КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ УПРАВЛЕНИЯ, ЭКОНОМИКИ И ФИНАНСОВ

Кафедра экономико-математического моделирования

И. И. ИСМАГИЛОВ, Е.И. КАДОЧНИКОВА

МНОГОФАКТОРНАЯ РЕГРЕССИЯ В СРЕДЕ GRETL

Учебно-методическое пособие для студентов,
обучающихся по направлению 38.04.01 «Экономика»

Казань 2016

УДК 330.43

ББК Ув631я73-1

Рекомендовано к публикации на заседании кафедры

экономико-математического моделирования

Протокол № 7 от 30 марта 2016 года

Рецензенты:

доктор экономических наук,

заведующий кафедрой бизнес-статистики и

математических методов в экономике

Казанского национального исследовательского технологического университета

А. В. Аксянова

кандидат технических наук,

доцент кафедры экономико-математического моделирования ИУЭиФ КФУ

В. А. Талызин

Исмагилов И.И., Кадочникова Е.И. Многофакторная регрессия в среде Gretl: учебно-методическое пособие для студентов, обучающихся по направлению 38.04.01 «Экономика» / И.И. Исмагилов, Е.И. Кадочникова – Казань: Казан. ун-т, 2016. – 62 с.

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

Содержание

Введение

Учебно-методическое пособие подготовлено в соответствии с программой дисциплины «Эконометрика (продвинутый уровень)», и требованиями действующего Федерального государственного образовательного стандарта высшего образования для магистерских программ ИУЭиФ КФУ.

Цель учебно-методического пособия – развить практические умения и навыки построения моделей многофакторной регрессии средствами пакета Gretl. Практические занятия по дисциплине «Эконометрика (продвинутый уровень)» ориентированы на развитие у студентов компетенций: ПК-3 **(**способность проводить самостоятельные исследования в соответствии с разработанной программой), ПК-8 **(**способность готовить аналитические материалы для оценки мероприятий в области экономической политики и принятий стратегических решений на микро- и макроуровне), ПК-9 **(**способность анализировать и использовать различные источники информации для проведения экономических расчетов), ПК-10 **(**способность составлять прогноз основных социально-экономических показателей деятельности предприятия, отрасли, региона и экономики в целом). На практических занятиях формируются 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 учебно-методическое 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 эко-

номической деятельности.

Решения рассматриваемых задач выполнены в пакете Gretl. Однако это не исключает возможность выполнения этих же заданий в других статистических и эконометрических пакетах (STATISTICA, SPSS, STATA, EViews). Ориентация на Gretl обусловлена следующими моментами. Во-первых, это 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.

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 учебно-методическое пособие для самостоятельной работы студентов заочного обучения.

1. Линейная многофакторная регрессия. Проверка качества регрессии и линейных ограничений на параметры

Расчетные формулы

Регрессионное уравнение – это уравнение вида

$$Y_t = \beta_0 + \beta_1 \cdot X_{t1} + \beta_1 \cdot X_{t1} + \ldots + \beta_m \cdot X_{tm} + \varepsilon_t, t = \overline{1,n}, \qquad (1.1)$$

где n – число наблюдений;

m - число регрессоров в модели;

ε - ошибка регрессии.

Модель такого вида имеет линейную спецификацию и называется классической линейной регрессионной моделью при выполнении следующих предпосылок:

1. $X_{t1}, X_{t1}, X_{tm}, t \in 1..n$ - детерминированные величины.

2. $E(\varepsilon_t) = 0$ – математическое ожидание ошибок равно нулю.

$E(\varepsilon^2_t) = \sigma^2$ – дисперсия ошибок не зависит от номера наблюдения.

3. $E(\varepsilon_t, \varepsilon_s) = 0$ – совместное математическое ожидание ошибок разных наблюдений равно нулю.

4. $\varepsilon_t \in N(0, \sigma^2)$ – ошибки регрессии, имеющие нормальное распределение.

Для получения оценок такой регрессионной модели применяется классический метод наименьших квадратов (МНК). МНК-оценки параметров регрессии можно определить векторно-матричным способом по следующей формуле:

$$B = \left(X'X\right)^{-1} X'Y,$$

где $B = [a, b_1, b_2, \ldots, b_m]$ – m+1-мерный вектор-столбец параметров эмпирического уравнения регрессии $y = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + e$;

$Y = [y_1, y_2, \ldots, y_n]$ –n-мерный вектор-столбец наблюдений зависимой переменной;

$\left(X'X\right)^{-1}$ - матрица, обратная к $X'X$;

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$ матрица регрессоров размерности $n \times (m+1)$.

