моделей временных рядов сводится к расчету значений *T*, *S* или *E* для каждого уровня ряда. Процесс построения модели включает в себя следующие этапы:

1. Выравнивание исходного ряда методом скользящей средней.

2. Расчет значений сезонной компоненты *S*.

3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных $(T+E)$ в аддитивной или $(T \cdot E)$ в мультипликативной модели.

4. Аналитическое выравнивание уровней $(T+E)$ или $(T \cdot E)$ и расчет значений *T* с использованием полученного уравнения тренда.

5. Расчет полученных по модели значений $(T+S)$ или $(T \cdot S)$.

6. Расчет абсолютных и относительных ошибок.

Применение трендовых временных рядов в качестве зависимой и объясняющих переменных имеет важную особенность. Пусть $Y_t = \alpha_0 + \alpha_1 \cdot t + u_t$, $X_t = \gamma_0 + \gamma_1 \cdot t + v_t$ и оценивается линейная модель регрессии $Y_t = \beta_0 + \beta_1 \cdot x_t + e$. Тогда невключение значимого фактора времени (*t*) приводит к смещению *OLS*-оценок параметров регрессии, в частности, коэффициент β_1 может оказаться значимым, хотя из экономических соображений факторы должны быть независимыми. Описанная проблема называется ложной регрессией. Необходимо учесть тренд (включить в модель значимый фактор времени) и оценивать регрессию $Y_t = \beta_0 + \beta_1 \cdot x_t + \beta_2 \cdot t + e$.

Задание для выполнения в аудитории

Задание 4.1. На основе поквартальных данных о выручке от продаж, млн. руб. (таблица 4.1), необходимо построить аддитивную тренд-сезонную модель. Используя построенную модель выполнить прогнозирование выручки на 1, 2, 3, 4 кварталы следующего года.

Таблица 4.1

t	1	2	3	4	5	6	7	8	9
Y_t	6	4,4	5	9	7,2	4,8	6	10	8
t	10	11	12	13	14	15	16		
Y_t	5,6	6,4	11	9	6,6	7	10,8		

Методические указания для выполнения задания

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «Занятие_Тренд.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1.

2. Интерпретировать данные как временной ряд (рис. 4.1).

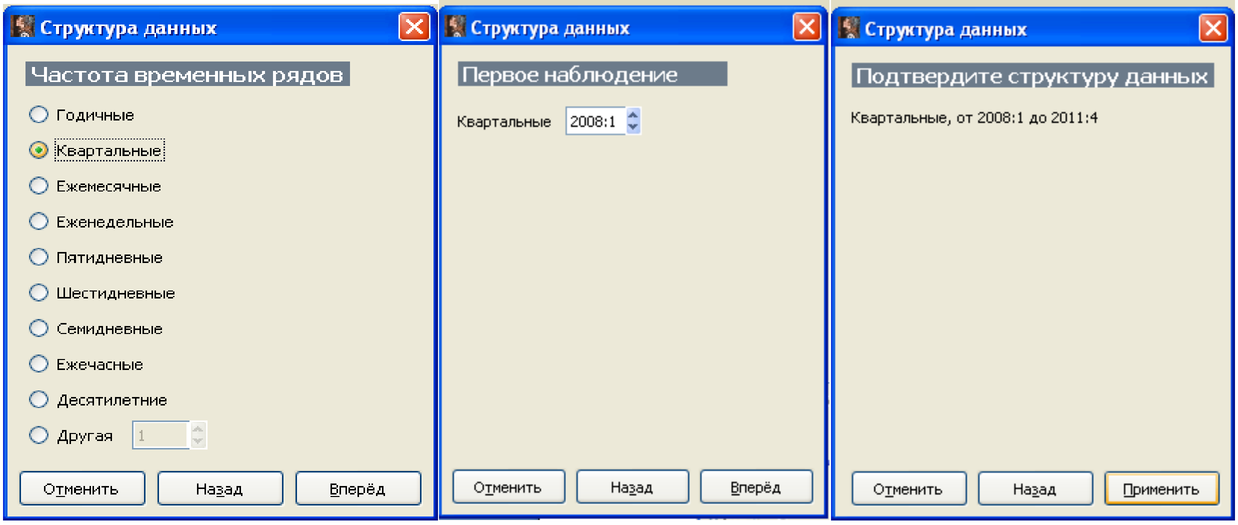


Рис. 4.1. Окно импорта временного ряда

1. Проведем визуальный анализ временного ряда: Вид/График/График временного ряда (рис. 4.2).

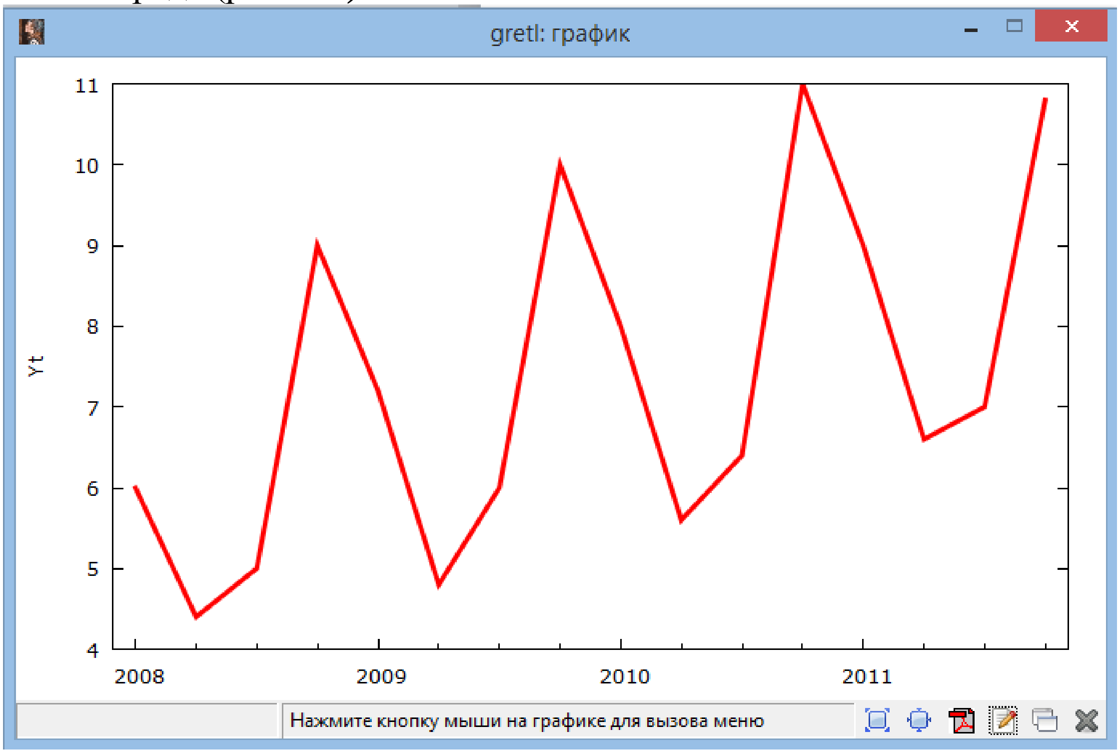
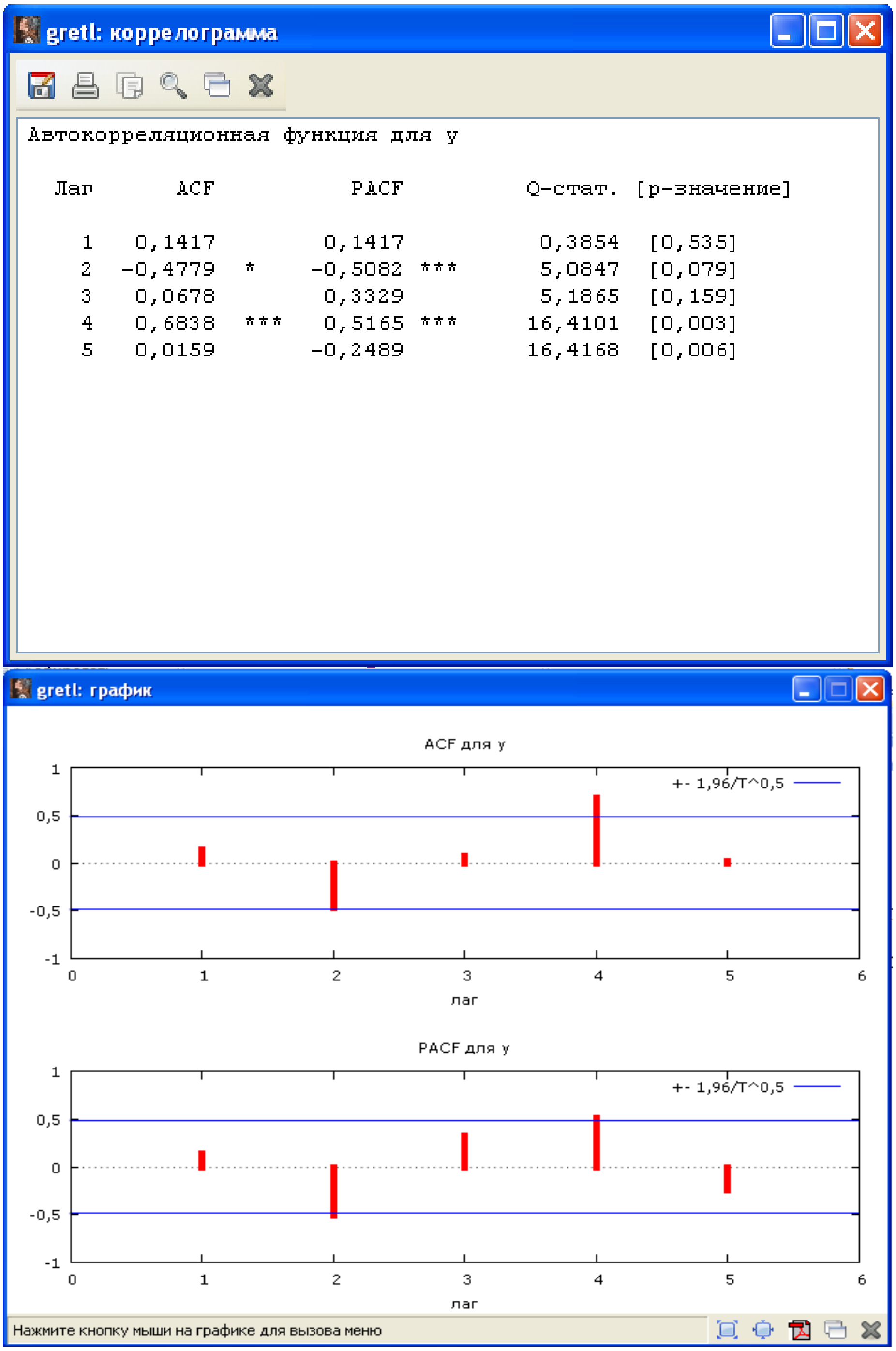


Рис. 4.2. График исходного временного ряда

График показывает, что временной ряд имеет тренд и сезонные колебания.

4. Проведем анализ автокорреляционной функции: Переменная / Коррелограмма, лаг=5 (рис. 4.3).


gretl: коррелограмма

Автокорреляционная функция для y

Лаг	ACF		PACF		Q-стат.	[p-значение]
1	0,1417		0,1417		0,3854	[0,535]
2	-0,4779	*	-0,5082	***	5,0847	[0,079]
3	0,0678		0,3329		5,1865	[0,159]
4	0,6838	***	0,5165	***	16,4101	[0,003]
5	0,0159		-0,2489		16,4168	[0,006]

Рис. 4.3. Коррелограмма исходного временного ряда

На автокорреляционной функции видно, что значимы коэффициенты автокорреляции 2 и 4 лага, значит, временной ряд имеет сезонные колебания с периодичностью в 4 квартала (год).

5. Введем условную переменную time: Добавить/Временной тренд.

6. Определяем центрированную скользящую среднюю: Переменная/Фильтр/Простая скользящая средняя (рис. 4.4).

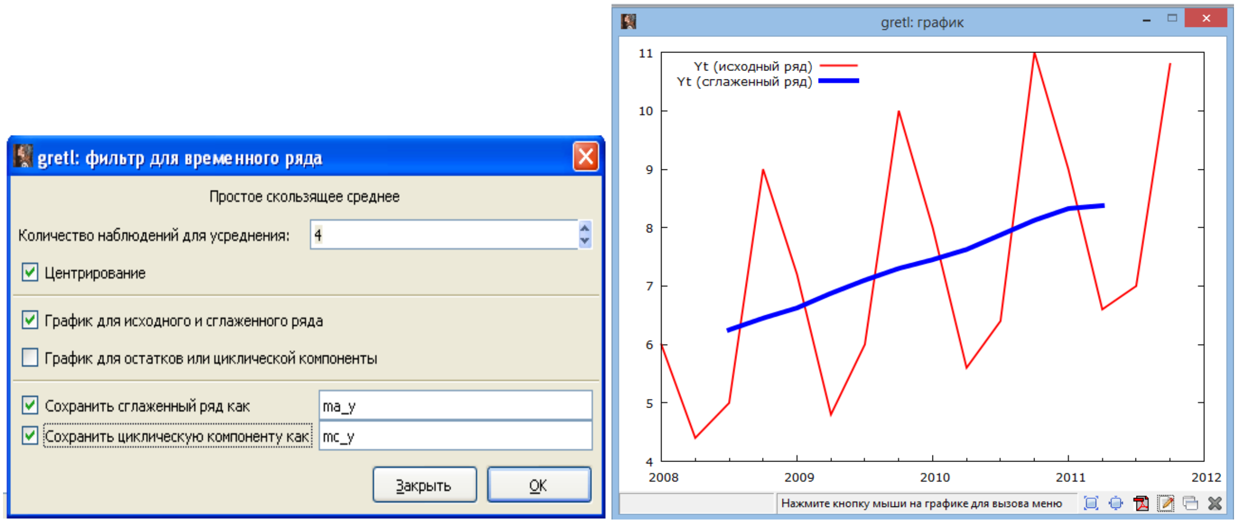


Рис. 4.4. График скользящей средней

Чтобы обнулить пропущенные значения ma_y, mc_y: Данные/Изменить значения.

7. Добавляем фиктивные переменные для периодов $dq1$, $dq2$, $dq3$, $dq4$: Добавить/Фиктивные переменные для периодов.

8. Находим частные подвыборки для оценки сезонных колебаний по 4 кварталам $OSK1$, $OSK2$, $OSK3$, $OSK4$: Добавить / Добавить новые переменные (рис. 4.5).

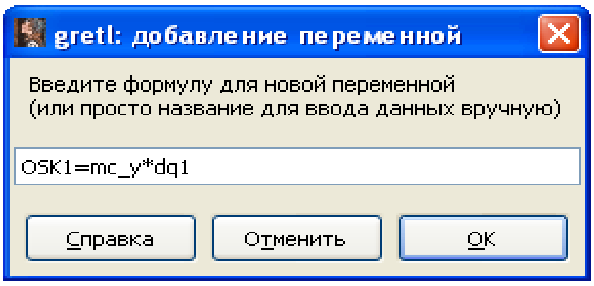


Рис. 4.5. Окно добавления новой переменной для частных подвыборок

9. Определяем сезонные коэффициенты $SK1$, $SK2$, $SK3$, $SK4$ через скаляры (рис. 4.6).