<u>Тест Фишера (Fisher test)</u> позволяет проверить статистическую незначимость регрессии в целом, то есть установить, равны ли коэффициенты одновременно при всех регрессорах нулю на заданном уровне значимости α. Если коэффициенты регрессии признаются равными нулю, то регрессия считается незначимой. Регрессия значима, если коэффициент хотя бы при одном регрессоре отличен от нуля. Для проверки нулевой гипотезы $H_0 : \beta_1 = ... = \beta_m = 0$ используется расчетная статистика:

$$F = \frac{ESS / m}{RSS / (n - m - 1)} \qquad (1.2)$$

где ESS – факторная сумма квадратов отклонений;

RSS –сумма квадратов остатков.

Гипотеза H_0 отвергается, если $F > F_{\alpha(m,\ n-m-1)}$,

где α – уровень значимости, т. е. вероятность отвергнуть правильную нулевую гипотезу (обычно равен 0,1; 0,05; 0,01). Выбор значения α выполняется исследователем и во многом определяется размером выборки.

<u>Тест Стьюдента (t-test)</u> позволяет проверить незначимость отдельного коэффициента при регрессоре. Если коэффициент признается равным нулю, то регрессор считается незначимым, если коэффициент отличен от нуля - регрессор значим. Для проверки нулевой гипотезы $H_0 : \beta_j = 0$ используется расчетная статистика: $t_{bj} = \dfrac{b_j}{m_{bj}}$. (1.3)

Гипотеза H_0 отвергается, если $|t_{bj}| > t_{\alpha/2, n-m-1}$.

<u>Тест Стьюдента (t-test)</u> также позволяет проверить, равен ли коэффициент при соответствующем регрессоре некоторому значению, определенному иссле-

дователем. Для проверки нулевой гипотезы $H_0: \beta_j = \beta^*$ используется расчетная статистика $t_{bj} = \dfrac{b_j - \beta^*}{m_{bj}}$. (1.4)

Гипотеза H_0 отвергается, если $\left|t_{bj}\right| > t_{\alpha/2, n-m-1}$.

<u>Доверительный интервал для оценки коэффициента</u> позволяет определить границы, в которых истинное значение коэффициента находится с доверительной вероятностью 1-α, где α – уровень значимости:

$$\begin{aligned} & b_j - \Delta b_j < \beta_j < b_j + \Delta b_j \\ & \Delta b_j = t_{\alpha/2, n-m-1} \cdot m_{b\,j} \end{aligned} \quad . \qquad (1.5)$$

<u>Тест Фишера (Fisher test)</u> на совместную незначимость позволяет проверить статистическую незначимость нескольких коэффициентов в модели, то есть установить, равны ли коэффициенты одновременно при этих регрессорах нулю. Если коэффициент хотя бы при одном регрессоре отличен от нуля, регрессоры совместно значимы и нужно проводить дальнейший анализ, чтобы понять, какие именно регрессоры из незначимых в отдельности можно исключить из модели. Для проверки нулевой гипотезы $H_0: \beta_{j+1} = \ldots = \beta_{j+q} = 0$ используется расчетная статистика:

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m-1)}, \qquad (1.6)$$

где RSS_R – сумма квадратов остатков для модели с ограничением, в которой все регрессоры, подозрительные на совместную незначимость коэффициентов, исключены;

RSS_{UR} – сумма квадратов остатков для модели без ограничения, в которой включены все регрессоры;

q – количество подозрительных на совместную незначимость регрессоров.

<u>Коэффициент детерминации (R^2)</u> показывает, какая доля дисперсии зависимой переменной объясняется регрессионным уравнением.

$$R^2 = \frac{ESS}{TSS},$$

где TSS – общая сумма квадратов отклонений зависимой переменной.

Коэффициент детерминации изменяется от 0 до 1 и при наличии константы в линейной модели определяется также по формуле:

$$R^2 = 1 - \frac{RSS}{TSS} \qquad (1.7)$$

Информационный критерий Акаике (AIC) показывает качество подгонки модели (goodness of fit) и при этом вводит штраф за переизбыток регрессоров в модели. При прочих равных условиях следует предпочитать модель, в которой значение критерия более низкое:

$$AIC = 2 \cdot m - 2 \cdot \ln(L), \qquad (1.8)$$

где m – число регрессоров модели;

L – значение функции максимума правдоподобия для оцененной модели.

Информационный критерий Хеннана-Куинна (Hannan-Quinn information criterion – HQC) - при прочих равных условиях следует предпочитать модель, в которой значение критерия более низкое:

$$HQC = n \cdot \log\left(\frac{ESS}{n}\right) + 2 \cdot m \cdot \log(n), \qquad (1.9)$$

где m – число регрессоров модели;

ESS – факторная сумма квадратов отклонений;

n - число наблюдений.