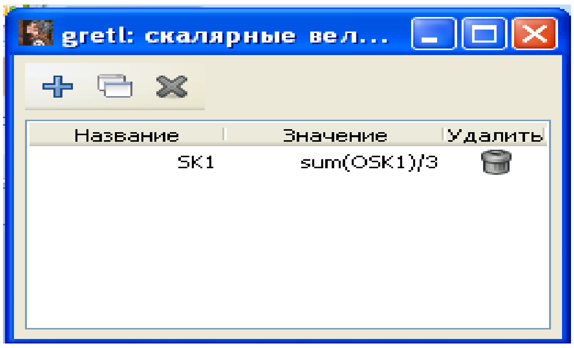


Рис. 4.6. Окно ввода скаляров

Находим сумму сезонных коэффициентов $S0$ и определяем корректировочный коэффициент, разделив $S0$ на 4 (рис. 4.7).

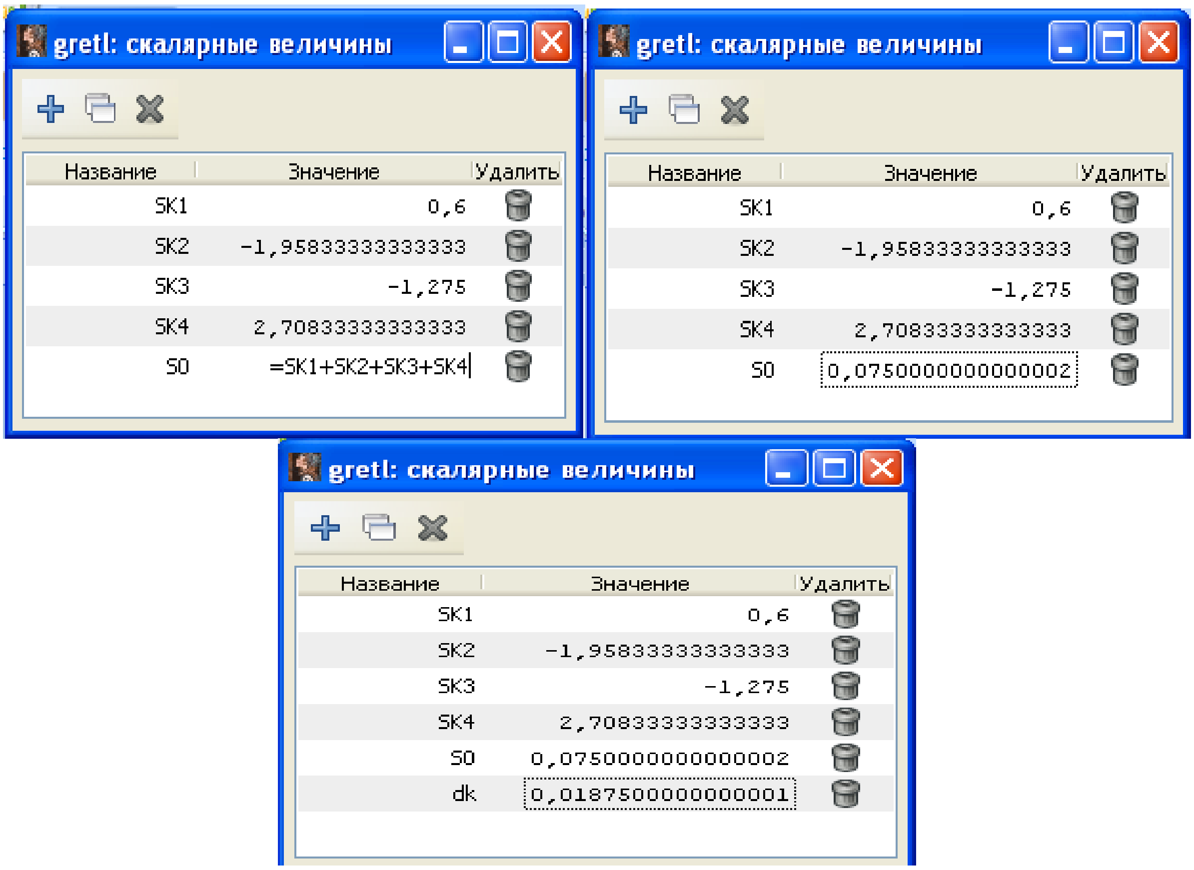


Рис. 4.7. Окно скаляров с корректировочным коэффициентом

Находим скорректированные сезонные коэффициенты $SSK1$, $SSK2$, $SSK3$, $SSK4$: $SSK1 = SK1 - dk$ и т.д. (рис. 4.8).

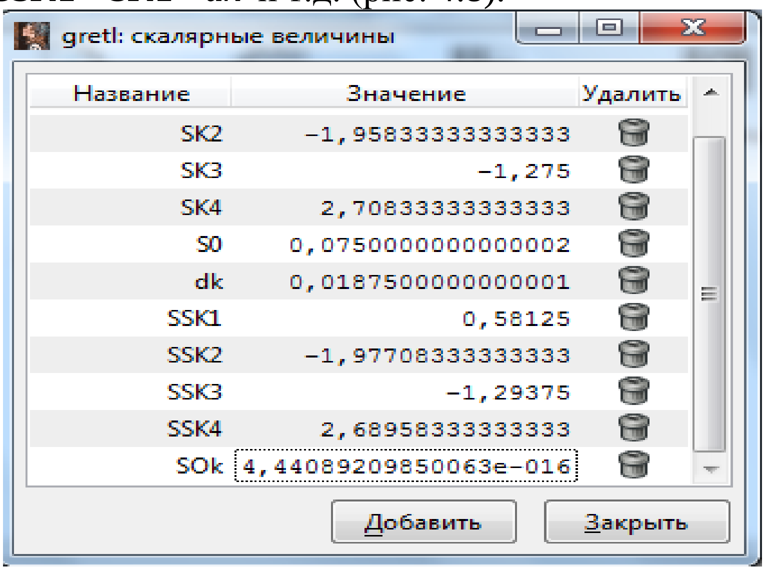


Рис. 4.8. Окно скаляров с сезонными коэффициентами

10. Формируем значения сезонной компоненты: Добавить / Добавить новую переменную (рис. 4.9, рис. 4.10).

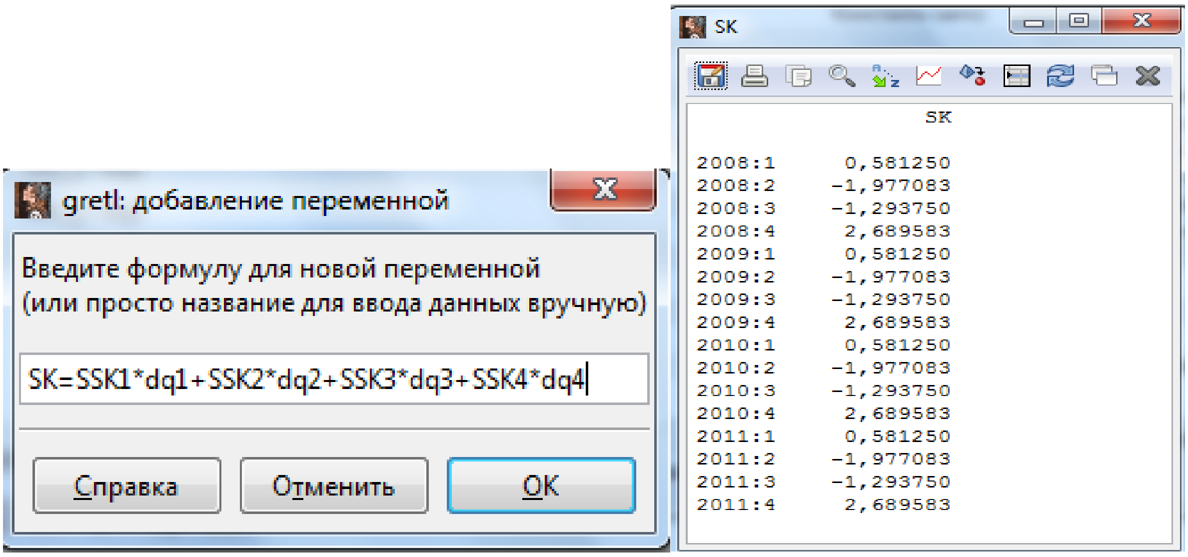


Рис. 4.9. Окно добавления новой переменной для сезонной компоненты

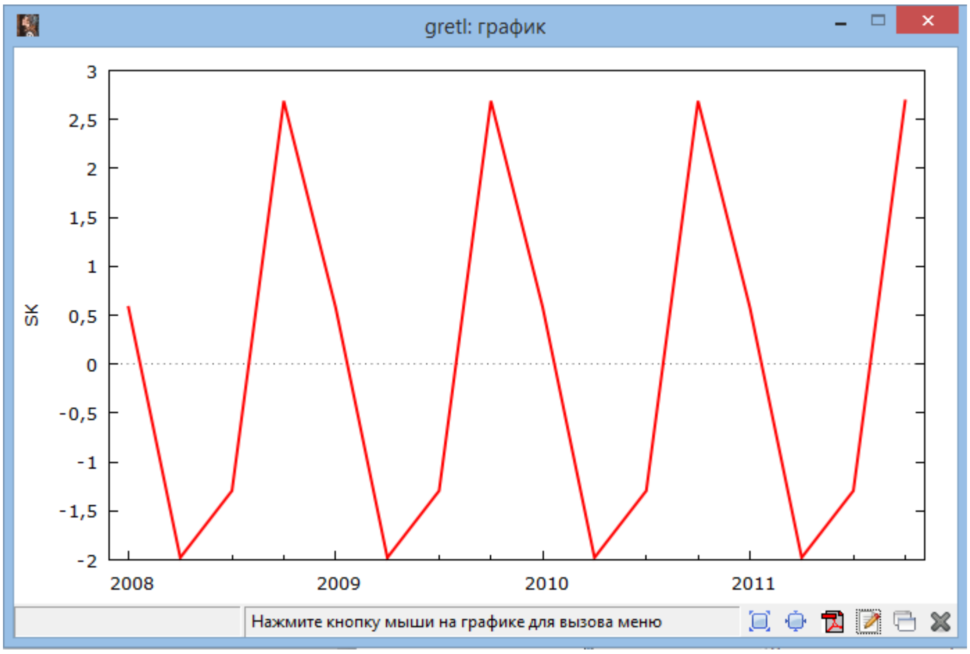


Рис. 4.10. График сезонной компоненты SK

11. Находим $TK + E = Y - SK$: Добавить / Добавить новую переменную (рис. 4.11).

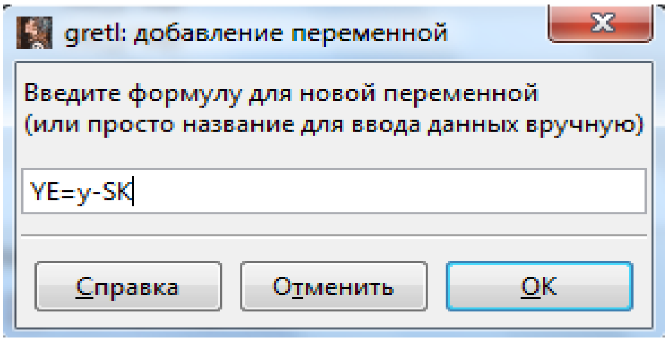


Рис. 4.11. Окно добавления новой переменной для трендовой и случайной компонент

12. Находим параметры линейного тренда обычным МНК: Модель/Метод наименьших квадратов (рис. 4.12).

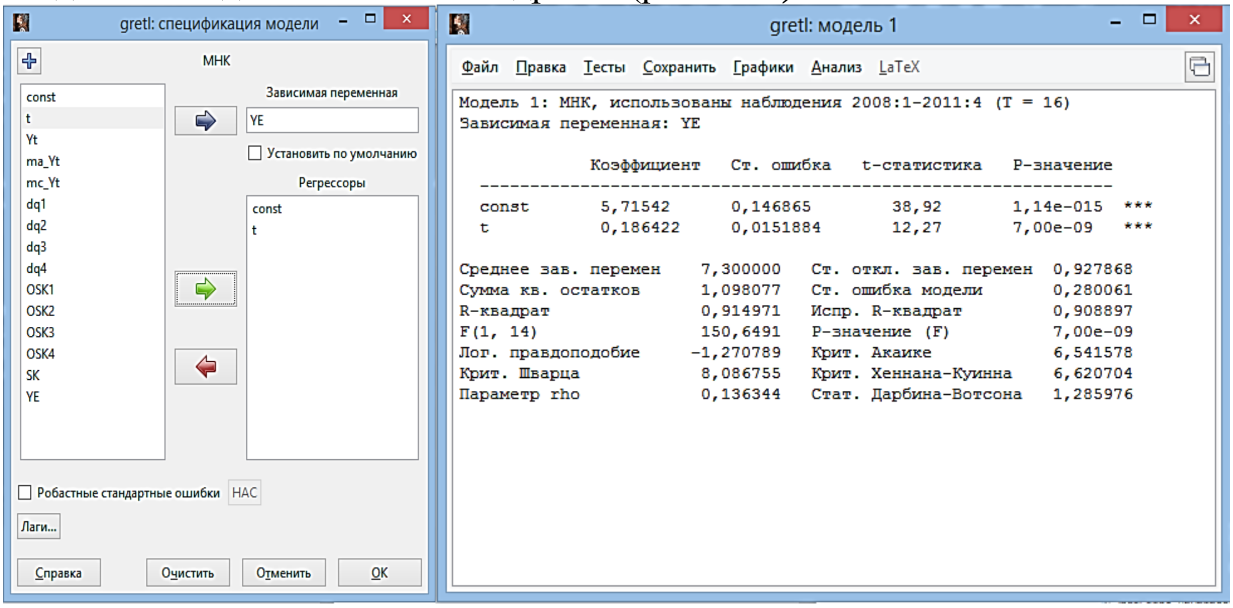


Рис. 4.12. Модель линейного тренда

Сохраним модель: В окне модели: Файл / Сохранить в текущей сессии.

13. Находим трендовую компоненту.

Вводим скаляры: $a0 = 5{,}71542$, $a1 = 0{,}186422$. Определяем значения трендовой компоненты TK: Добавить/Добавить новую переменную: $TK = a0 + a1 \cdot time$ (рис. 4.13).