Байесовский информационный критерий (BIC) или критерий Шварца (SC) - при прочих равных условиях следует предпочитать модель, в которой значение критерия более низкое:

$$BIC = SC = m \cdot \ln(n) - 2 \cdot \ln(L), \qquad (1.10)$$

где m – число регрессоров модели;

L – значение функции максимума правдоподобия для оцененной модели.

Для принятия решения можно руководствоваться результатами всех трех критериев и выбирать ту модель, в которой хотя бы два из трех критериев имеют наименьшее значение.

Тест Фишера (Fisher test) для линейных ограничений позволяет ответить на вопросы о том, связаны ли оценки коэффициентов регрессии между собой некоторым линейным соотношением. Для проверки нулевой гипотезы о том, что линейные ограничения верны используется расчетная статистика:

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-m-1)} \qquad (1.11)$$

где RSS_R – сумма квадратов остатков для модели с линейным ограничением для регрессоров;

RSS_{UR} – сумма квадратов остатков для модели без линейного ограничения;

q – количество линейных ограничений.

Задание 1.1. Имеются поквартальные данные об объеме продаж, тыс. руб. (Y) компании «Темп», сквозном номере квартала (X1), расходах на рекламу, тыс. руб. (X2), цене на товар, руб. (X3), цене конкурента, руб. (X4), индексе потребительских расходов, % (X5). По данным таблицы 1.2 оценить линейную модель многофакторной регрессии, построить доверительный интервал для коэффициента регрессии. С надежностью 95% проверить предположение о том, что увеличение расходов на рекламу на 1 тыс. рублей приведет к увеличению объема продаж на 12 тыс. рублей. Проверить, верно ли, что влияние расходов на рекламу на объем продаж в три раза меньше, чем влияние индекса потребительских расходов. Определить прогноз объема продаж в следующем квартале при возможных значениях факторов: расходы на рекламу запланированы на уровне 6 тыс. руб., цена на товар – 17 руб., цена конкурента – 18 руб., индекс потребительских расходов – 113,2%.

Таблица 1.2

Данные об объеме продаж компании «Темп»

Y	X1	X2	X3	X4	X5	Y	X1	X2	X3	X4	X5
126	1	4	15	17	100	367	9	19,8	15,8	18,2	108,3
137	2	4,8	14,8	17,3	98,4	367	10	10,6	16,9	16,8	109,2
148	3	3,8	15,2	16,8	101,2	321	11	8,6	16,3	17	110,1
191	4	8,7	15,5	16,2	103,5	307	12	6,5	16,1	18,3	110,7
274	5	8,2	15,5	16	104,1	331	13	12,6	15,4	16,4	110,3
370	6	9,7	16	18	107	345	14	6,5	15,7	16,2	111,8
432	7	14,7	18,1	20,2	107,4	364	15	5,8	16	17,7	112,3
445	8	18,7	13	15,8	108,5	384	16	5,7	15,1	16,2	112,9

Решение в среде Gretl

1. Создание рабочего листа 1 с исходными данными в Excel и его сохранение в файле «ЗанятиеМР.xlsx».

2. Импорт данных из таблицы Excel. В основном меню выберем пункт: Файл / Открыть / Импорт / Excel / лист 1.

3. Построим линейную модель многофакторной регрессии классическим методом наименьших квадратов (МНК), включив все факторы. Корреляционный анализ в данном задании не проводится. Выбор наиболее подходящей спецификации модели выполняется за счет теста на избыточные переменные. В основном меню выберем пункт: Модель / Метод наименьших квадратов и построим линейную модель с полным набором факторов (рис. 1.1).