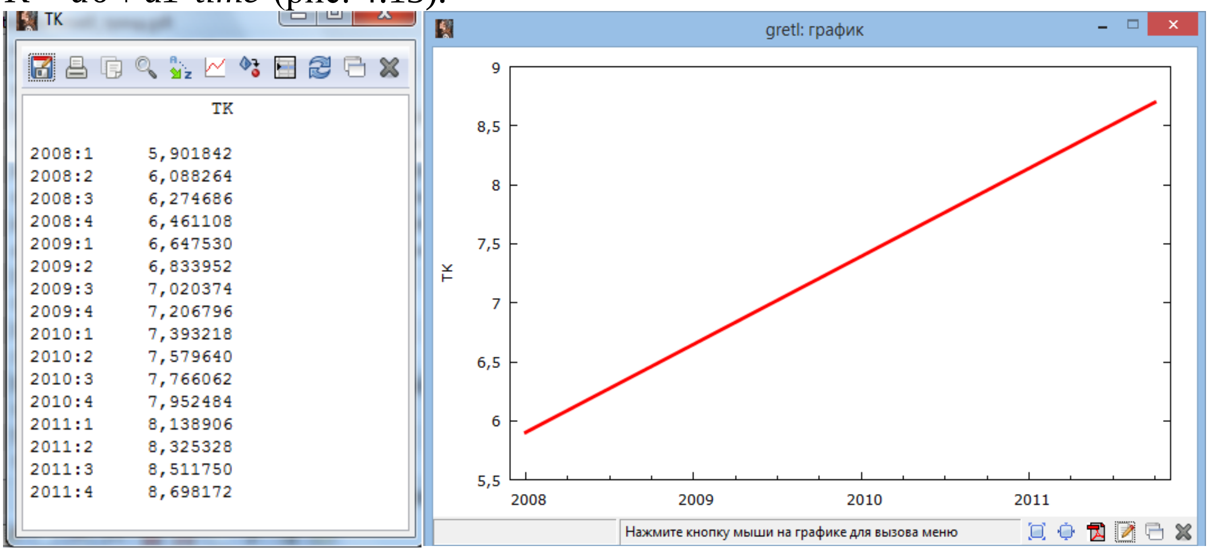


Рис. 4.13. Значения и график трендовой компоненты *TK*

14. Определяем значения суммы трендовой компоненты и сезонной компоненты: $TSK = TK + SK$ (рис. 4.14).

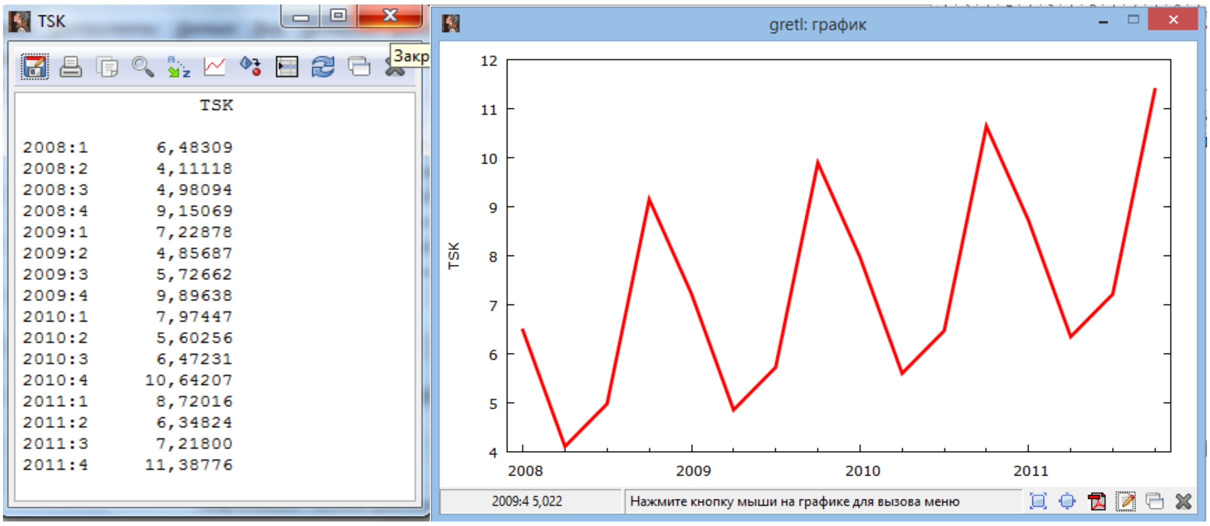


Рис. 4.14. Значения и график трендовой и сезонной компонент

15. Определяем прогнозную оценку объема продаж на 1, 2, 3, 4 кварталы следующего года. Изменим диапазон *time* для прогноза: Данные/Изменить значения/Добавить наблюдения (рис. 4.15).

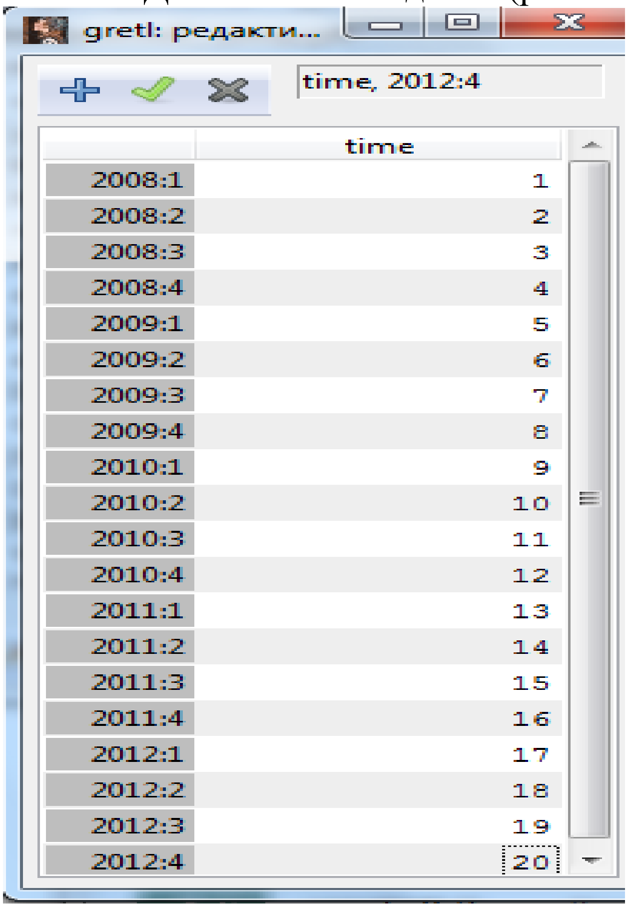


Рис. 4.15. Окно добавления наблюдений в переменную *time*

16. В окне «Просмотр сессии» откроем модель линейного тренда. В окне модели выбираем: Анализ / Прогнозы, и задаем горизонт прогнозирования (рис. 4.16, рис. 4.17).

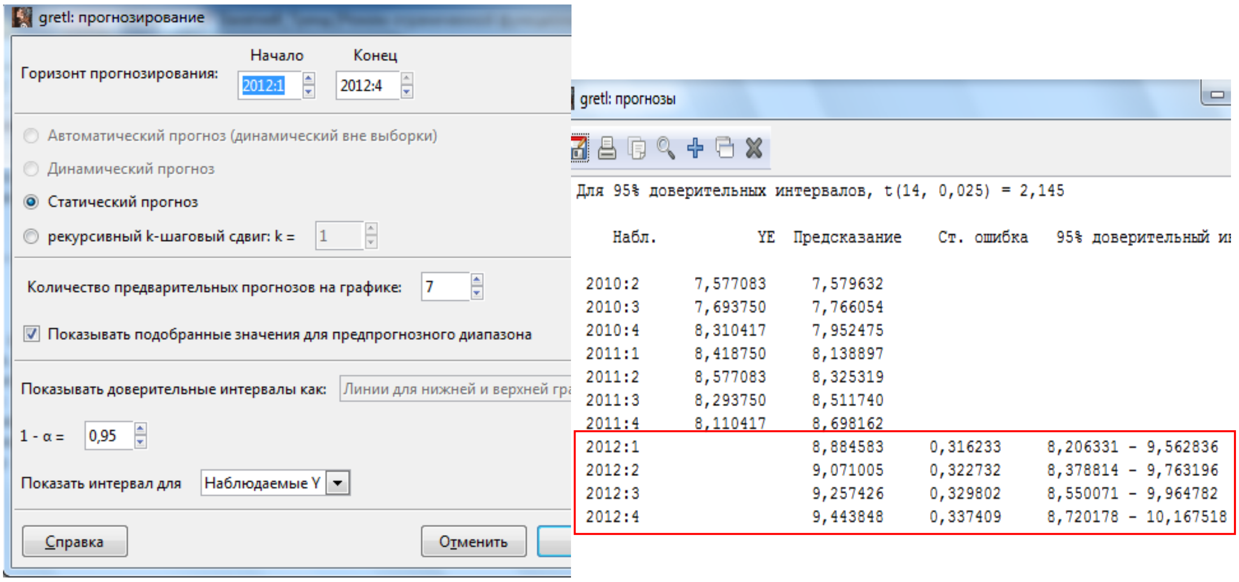


Рис. 4.16. Прогноз выручки в 1, 2, 3, 4 кварталах следующего года

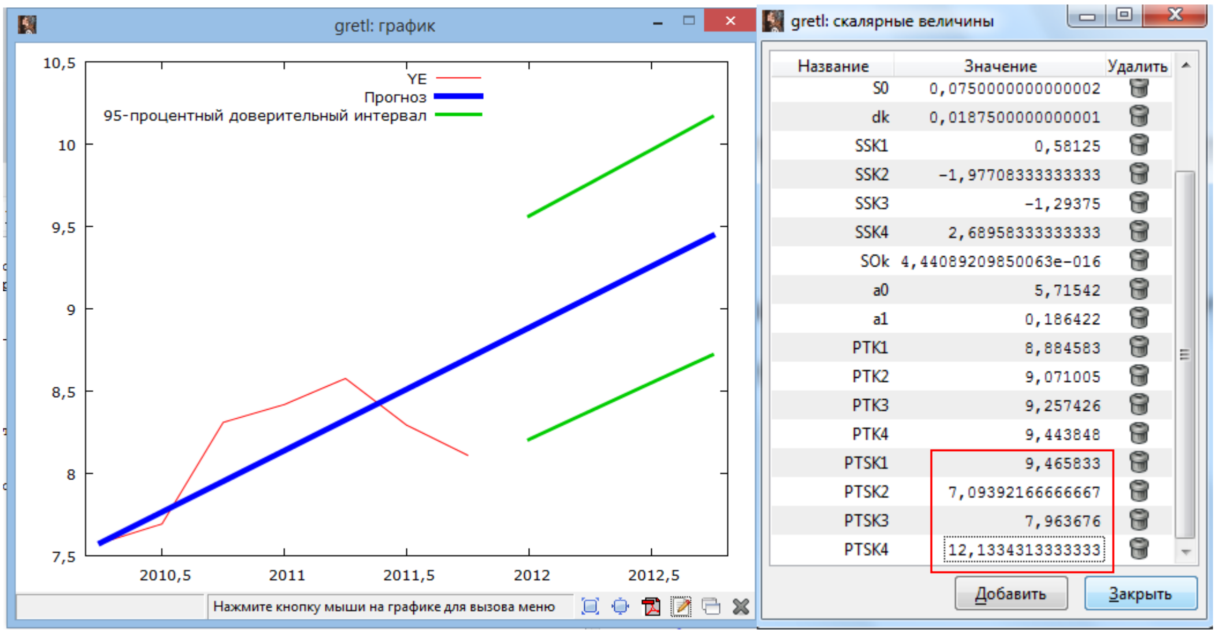


Рис. 4.17. График прогноза трендовой компоненты и суммы трендовой и сезонной компонент

17. К каждому прогнозному значению тренда прибавим величину сезонной компоненты. Введем скаляры: $PTK1$, $PTK2$, $PTK3$, $PTK4$. Затем введем скаляры: $PTSK1$, $PTSK2$, $PTSK3$, $PTSK4$: $PTSK1 = PTK1 + SSK1$ и т.д.

Задания для самостоятельного выполнения

Задание 4.2. Исследуется регрессионная модель влияния численности населения (*Xt*) на ВВП (*Yt*) на основе временных рядов

$$LnY_t = \beta_0 + \beta_1 lnX_t + e.$$

Однако временные ряды для ВВП и численности населения, как правило, имеют тенденцию к росту. Как это повлияет на статистические выводы для регрессионной модели?

Задание 4.3. Н основе данных об объеме продаж фирмы (таблица 4.2) проанализировать имеющиеся данные, выбрать модель для расчета прогноза, обосновать выбор модели, построить прогноз объема продаж на декабрь.

Таблица 4.2

Месяц	1	2	3	4	5	6	7	8	9	10
Объем продаж	95	63	56	54	60	49	46	49	62	56

Задание 4.4. Имеются поквартальные данные об объеме продаж товара за три года (таблица 4.3). Используя мультипликативную модель построить прогноз объема продаж на 1, 2 и 3 кварталы следующего года. Обосновать выбор модели.

Таблица 4.3

Период времени	1 год				2 год				3 год			
	1	2	3	4	1	2	3	4	1	2	3	4
Объем продаж	3	5,5	8	6	5	9,5	13	11	9	15	22	20

Контрольные вопросы:

1. Какие свойства имеет коэффициент автокорреляции?
2. Что такое коррелограмма? Что выявляют при помощи анализа коррелограммы?
3. Что такое аналитическое выравнивание временного ряда?
4. Какие этапы содержит процедура построения тренд-сезонных моделей временных рядов?
5. В чем отличие аддитивной и мультипликативной моделей временных рядов?
6. Чему равна сумма сезонных компонент в аддитивной модели временного ряда?
7. Как осуществляется прогнозирование на основе трендовой и тренд-сезонной моделей временных рядов?

5. Модели ARMA и ARIMA

Занятие проводится в форме кресельного кейс-метода. Условие задачи содержит реальный практический смысл. Поэтому преподаватель предлагает выполнить комплекс эконометрических расчетов, идентифицировать и верифицировать результаты с целью их экономической интерпретации.

На первом этапе преподаватель объявляет группы из двух-трех магистрантов для совместного выполнения практического задания. Роль преподавателя на данном этапе заключается в постоянном поддержании активного внутригруппового взаимодействия, снятии напряженности во взаимоотношениях между участниками, оперативном вмешательстве в случае возникновения непредвиденных трудностей, а также в целях пояснения новых положений учебной программы.

На втором этапе преподаватель назначает лидера для руководства ходом обсуждения результатов выполнения задания. Лидер, применяя уникальное сочетание компьютерных и традиционных методов организации учебной деятельности демонстрирует основные формулы, используемые в решении задания. Совместно с преподавателем лидер руководит групповым обсуждением области применения формул, полученных эконометрических оценок и их качества, экономической интерпретацией результатов. Преподаватель побуждает магистрантов к обоснованию в аудитории полученных результатов, конструктивной критике и поддержке выводов одногруппников с целью формирования навыков изучения реальных социально-экономических явлений и процессов, проблемных ситуаций, высказывания и отстаивания собственной точки зрения, концентрируя внимание на следующих вопросах:

1. Стационарные и нестационарные дискретные случайные процессы.
2. Тестирование временного ряда на стационарность.
3. Информационные критерии для качества подгонки моделей.

В конце занятия преподаватель подводит итоги и оценивает каждого студента в зависимости от его участия в выполнении заданий и обсуждении.

Расчетные формулы

Модель скользящей средней в качестве объясняющих переменных содержит комбинацию белых шумов, то есть ряд Y_t описывается процессом $MA\left(\tau\right)$:

$$Y_t = \mu + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + ... + \alpha_q u_{t-q} = \mu + u_t + \sum_{i=1}^{q} \alpha_i u_{t-i} \ ,$$

где u_t – «белый шум».

Авторегрессионная модель порядка τ – $AR\left(\tau\right)$ имеет следующий вид:

$$Y_t = \mu + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-\tau} + u_t = \mu + \sum_{i=1}^{\tau} \beta_i y_{t-1} + u_t ,$$

где u_t – «белый шум».

Процесс $MA\left(\tau\right)$ всегда стационарен. Процесс $AR\left(\tau\right)$ либо стационарен и может быть представлен в виде скользящего среднего, либо не стационарен.

Модель $ARMA\left(p,q\right)$ выглядит следующим образом:

$$Y_t = \mu + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + ... + \alpha_q u_{t-q} .$$

Выбор параметров p и q модели ARMA определяется как этап идентификации процесса путем сравнения функции автокорреляции ACF и частной автокорреляции PACF. Для любого стационарного авторегрессионного процесса автокорреляционная функция (ACF) будет уменьшаться по экспоненте. Частная автокорреляционная функция τ_{kk} (PACF) определяет корреляцию между текущим и произошедшим k периодов назад наблюдением после удаления косвенного влияния других наблюдений и определяет порядок авторегрессионного процесса. Важнейшие свойства этих функций представлены в таблице 5.1.

Таблица 5.1

Свойства теоретических функций ACF и PACF процесса ARMA

Процесс	Функция автокорреляции ACF	Функция частной автокорреляции PACF
$AR\left(p\right)$	Бесконечная, убывающая (убывающие показательные функции и/или синусоиды с возмущениями)	Конечная, обрывается после p периодов (убывающие показательные функции и/или синусоиды с возмущениями)
$MA\left(q\right)$	Конечная, обрывается после q периодов	Бесконечная, убывающая (убывающие показательные функции и/или синусоиды с возмущениями)
$ARMA\left(p,q\right)$	Бесконечная, убывающая (после первых p-q периодов убывающие показательные функции и/или синусоиды с возмущениями)	Бесконечная, убывающая (после первых p-q периодов убывающие показательные функции и/или синусоиды с возмущениями)

Кроме анализа графиков ACF и PACF для идентификации модели ARMA можно использовать информационные критерии Акайке, Шварца, Ханнан-Куина. Цель – получить модель с минимальными значениями информационных критериев.

Критерий Акайке:

$$AIC = \ln\left(\hat{\sigma}^2\right) + \frac{2k}{T} \text{ или } AIC = \ln s^2(m) + \frac{2m}{n} \to \min.$$

Критерий Шварца:

$$SIC = \ln\left(\hat{\sigma}^2\right) + \frac{k}{T}\ln T \text{ или } SIC = \ln s^2(m) + \frac{m \ln n}{n} \to \min.$$

Критерий Хеннана-Куина:

$$HQIC = \ln\left(\hat{\sigma}^2\right) + \frac{2k}{T}\ln\left[\ln(T)\right], \text{ или}$$

$$HQ(m) = \ln s^2(m) + 2c\frac{m \ln\ln n}{n} \to \min,\ c > 1,$$

где $\hat{\sigma}^2$ – остаточная дисперсия; $k = p + q + 1$ – число оцениваемых параметров; T – размер выборки. Этот критерий недооценивает порядок модели при небольших объемах выборки.

Проверка адекватности выбранной модели основана на исследовании остатков на стационарность – наличие «белого шума», с помощью Q-статистики Бокса-Льюинга:

$$Q = n(n+2)\sum_{h=1}^{M}\frac{r_e^2(h)}{n-h}.$$

Нулевая гипотеза о независимости и одинаковой распределенности остатков (адекватности модели) отвергается, если $Q > \chi^2 кр\ (\alpha, M-p)$.

Для описания нестационарных однородных временных рядов применяется модель Бокса-Дженкинса (ARIMA-модель): (Auto Regressive Integrated Moving Average ($ARIMA(p,s,q)$). Символ I (Integrated) отвечает за порядок оператора последовательной разности.

Для измерения качества подгонки модели Бокса-Дженкинса можно использовать следующие информационные критерии:

1. Критерий Акайка (Akaike information criterion, AIC):

$$AIC = \frac{k+m}{n} + \ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right).$$

2. Критерий Шварца (Swarz criterion):

$$SIK = \frac{(p+q)\ln n}{n} + \ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right).$$

Выбор следует сделать в пользу модели с меньшим значением *AIC*, *SIK*.

Задания для выполнения в аудитории

Задание 5.1. Используя временной ряд цен на нефть, долл./бар. (таблица 5.2), оцените ARMA-модель и ARIMA-модель. На основе полученных моделей постройте динамический прогноз цен на нефть на три квартала вперед.

Таблица 5.2

Date	Oil Price	Date	Oil Price	Date	Oil Price	Date	Oil Price
1998-Jan	14,41	2001-Feb	26,03	2004-Mar	31,11	2007-Apr	63,49
1998-Feb	12,97	2001-Mar	22,01	2004-Apr	29,68	2007-May	63,6
1998-Mar	12,22	2001-Apr	24,27	2004-May	35,14	2007-Jun	67,18
1998-Apr	12,52	2001-May	26,8	2004-Jun	32,68	2007-Jul	73,76
1998-May	13,01	2001-Jun	25,83	2004-Jul	34,75	2007-Aug	69,18
1998-Jun	10,2	2001-Jul	23,17	2004-Aug	39,99	2007-Sep	73,7
1998-Jul	11,73	2001-Aug	24,66	2004-Sep	37,31	2007-Oct	78,32
1998-Aug	11,55	2001-Sep	24,43	2004-Oct	41,92	2007-Nov	89,31
1998-Sep	13,05	2001-Oct	19,93	2004-Nov	37,02	2007-Dec	87,28
1998-Oct	11,44	2001-Nov	18,24	2004-Dec	34,588	2008-Jan	90,31
1998-Nov	10,46	2001-Dec	18,11	2005-Jan	39,09	2008-Feb	91,44
1998-Dec	9,04	2002-Jan	18,58	2005-Feb	40,11	2008-Mar	100,15
1999-Jan	10,85	2002-Feb	18,73	2005-Mar	47,62	2008-Apr	103,51
1999-Feb	9,4	2002-Mar	22,26	2005-Apr	46,97	2008-May	117,43
1999-Mar	11,7	2002-Apr	23,64	2005-May	44,28	2008-Jun	127,36
1999-Apr	14,98	2002-May	23,83	2005-Jun	49,3	2008-Jul	132,05
1999-May	13,78	2002-Jun	23,06	2005-Jul	52,24	2008-Aug	112,87
1999-Jun	14,91	2002-Jul	24,82	2005-Aug	59,28	2008-Sep	97,5
1999-Jul	18,27	2002-Aug	25,86	2005-Sep	58,49	2008-Oct	73,82
1999-Aug	20,02	2002-Sep	27,09	2005-Oct	55,01	2008-Nov	51,2
1999-Sep	21,92	2002-Oct	27,02	2005-Nov	51,69	2008-Dec	39,55
1999-Oct	21,12	2002-Nov	24,39	2005-Dec	52,93	2009-Jan	41,14
1999-Nov	24,05	2002-Dec	27,54	2006-Jan	57,57	2009-Feb	42,56
1999-Dec	25,02	2003-Jan	30,23	2006-Feb	54,98	2009-Mar	44,76
2000-Jan	25,32	2003-Feb	31,76	2006-Mar	57,17	2009-Apr	48,33
2000-Feb	27,43	2003-Mar	27,61	2006-Apr	65,11	2009-May	54,79
2000-Mar	26,47	2003-Apr	22,56	2006-May	64,64	2009-Jun	68,32
2000-Apr	21,62	2003-May	23,65	2006-Jun	64,03	2009-Jul	64,5

2000-May	26,86	2003-Jun	25,97	2006-Jul	68,97	2009-Aug	71,7
2000-Jun	29,13	2003-Jul	27,14	2006-Aug	69,96	2009-Sep	67,82
2000-Jul	28,55	2003-Aug	28,33	2006-Sep	59,61	2009-Oct	71,26
2000-Aug	29,5	2003-Sep	25,32	2006-Oct	55,12	2009-Nov	76,08
2000-Sep	30,25	2003-Oct	28,08	2006-Nov	55,45	2009-Dec	73,12
2000-Oct	29,36	2003-Nov	27,1	2006-Dec	57,94	2010-Jan	76,25
2000-Nov	30,04	2003-Dec	28,11	2007-Jan	49,27	2010-Feb	72,58
2000-Dec	23,9	2004-Jan	28,81	2007-Feb	53,22	2010-Mar	76,19
2001-Jan	24,6	2004-Feb	27,57	2007-Mar	58,34	2010-Apr	81,29

Методические указания для выполнения задания

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «Занятие_ARIMA.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1. Распознать данные как месячный временной ряд, построить график и автокорреляционную функцию исходного временного ряда (рис. 5.1).

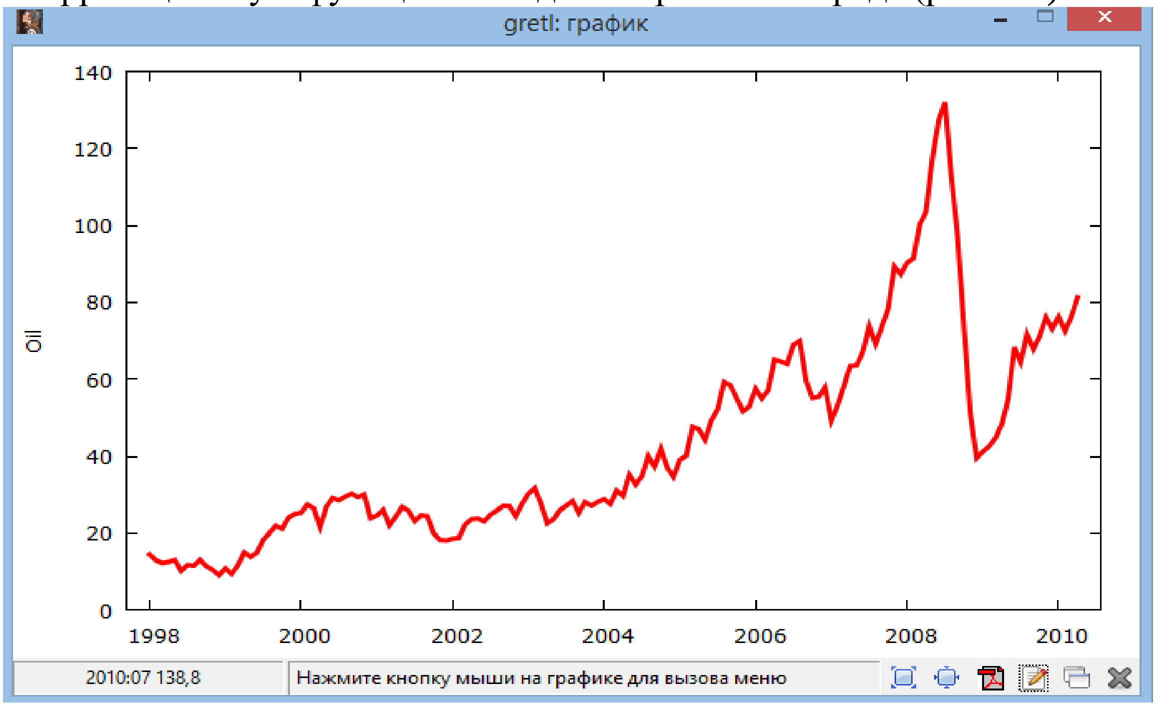
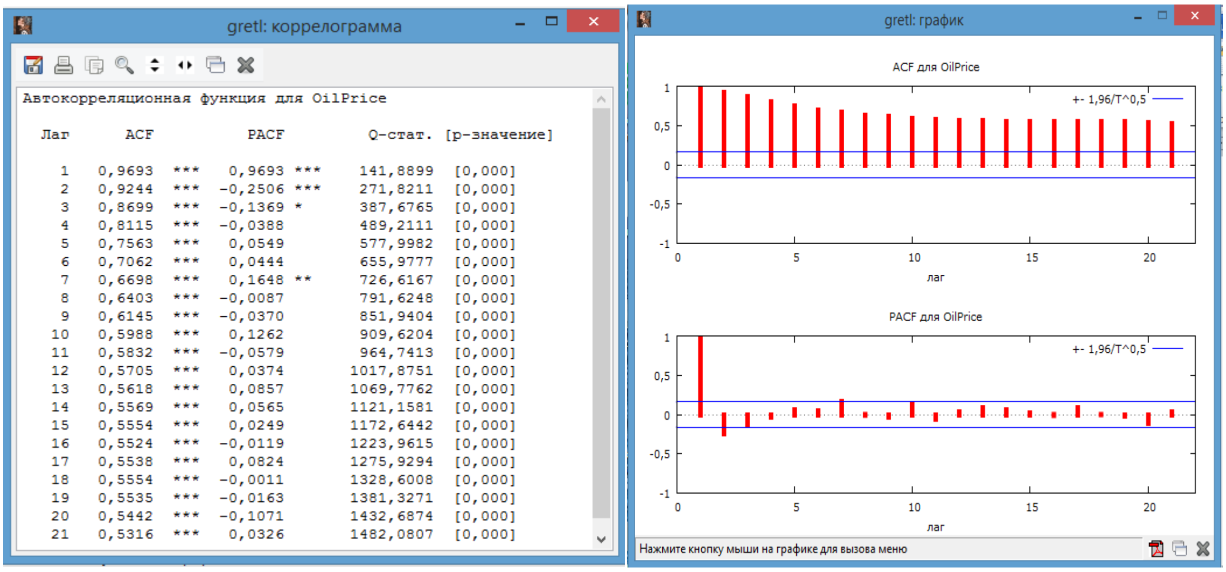


Рис. 5.1. График исходного временного ряда



Автокорреляционная функция для OilPrice

Лаг	ACF		PACF		Q-стат.	[p-значение]
1	0,9693	***	0,9693	***	141,8899	[0,000]
2	0,9244	***	-0,2506	***	271,8211	[0,000]
3	0,8699	***	-0,1369	*	387,6765	[0,000]
4	0,8115	***	-0,0388		489,2111	[0,000]
5	0,7563	***	0,0549		577,9982	[0,000]
6	0,7062	***	0,0444		655,9777	[0,000]
7	0,6698	***	0,1648	**	726,6167	[0,000]
8	0,6403	***	-0,0087		791,6248	[0,000]
9	0,6145	***	-0,0370		851,9404	[0,000]
10	0,5988	***	0,1262		909,6204	[0,000]
11	0,5832	***	-0,0579		964,7413	[0,000]
12	0,5705	***	0,0374		1017,8751	[0,000]
13	0,5618	***	0,0857		1069,7762	[0,000]
14	0,5569	***	0,0565		1121,1581	[0,000]
15	0,5554	***	0,0249		1172,6442	[0,000]
16	0,5524	***	-0,0119		1223,9615	[0,000]
17	0,5538	***	0,0824		1275,9294	[0,000]
18	0,5554	***	-0,0011		1328,6008	[0,000]
19	0,5535	***	-0,0163		1381,3271	[0,000]
20	0,5442	***	-0,1071		1432,6874	[0,000]
21	0,5316	***	0,0326		1482,0807	[0,000]

Окончание рис. 5.1. Автокорреляционная функция исходного временного ряда

Визульный анализ графика временного ряда и коррелограмм ACF и PACF обнаруживает процесс случайного блуждания. ACF монотонно убывает, PACF – конечная, с обрывом после 1 периода.

3. Построение ARIMA модели: Модель / Временной ряд / ARIMA / Разность=1. Устанавливаем требуемый порядок: Порядок *AR*, Порядок *MA*: p=1,s=1,q=0 (рис. 5.2).

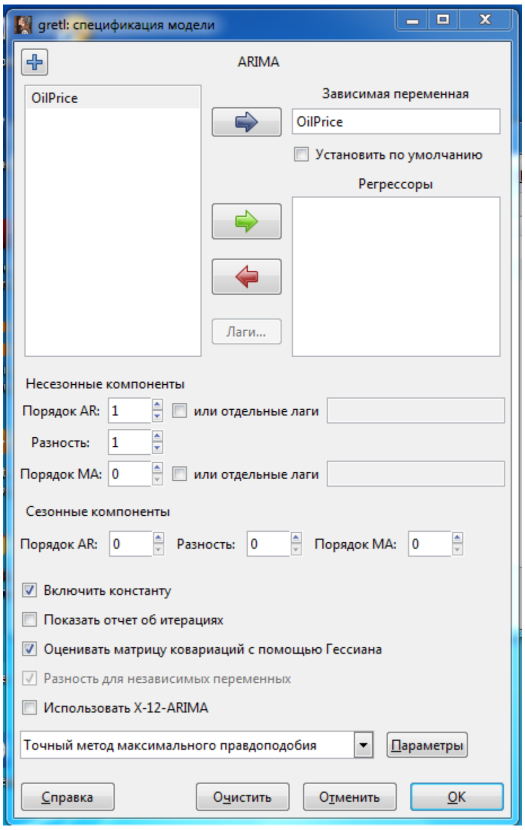
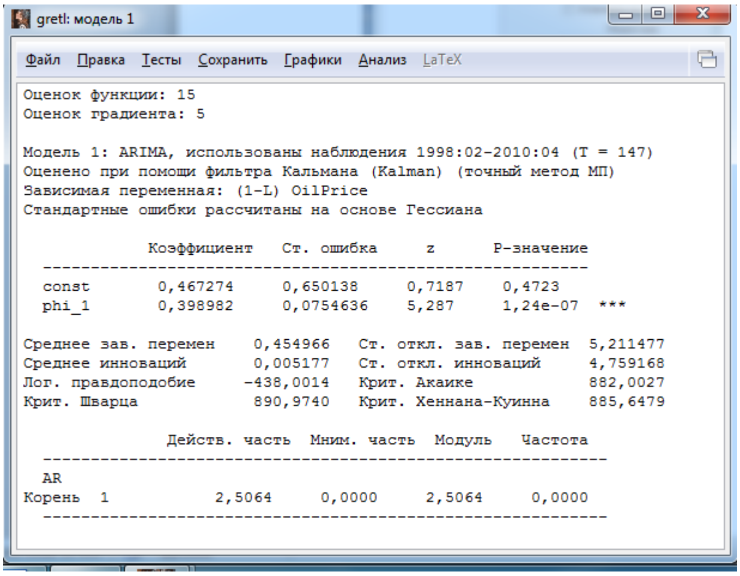


Рис. 5.2. ARMA-модель (p=1, s=1,q=0)



Окончание рис. 5.2. ARMA-модель (p=1, s=1,q=0,)

Для сравнения построим модель ARMA (p=0, q=0) (рис 5.3)

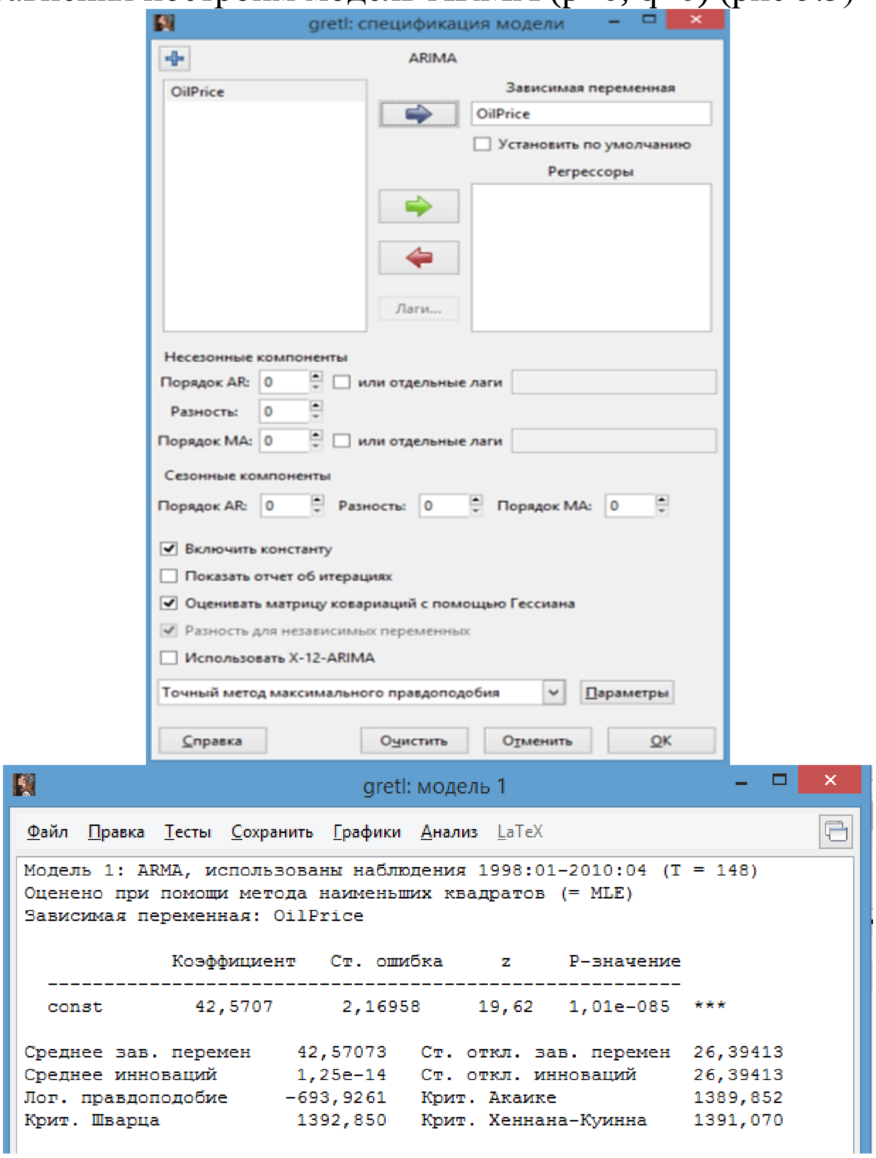


Рис. 5.3. ARMA-модель ($p=0,\ q=0$)

Результаты оформляем в виде таблицы 5.3:

Таблица 5.3

Сводная таблица ARMA-моделей

p	*s*	*q*	Критерий Акаике	Критерий Шварца
0	0	0	1389	1391
1	1	0	882	890

Выбираем лучшую модель по минимуму информационных критериев Акаике, Шварца.

3. Построение модели авторегрессии при лагах 1, 2, 7. Поставить флажок в строку «Отдельные лаги» и прописать их через пробел (рис. 5.4).

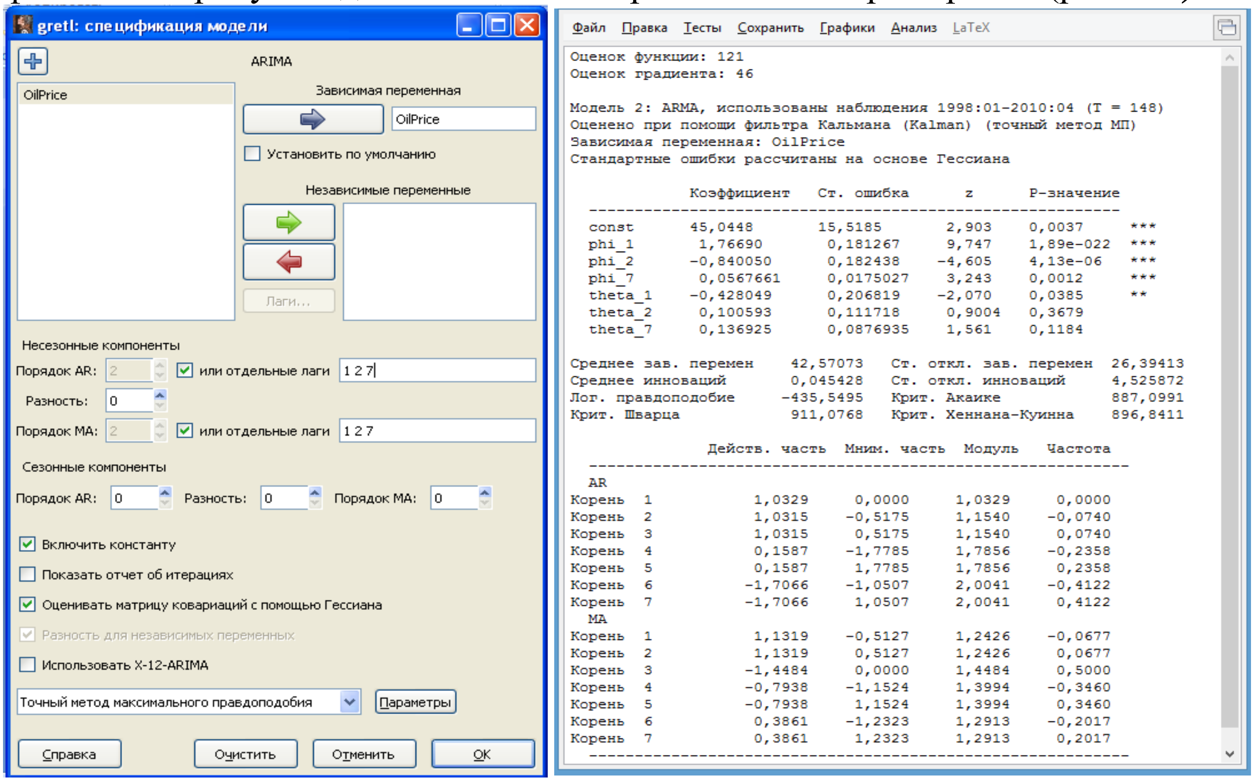


Рис. 5.4. ARMA-модель при лагах 1, 2, 7

Если точный метод максимального правдоподобия не дает результатов (не может найти гессиан), то переходим к условному методу.

4. Строим прогноз на три периода (квартала) вперед. Это динамический прогноз, поскольку в нем участвуют лаговые переменные. В окне Модели: Анализ / Прогнозы / Добавить 3 наблюдения (рис. 5.5).

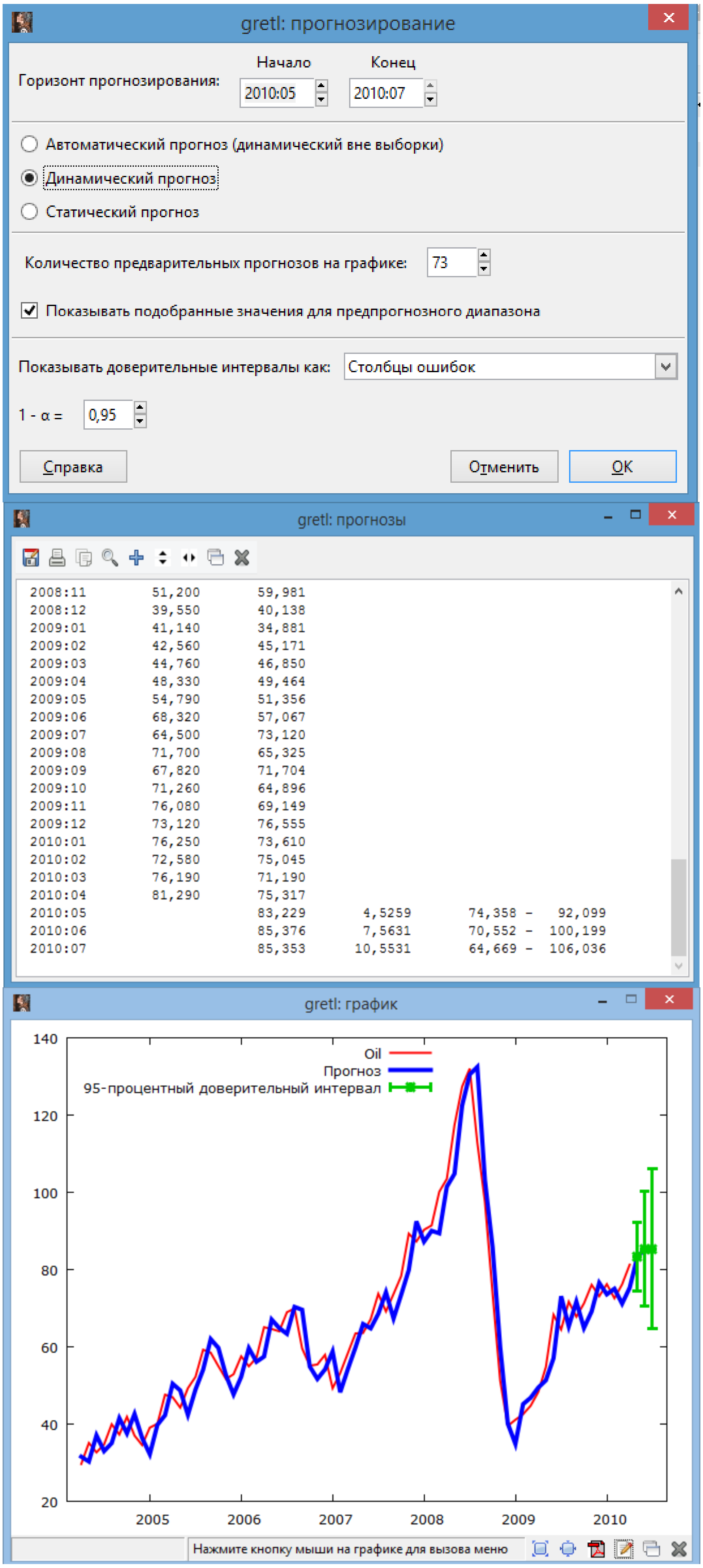


Рис. 5.5. Прогноз цены нефти и его доверительный интервал

Задание 6.2. Используя временной ряд цен на золото, долл./унц. (таблица 5.4), оцените ARIMA-модель. На основе полученной модели постройте динамический прогноз цен на золото на четыре недели вперед.

Таблица 5.4

t	*Y*	*t*	*Y*	*t*	*Y*	*t*	*Y*
26.12.2003	77,8	02.04.2004	83,6	09.07.2004	89,2	15.10.2004	100,1
02.01.2004	77,1	09.04.2004	83,5	16.07.2004	89,2	22.10.2004	105,2
09.01.2004	78,9	16.04.2004	83,2	23.07.2004	88,9	29.10.2004	107,3
16.01.2004	79,1	23.04.2004	82,8	30.07.2004	88,7	05.11.2004	112,8
23.01.2004	82,7	30.04.2004	82,7	06.08.2004	89	12.11.2004	113,1
30.01.2004	84,1	07.05.2004	83,4	13.08.2004	89,6	19.11.2004	113,9
06.02.2004	84,3	14.05.2004	82,7	20.08.2004	88,3	26.11.2004	117,1
13.02.2004	88	21.05.2004	83,2	27.08.2004	88,8	03.12.2004	121,6
20.02.2004	86,7	28.05.2004	85,4	03.09.2004	89,1	10.12.2004	120,3
27.02.2004	86,4	04.06.2004	85,6	10.09.2004	90	17.12.2004	119,8
05.03.2004	84,6	11.06.2004	86,2	17.09.2004	92,6	24.12.2004	120,7
12.03.2004	84,6	18.06.2004	87,4	24.09.2004	94,3	31.12.2004	124,5
19.03.2004	84,8	25.06.2004	87,9	01.10.2004	95,3	07.01.2005	124,6
26.03.2004	83,7	02.07.2004	88,3	08.10.2004	98,3		

Методические указания для выполнения задания

1. Импорт данных из таблицы Excel: Файл/Открыть/Импорт/Excel/«Занятие_ARMA.xls»/gold. Распознать данные как еженедельный временной ряд. Начало: пятница, 26.12.2003 (рис. 5.5).

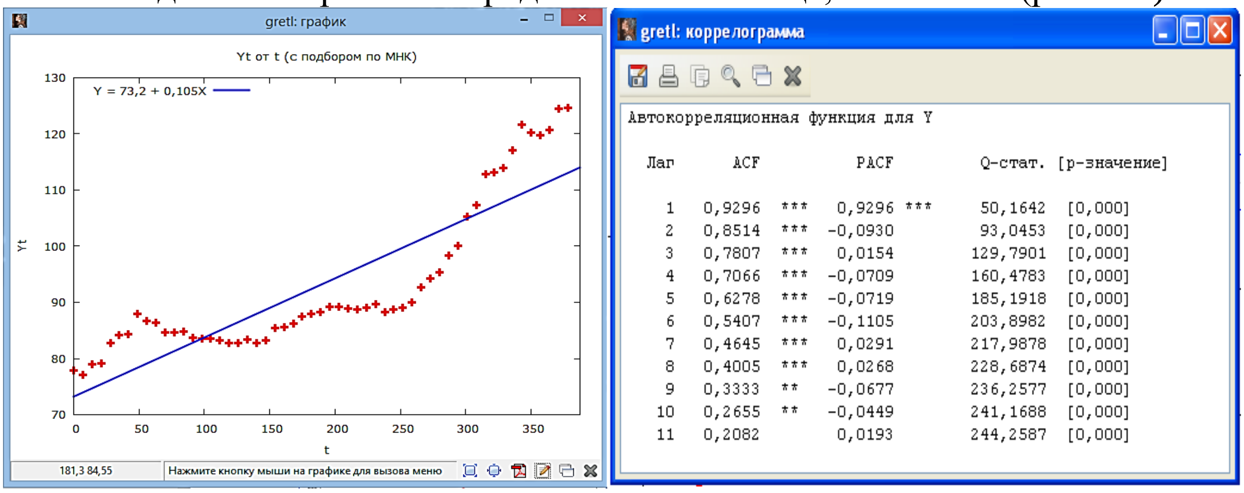


Рис. 5.5. График и автокорреляционная функция исходного временного ряда

Визуальный анализ графика временного ряда и коррелограммы отчетливо обнаруживает тренд, на который накладываются случайные колебания. Имеется ACF с 10 значимыми коэффициентами. Медленное спадание ACF связано с наличием тренда. PACF имеет один значимый коэффициент. Приемлемы: $ARIMA\ (0,1,1)$, $ARIMA\ (1,1,0)$.

2. Построим модели $ARIMA\ (0,1,1)$, $ARIMA\ (1,1,0)$: Модель/Временной ряд/ARIMA (рис. 5.6, рис. 5.7).

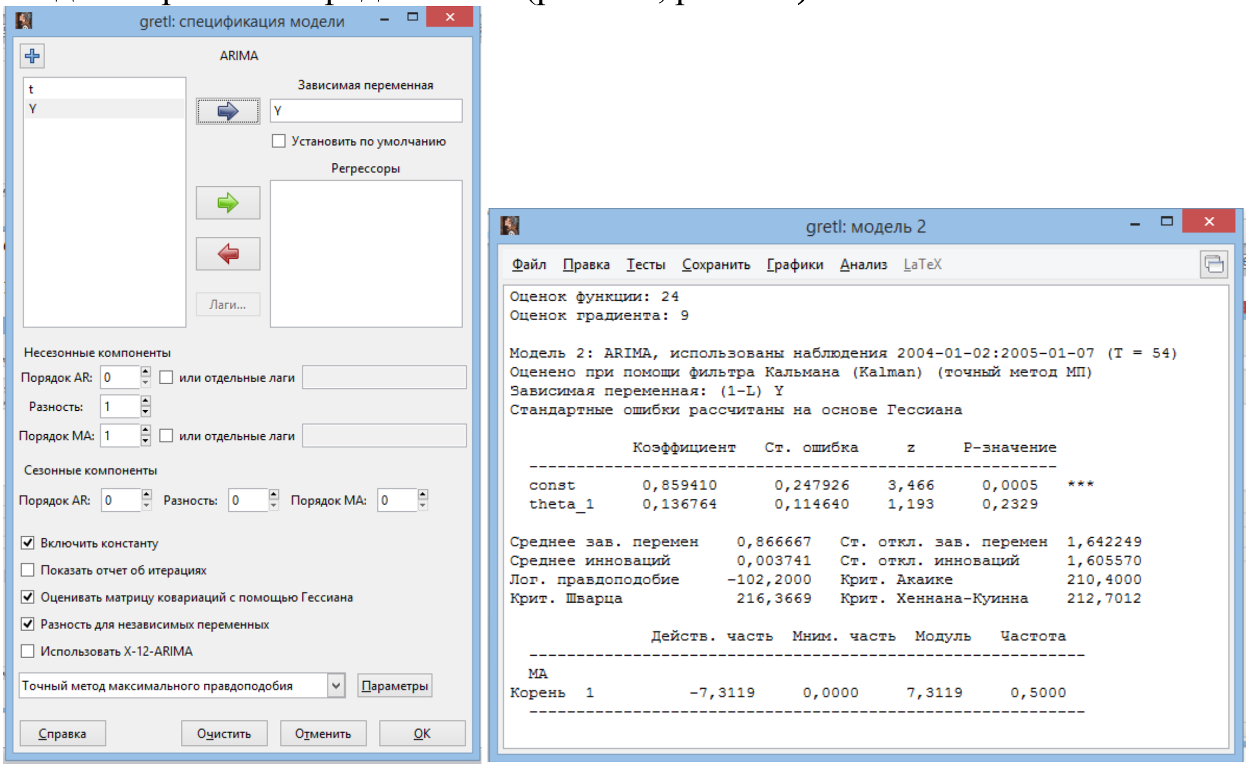


Рис. 5.6. ARIMA-модель для $p=1,\ s=1,\ q=0$

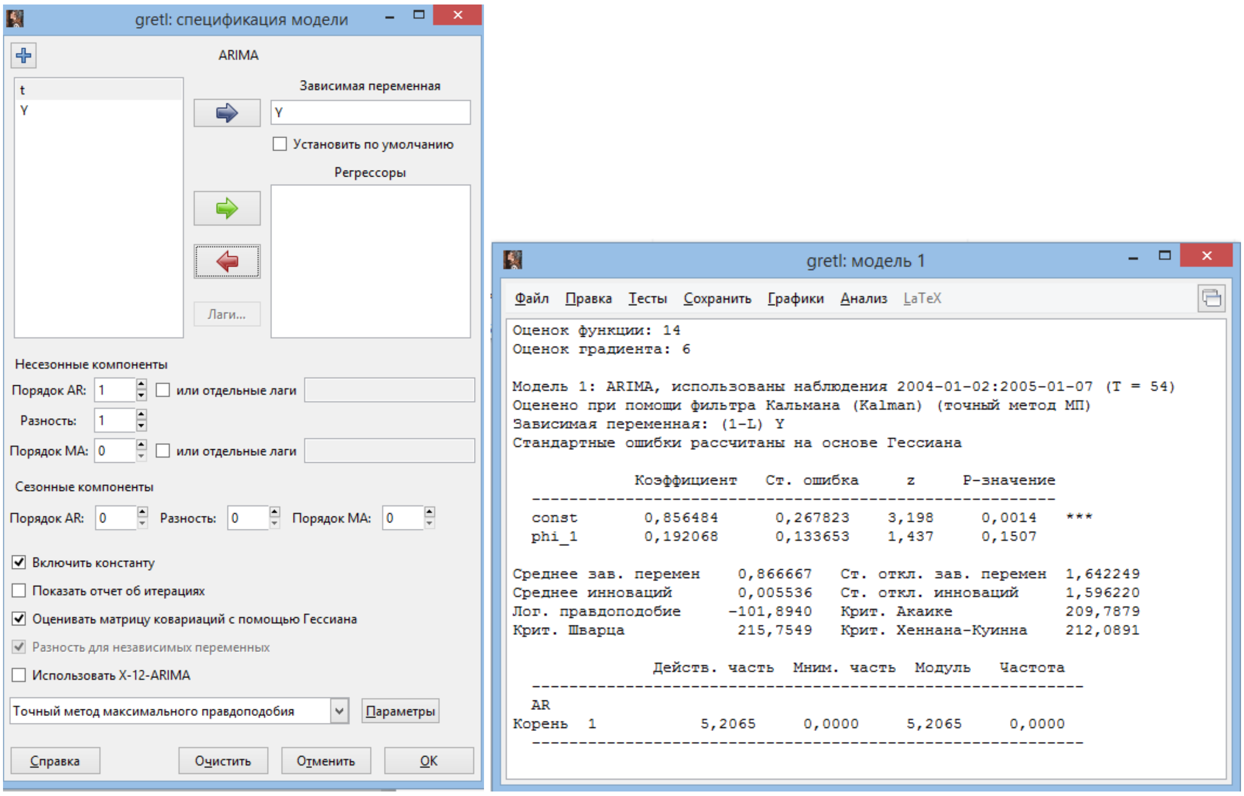
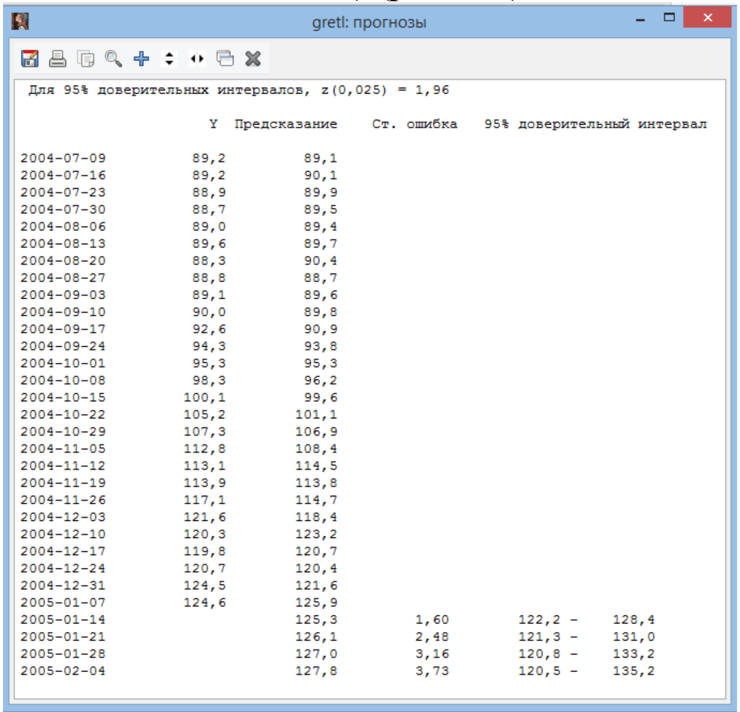


Рис. 5.7. ARIMA-модель для $p=0,\ s=1,\ q=1$

Построим прогноз на 4 недели вперед. Предварительно добавим 4 наблюдения. В окне Модели: Анализ/Прогнозы/Динамический прогноз (начало: 2005.01.14, конец: 2005.02.04) (рис. 5.8).



gretl: прогнозы

Для 95% доверительных интервалов, z(0,025) = 1,96

	Y	Предсказание	Ст. ошибка	95% доверительный интервал
2004-07-09	89,2	89,1		
2004-07-16	89,2	90,1		
2004-07-23	88,9	89,9		
2004-07-30	88,7	89,5		
2004-08-06	89,0	89,4		
2004-08-13	89,6	89,7		
2004-08-20	88,3	90,4		
2004-08-27	88,8	88,7		
2004-09-03	89,1	89,6		
2004-09-10	90,0	89,8		
2004-09-17	92,6	90,9		
2004-09-24	94,3	93,8		
2004-10-01	95,3	95,3		
2004-10-08	98,3	96,2		
2004-10-15	100,1	99,6		
2004-10-22	105,2	101,1		
2004-10-29	107,3	106,9		
2004-11-05	112,8	108,4		
2004-11-12	113,1	114,5		
2004-11-19	113,9	113,8		
2004-11-26	117,1	114,7		
2004-12-03	121,6	118,4		
2004-12-10	120,3	123,2		
2004-12-17	119,8	120,7		
2004-12-24	120,7	120,4		
2004-12-31	124,5	121,6		
2005-01-07	124,6	125,9		
2005-01-14		125,3	1,60	122,2 - 128,4
2005-01-21		126,1	2,48	121,3 - 131,0
2005-01-28		127,0	3,16	120,8 - 133,2
2005-02-04		127,8	3,73	120,5 - 135,2

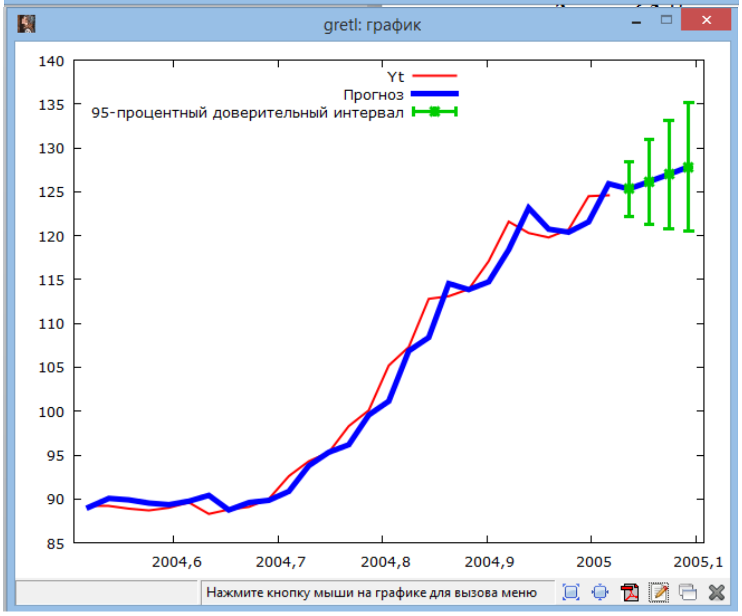


Рис. 5.8. Прогноз цены золота и его доверительный интервал

Задания для самостоятельного выполнения

Задание 5.3. По временному ряду длины $n = 100$ были оценены авторегрессионные модели до четвертого порядка и для них получены следующие оценки дисперсий остатков: $s^2(1) = 0,9, s^2(2) = 0,7, s^2(3) = 0,5, s^2(4) = 0,46$. Выберите порядок модели авторегрессии с помощью информационных критериев Шварца и Акаике.

Задание 5.4. Для временного ряда длины $n = 100$ была оценена модель второго порядка ($p = 2$) и вычислены коэффициенты автокорреляции остатков,
$r_e(1) = 0{,}001;\ r_e(2) = 0{,}001;\ r_e(3) = 0{,}0006;\ r_e(4) = 0{,}0004;\ r_e(5) = 0{,}0003.$
Проверьте адекватность модели по критерию Бокса-Льюинга.

Задание 5.5. По временному ряду длины $n = 60$ были оценены следующие авторегрессионные модели:

1) $X_t = 2 + 0{,}7X_{t-1};\ s^2 = 2{,}1;$

2) $X_t = 2{,}3 + 0{,}6X_{t-1} - 0{,}3X_{t-2};\ s^2 = 1{,}9;$

3) $X_t = 1{,}8 + 0{,}55X_{t-1} - 0{,}25X_{t-2} + 0{,}01X_{t-3};\ s^2 = 1{,}85.$

Какую модель выбрать для прогнозирования?

Задание 5.6. Для временного ряда (таблица 5.5) оценить модели $AR(1)$, $AR(2)$, $AR(3)$. Используя информационные критерии, выбрать одну из этих моделей и проверить ее адекватность.

Таблица 5.5

t	*Yt*	*t*	*Yt*	*t*	*Yt*	*t*	*Yt*	*t*	*Yt*
1	5,84	9	7,58	17	3,61	25	3,93	33	7,08
2	5,92	10	6,47	18	5,02	26	2,96	34	4,94
3	4,64	11	5,42	19	5,79	27	2,98	35	2,90
4	4,77	12	4,70	20	5,65	28	1,70		
5	4,64	13	5,08	21	5,91	29	3,26		
6	6,28	14	4,23	22	6,56	30	6,56		
7	6,71	15	4,03	23	6,31	31	8,26		
8	7,47	16	2,17	24	5,99	32	9,21		

Контрольные вопросы:

1. Как определяется модель ARMA?
2. Как интерпретируют параметры моделей авторегрессии?
3. Что означает стационарность временного ряда?
4. Какой стационарный процесс называется «белым шумом»?
5. Какие типы включают модели стационарных временных рядов?
6. Какие типы включают модели нестационарных временных рядов?
7. Как определяется ARIMA-модель?

4. Многомерные динамические модели временных рядов с распределенным лагом

Расчетные формулы

Модели с распределенными лагами содержат в качестве лаговых переменных лишь независимые переменные:

$$y = a + b_0 x_t + b_1 x_{t-1} + ... + b_{p-l} + \varepsilon_t$$

Интерпретация: коэффициент b_0 называется краткосрочным мультипликатором, т.к. он характеризует изменение среднего значения y при единичном изменении x в тот же самый момент времени. $\sum_{j=0}^{p} b_j$ - долгосрочный мультипликатор, он характеризует изменение y под воздействием единичного изменения x в каждом из моментов времени. Любая сумма $\sum_{j=0}^{k} b_j$ $(k < l)$ называется промежуточным мультипликатором.

Относительные коэффициенты модели с распределенным лагом определяются выражениями: $\beta_j = \frac{b_j}{\sum_{j=0}^{p} b_j}$; $\sum_{j=0}^{p} \beta_j = 1$. Средний лаг определяется по формуле средней арифметической взвешенной: $\bar{l} = \sum_{j=0}^{p} j \cdot \beta_j$.

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

Метод Алмон. Предполагается, что в модели с конечной максимальной величиной лага l значения коэффициентов b_j описываются полиномом к – й степени: $b_j = c_0 + c_1 j + c_2 j^2 + ... + c_k j^k$.

1.Определяется максимальная величина лага l и степень полинома k.

2.Рассчитываются значения вспомогательных переменных $z_0, z_1, ..., z_k$:

$$z_0 = x_t + x_{t-1} + x_{t-2} + ... + x_{t-l};$$
$$z_1 = 0 \cdot x_t + 1 \cdot x_{t-1} + 2 \cdot x_{t-2} + 3 \cdot x_{t-3} + ... + l \cdot x_{t-l};$$
$$z_2 = x_{t-1} + 4 \cdot x_{t-2} + 9 \cdot x_{t-3} + ... + l^2 \cdot x_{t-l};$$
....
$$z_k = x_{t-1} + 2^k \cdot x_{t-2} + 3^k \cdot x_{t-3} + ... + l^k \cdot x_{t-l}.$$

$$y_t = a + c_0 \cdot z_0 + c_1 \cdot z_1 + c_2 \cdot z_2 + ... + c_k \cdot z_k + \varepsilon_t.$$

3. Обычным МНК определяются параметры уравнения регрессии: $y_t = a + c_0 z_0 + c_1 z_1 + c_2 z_2 + \ldots + c_k z_k + \varepsilon_t$

4. С помощью соотношений определяются коэффициенты b_j исходной модели с распределенным лагом:

$$b_0 = c_0;$$
$$b_1 = c_0 + c_1 + \ldots + c_k;$$
$$b_2 = c_0 + 2c_1 + 4c_2 + \ldots + 2^k c_k;$$
$$b_3 = c_0 + 3c_1 + 9c_2 + \ldots + 3^k c_k;$$
....
$$b_j = c_0 + lc_1 + l^2 c_2 + \ldots + l^k c_k.$$

Метод Койка. Этот метод применяется в модели с бесконечным лагом:

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

1. Обычным МНК выполняется оценка параметров модели Койка:

$$y_t = a(1-\lambda) + b_0 x_t + \lambda y_{t-1} + \varepsilon_t - \lambda \varepsilon_{t-1} \quad .$$

2. На основе модели Койка определяются параметры λ, *a* и b_0.

3. С помощью соотношений $b_j = b_0 \cdot \lambda^j$ определяются коэффициенты b_j исходной модели с бесконечным лагом. Данная модель позволяет определить долгосрочный мультипликатор $\sum_{j=0}^{\infty} b_j = b_0 \frac{1}{1-\lambda}$ и средний лаг $\bar{l} = \frac{\lambda}{1-\lambda}$.

Задания для аудиторной работы

Задание 6.1. Компания «Прогресс» с целью увеличения своей прибыли периодически проводит маркетинговые исследования, ориентированные на выявление изменений в предпочтениях потребителей, а также анализ динамики рыночной конъюнктуры. Для этого необходимо проанализировать данные, отражающие зависимость прибыли компании от расходов на маркетинговые исследования (таблица 6.1). Руководство компании заинтересовано в получении ответа на вопрос: какой эффект дает дополнительное вложение в маркетинговые исследования одной тысячи рублей и каков средний лаг, существующий между вложением средств в маркетинговые исследования и получением прибыли от этих вложений.

Таблица 6.1

Период	Прибыль компании, тыс. руб.	Расходы на маркетинговые исследования, тыс. руб.	Период	Прибыль компании, тыс. руб.	Расходы на маркетинговые исследования, тыс. руб.
1	988	60	11	1281	78
2	1035	66	12	1253	67
3	1089	73	13	1302	74
4	1082	67	14	1382	94
5	1073	54	15	1426	93
6	1126	65	16	1468	91
7	1177	75	17	1513	93
8	1234	83	18	1593	96
9	1265	83	19	1612	98
10	1258	74	20	1628	93

Методические указания для выполнения задания

Метод Алмон

Применим метод Алмон. Выберем лаг, равный 4, степень полинома, равную 2. Тогда динамическая модель распределенных лагов имеет вид:

$$Y_t = a_0 + b_0 X_t + b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} + b_4 X_{t-4} + \varepsilon .$$

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «ЗанятиеАлмон.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1.

Распознать как: временной ряд/квартальный/начало 2013:1.

2. Преобразуем исходные переменные X в переменные $Z0$, $Z1$, $Z2$. Предварительно создадим переменные X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}. Для этого необходимо выделить переменную X.

Способ 1: В меню: Добавить/Добавить лаги для выделенных переменных/4.

Способ 2: В меню: Добавить/Добавить новую переменную:

$$X\_1 = X(-1);\ X\_2 = X(-2);\ X\_3 = X(-3);\ X\_4 = X(-4).$$

3. Изменим диапазон переменных X, X_1, X_2, X_3, X_4: в меню: Выборка/Установить диапазон: начало: 2014:1.

4. Создадим переменные $Z0$, $Z1$, $Z2$. В меню: Добавить/Добавить новую переменную:

$$Z0 = X + X\_1 + X\_2 + X\_3 + X\_4,$$

$$Z1 = 0 \cdot X + 1 \cdot X\_1 + 2 \cdot X\_2 + 3 \cdot X\_3 + 4 \cdot X\_4,$$

$$Z2 = 0 \cdot X + 1 \cdot X\_1 + 4 \cdot X\_2 + 9 \cdot X\_3 + 16 \cdot X\_4 .$$

5. Определяем параметры модели распределенных лагов. В меню: Модель/МНК/ (рис. 6.1, рис. 6.2)).

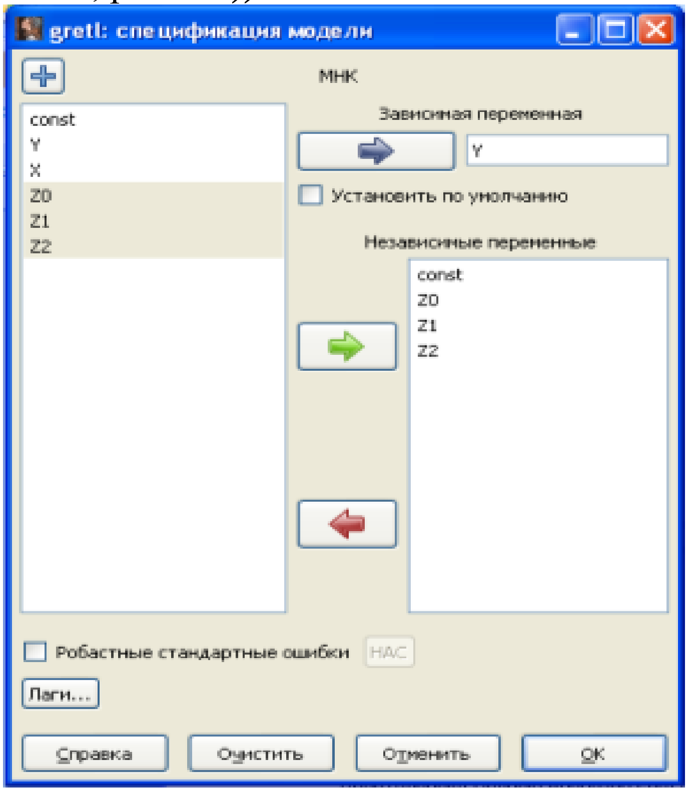


Рис. 6.1. Окно ввода параметров линейной модели с регрессорами *Z*0, *Z*1, *Z*2

```
gretl: модель 1
Файл  Правка  Тесты  Сохранить  Графики  Анализ  LaTeX

Модель 1: МНК, использованы наблюдения 2008:1-2011:4 (T = 16)
Зависимая переменная: Y

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        31,6184      49,3516        0,6407       0,5338
  Z0            5,93328      0,532509     11,14         1,10e-07  ***
  Z1           -3,19784      0,757169     -4,223        0,0012    ***
  Z2            0,631660     0,184014      3,433        0,0050    ***

Среднее зав. перемен    1349,438   Ст. откл. зав. перемен  174,1107
Сумма кв. остатков      7094,468   Ст. ошибка модели       24,31472
R-квадрат               0,984398   Испр. R-квадрат         0,980498
F(3, 12)                252,3789   P-значение (F)          4,20e-11
Лог. правдоподобие     -71,45887   Крит. Акаике            150,9177
Крит. Шварца            154,0081   Крит. Хеннана-Куинна    151,0760
Параметр rho            0,303520   Стат. Дарбина-Вотсона   1,297568

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 10,895
  p-значение = P(Хи-квадрат(9) > 10,895) = 0,282976

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 3,35742
  p-значение = P(Хи-квадрат(3) > 3,35742) = 0,33973

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 4,42957
  p-значение = P(Хи-квадрат(3) > 4,42957) = 0,218659

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 0,673151
  p-значение = 0,714212

LM тест на наличие автокорреляции до порядка 4 -
  Нулевая гипотеза: автокорреляция отсутствует
  Тестовая статистика: LMF = 0,546804
  p-значение = P(F(4,8) > 0,546804) = 0,706923
```

Рис. 6.2. Линейная модель с регрессорами *Z*0, *Z*1, *Z*2

Получим модель: $Y = 31,618 + 5,933Z0 - 3,198Z1 + 0,632Z2 + \varepsilon$.

Параметры модели значимы, гетероскедастичность и автокорреляция в остатках отсутствуют, остатки имеют нормальный закон распределения.

6. Найдем через скаляры параметры исходной модели распределенных лагов: $a = 31,618$; $b0 = c0 = 5,9333$; $b1 = c0 + c1 + c2$; $b2 = c0 + 2c1 + 4c2$; $b3 = c0 + 3c1 + 9c2$; $b4 = c0 + 4c1 + 16c2$ (рис. 6.3).

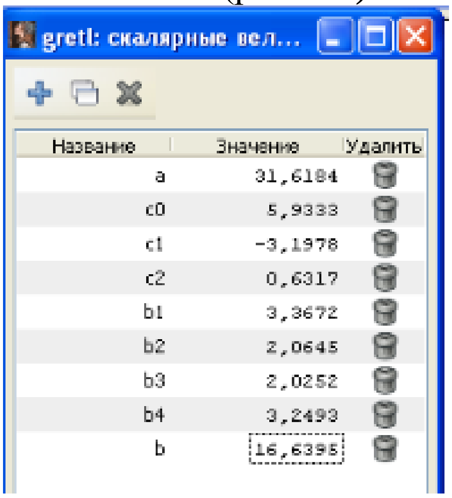


Рис. 6.3. Окно скаляров с параметрами исходной модели распределенных лагов (метод Алмон)

Следовательно, модель с распределенным лагом имеет вид:

$$Y_t = 31,6184 + 5,9333X_t + 3,3672X_{t-1} + 2,0645X_{t-2} + 2,0252X_{t-3} + +3,2493X_{t-4} + \varepsilon.$$

7. Расчет долгосрочного мультипликатора:

$$b = 5,9333 + 3,3672 + 2,0645 + 2,0252 + 3,2493 = 16,6395.$$

Мультипликатор показывает, что на 1 тыс. руб. средств, вложенных в текущем периоде в маркетинговые исследования через 4 квартала (через год) будет получено 16,640 тыс. рублей прибыли.

8. Расчет относительных мультипликаторов и среднего лага (рис. 4.4): $\beta 0 = b0 / b$; $\beta 1 = b1 / b$; $\beta 2 = b2 / b$; $\beta 3 = b3 / b$; $\beta 4 = b4 / b$. Средний лаг $l = 0 \cdot \beta 0 + 1 \cdot \beta 1 + 2 \cdot \beta 2 + 3 \cdot \beta 3 + 4 \cdot \beta 4 = 1,597$.

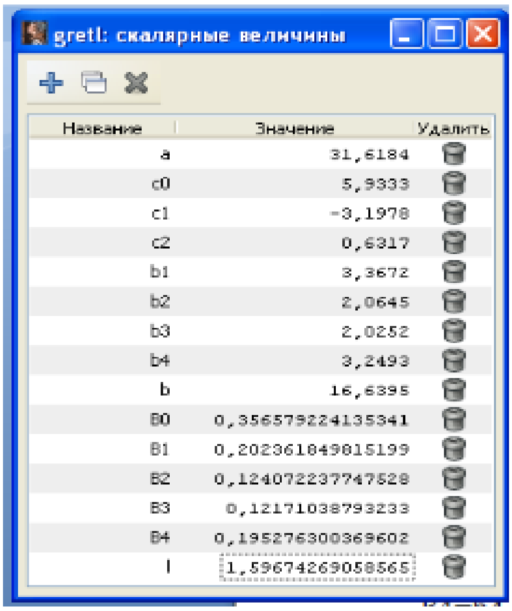


Рис. 6.4. Окно скаляров с относительными мультипликаторами модели

Таким образом, в среднем увеличение расходов на маркетинговые исследования приведет к увеличению прибыли через 1,6 квартала.

Задание 6.2. Администрация региона проводит комплексное исследование социально-экономической ситуации с целью выработки обоснованной политики его развития. Одной из поставленных задач исследования является определение среднего срока воздействия инфляции на реальные доходы населения. Для анализа ситуации сделана выборка наблюдений за 19 лет (таблица 6.2). Специалисты выдвинули гипотезу о том, что годовой уровень инфляции оказывает воздействие на реальные доходы населения с бесконечным временным лагом, который имеет геометрическую структуру.

Таблица 6.2

Год	Уровень инфляции,%	Реальные доходы, млн. руб.	Год	Уровень инфляции,%	Реальные доходы, млн. руб.
1	13,9	1704	11	10,6	2198
2	12,6	1749	12	9,6	2297
3	11,8	1821	13	8,5	2407
4	13,3	1870	14	9,2	2468
5	13,2	1869	15	8,8	2455
6	12,0	1927	16	7,4	2498
7	10,4	2020	17	6,8	2444
8	8,7	2125	18	6,5	2472
9	10,8	2111	19	5,9	2692
10	11,8	2094			

Метод Койка

Применим метод Койка для вычисления параметров исходной модели с бесконечным числом лаговых переменных: $Y_t = a_0 + b_0 X_t + b_1 X_{t-1} + b_2 X_{t-2} + \ldots + \varepsilon.$

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «ЗанятиеКойка.xlsx». Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл/Открыть/Пользовательские/лист 1.

2. Распознать данные как: Временной ряд/Годичные/Начало: 1992.

3. Расчет параметров модели авторегрессии: $Y_t = (1-\lambda)a + b_0 X_t + \lambda Y_{t-1} + v_t$

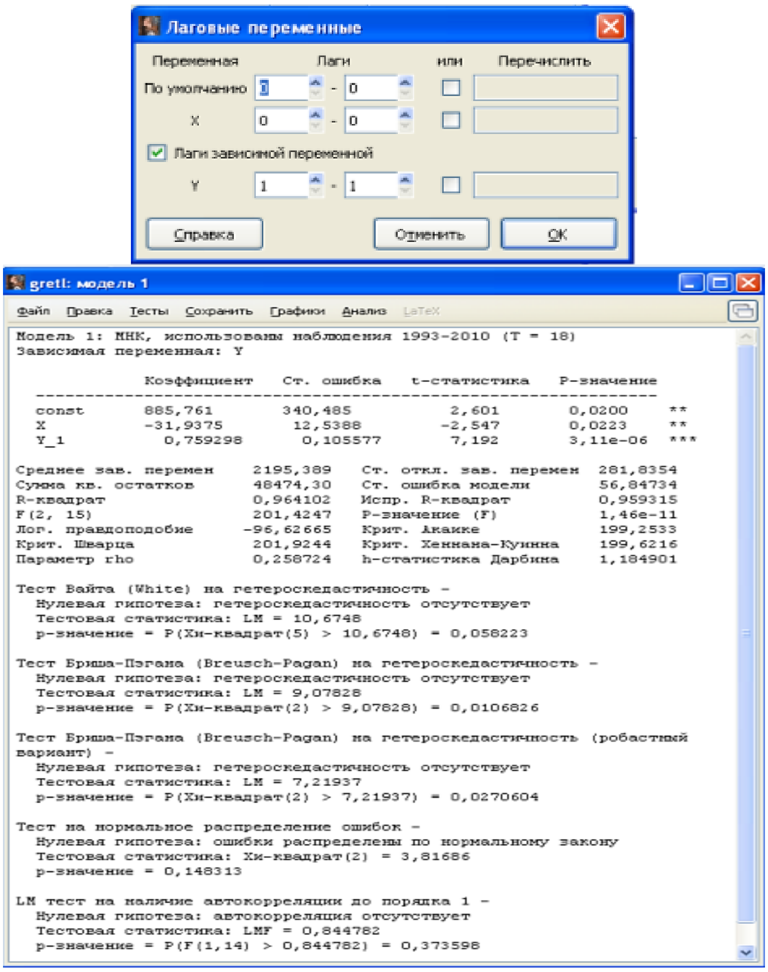


Рис. 6.5. Модель авторегрессии

В окне меню: Модель / Метод наименьших квадратов. Дополнительно в окне спецификации модели: Лаги / Лаги зависимой переменной / 1 (рис. 6.5).

Таким образом, уравнение, полученное в результате преобразования Койка имеет вид:

$$Y_t = 885{,}761 - 31{,}9375 X_t + 0{,}7593 Y_{t-1} + \varepsilon_t.$$

Параметры уравнения значимы, автокорреляция не обнаружена, остатки имеют нормальный закон распределения, но присутствует гетероскедастичность.

4. Вычисление параметров исходной модели с бесконечным числом лаговых переменных: $Y_t = a_0 + b_0 X_t + b_1 X_{t-1} + b_2 X_{t-2} + \ldots + \varepsilon$.

Введем скаляры: $a = 885,76$; $\lambda = 0,7593$; $b0 = -31,9375$; $b1 = b0 \cdot \lambda$; $b2 = b0 \cdot \lambda^2$ и т.д. (рис. 6.6).

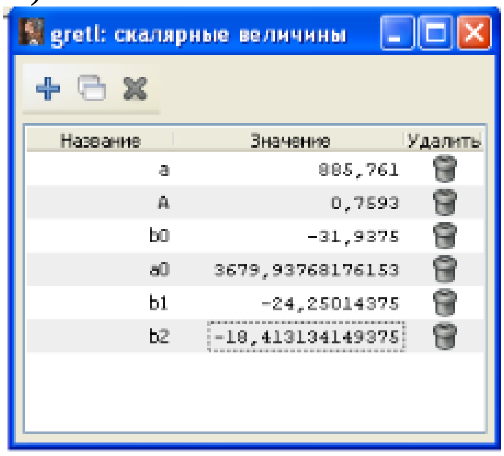


Рис. 6.6. Окно скаляров, полученных по модели авторегрессии

Таким образом, модель с бесконечным числом лаговых переменных в рассматриваемом случае записывается следующим образом:

$Y_t = 3679,9377 - 31,9375 X_t - 24,2501 X_{t-1} - 18,4131 X_{t-2} + \ldots + \varepsilon.$

5. Средний лаг модели: $l = \lambda / (1 - \lambda) = 3,15$ (рис. 6.7).

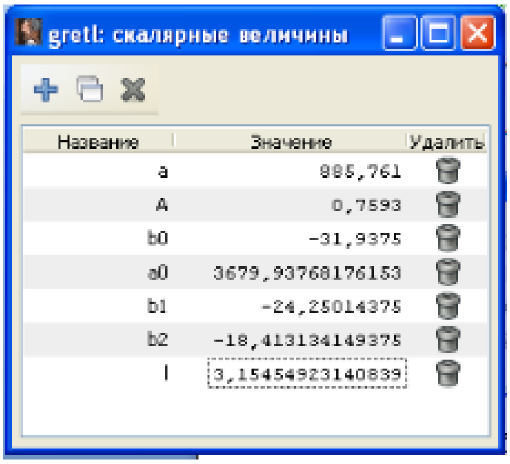


Рис. 6.7. Окно скаляров с параметрами исходной модели распределенных лагов (метод Койка)

Следовательно, очередной рост цен в среднем оказывает инфляционное воздействие на реальные доходы населения в течение отрезка времени, равного 3,15 года.

Задание для самостоятельного выполнения

Задание 6.3. Департамент экономического развития городской администрации проводит мониторинг социально-экономического развития региона. В частности, исследуется проблема взаимосвязи среднемесячной реальной заработной платы (*Y*, руб.) и уровня регистрируемой безработицы

(*X*, %) (таблица 6.3). Рассчитайте средний срок воздействия фактора безработицы на заработную плату, предварительно построив для этого модель распределенных лагов.

Таблица 6.3

Год	*Y*	*X*	Год	*Y*	*X*
1	21605	0,54	11	17948	1,17
2	21412	0,59	12	18094	1,31
3	21188	0,65	13	18214	1,20
4	20948	0,72	14	17314	0,96
5	21411	0,75	15	16517	1,15
6	21042	0,82	16	16020	1,33
7	21154	0,97	17	16028	1,47
8	20631	1,02	18	15608	1,47
9	19688	0,94	19	14991	1,31
10	18824	1,06	20	14605	1,40

Контрольные вопросы

1. Как определяются модели с распределенными лагами?
2. Как интерпретируют параметры модели с распределенным лагом?
3. Как определяются авторегрессионные модели?
4. Как интерпретируют параметры моделей авторегрессии?
5. В чем основная идея метода Алмон, к каким моделям он применяется?
6. Когда применяется преобразование Койка?

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Айвазян С.А. Методы эконометрики: Учебник /; Московская школа экономики МГУ им. М.В. Ломоносова (МШЭ). - М.: Магистр: ИНФРА-М, 2010. - 512 с.: 70x100 1/16. (переплет) ISBN 978-5-9776-0153-5, 1500 экз. Режим доступа: http://www.znanium.com/bookread.php?book=196548.

2. Айвазян, С. А. Эконометрика-2: продвинутый курс с приложениями в финансах [Электронный ресурс] : учебник / С. А. Айвазян, Д. Фантаццини. – М.: Магистр: НИЦ ИНФРА-М, 2014. – 944 с. – Режим доступа: http://znanium.com/bookread2.php?book=472607.

3. Балдин, К. В. Эконометрика [Электронный ресурс] : учеб. пособие / К. В. Балдин, О. Ф. Быстров, М. М. Соколов. – 2-е изд. перераб. и доп. – М.: ЮНИТИ-ДАНА, 2012. – 254 с. – Режим доступа: http://znanium.com/catalog.php?bookinfo=389655.

4. Берндт, Э. Р. Практика эконометрики: классика и современность: Учебник для студентов вузов/ Э. Р. Берндт; пер. с англ. под ред. проф. С. А. Айвазяна. - М.: ЮНИТИ-ДАНА, 2012. - 863 с. Режим доступа: http://www.znanium.com/bookread.php?book=389506.

5. Дайитбегов, Д. М. Компьютерные технологии анализа данных в эконометрике [Электронный ресурс] : монография / Д. М. Дайитбегов. – 3-е изд., испр. и доп. – М. : Вузовский учебник : НИЦ Инфра-М, 2013. – XIV, 587 с. – Режим доступа : http://znanium.com/bookread2.php?book=365692.

6. Исмагилов И. И., Кадочникова Е.И. Многофакторная регрессия в среде Gretl: учебно-методическое пособие. – Казань, изд-во К(П)ФУ, 2016 г. – 85 с. Режим доступа: http://dspace.kpfu.ru/xmlui/bitstream/handle/net/103980/Ekonometrika_PU_Praktikum_1_1_bibl.pdf;jsessionid=9BB1341CFA2B980810B2C4E89723E101?sequence=1

7. Картаев, Ф. С. Эконометрика [Электронный ресурс] : учеб. пособие / Ф. С. Картаев, Е. Н. Лукаш ; Московский государственный ун-т им. М. В. Ломоносова, Эконом. фак-т. – М. : Проспект, 2014. – 118 с. – Режим доступа: http://znanium.com/catalog.php?bookinfo=534320.

8. Крянев, А. В. Эконометрика : продвинутый курс [Электронный ресурс] : краткий конспект лекций / А. В. Крянев. – М. : КУРС : ИНФРА-М, 2017.– 62 с. – Режим доступа: http://znanium.com/bookread2.php?book=767248.

9. Куфель Т. Эконометрика. Решение задач с применением пакета программ Gretl. – М.: Горячая линия – Телеком, 2007. – 200 с.

10. Малова А. С. Основы эконометрики в среде Gretl: учебное пособие. – М.: Проспект, 2016. – 112 с.

11. Новиков, А. И. Эконометрика [Электронный ресурс] : учеб. пособие / А. И. Новиков. – 3-е изд., перераб. и доп. – М. : НИЦ ИНФРА-М, 2014. –

272 с. – Режим доступа: http://znanium.com/bookread2.php?book=437118

12. Плохотников К.Э. Основы эконометрики в пакете STATISTICA.: Учебное пособие / К.Э. Плохотников. - М.: Вузовский учебник, 2010. - 298 с.: Режим доступа: http://www.znanium.com/bookread.php?book=177719.

13. Подкорытова О. А. Анализ временных рядов: учебное пособие для бакалавриата и магистратуры с примерами в Gretl. –М.: Юрайт, 2016. – 266 с. – Серия: Бакалавр и магистр. Модуль.

14. Ратникова Т. А., Фурманов К. К. Анализ панельных данных и данных о длительности состояний: учебное пособие; НИУ «ВШЭ». – М.: Изд. дом ВШЭ, 2014 – 373 с.

15. Соколов, Г. А. Эконометрика : теоретические основы: учеб. пособие / Г. А. Соколов. – М. : ИНФРА-М, 2016. – 216 с. – Режим доступа: http://znanium.com/bookread2.php?book=503663

16. Эконометрика : учебник для магистров / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – М. : Юрайт, 2012. – 453 с.