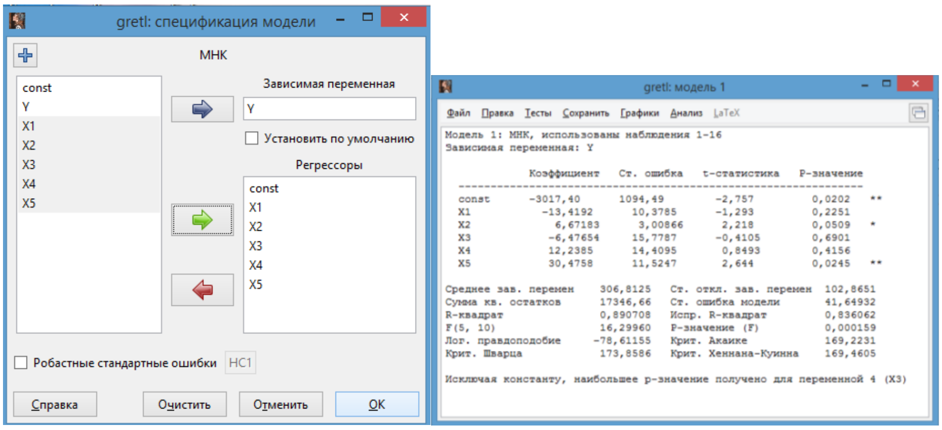


Рис. 1.1. МНК-оценки исходной регрессии

Обратим внимание на соответствие знаков коэффициентов регрессии экономическому смыслу задачи: с увеличением расходов на рекламу, цены конкурента и индекса потребительских расходов объем продаж увеличивается, а увеличение цены приводит к снижению объема продаж. Для проверки качества регрессии применим тест Фишера для проверки незначимости регрессии в целом и тест Стьюдента для проверки незначимости отдельного коэффициента при регрессоре.

4. Применение теста Фишера (F-теста). Самый быстрый и простой способ: если Р-значение (F) < 0,01, то модель значима на уровне значимости α=0,01 (рис. 1.12). Также можно воспользоваться другими возможностями

Gretl. В основном меню выберем пункт Инструменты / Критические значения / Фишера и введем нужные параметры распределения (рис.1.2).

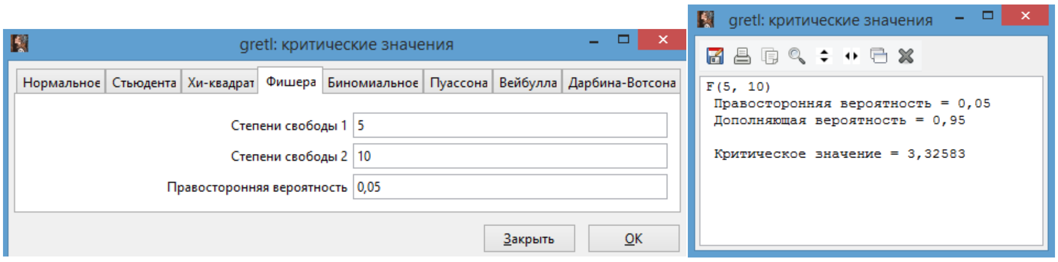


Рис. 1.2. Окно Инструмента «Критические значения» (Фишера)

Сравниваем расчетное значение F-статистики с критическим значением распределения Фишера при заданном уровне значимости α: Fрасч > F(α;m;n-m-1). При α=0,05: 16,2996>3,3258. Следовательно, нулевая гипотеза о незначимости регрессии в целом отвергается на уровне значимости α=0,05, то есть коэффициенты одновременно при всех регрессорах не равны нулю, что подтверждает совместное влияние факторов на зависимую переменную.

Обратим внимание на самый простой способ проверки нулевой гипотезы о незначимости отдельного коэффициента при регрессоре: если Р-значение (t) < 0,01, то коэффициент значим с надежностью 99%. При этом в строке коэффициента указаны одна звездочка (для значимости на уровне α=0,1), две звездочки (для значимости на уровне α=0,05), три звездочки (для значимости на уровне α=0,01). Также можно воспользоваться другими возможностями Gretl. В основном меню выберем Инструменты / Критические значения / Стьюдента и введем нужные параметры распределения. Заметим, что для распределения Стьюдента надо вводить не двухстороннюю вероятность 1-α=0,05, а только правостороннюю, то есть в нашем случае 0,025 (рис. 1.3).

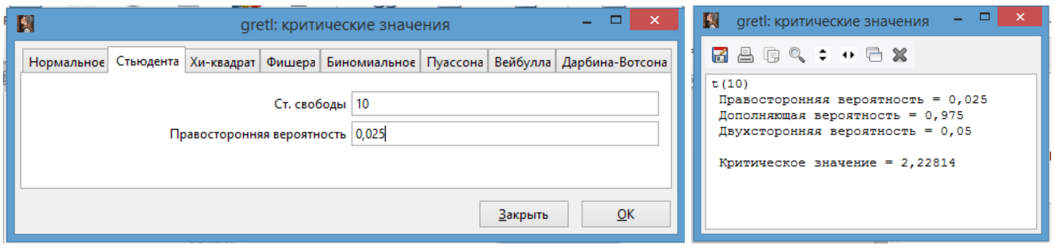


Рис. 1.3. Окно Инструмента «Критические значения» (Стьюдента)

Сравниваем расчетное и критическое значения t-статистик для каждой переменной. В нашем случае $|t_{b1}| < t_{0,05/2;10}$, $|t_{b2}| > t_{0,05/2;10}$, $|t_{b3}| < t_{0,05/2;10}$, $|t_{b5}| > t_{0,05/2;10}$, $|t_{b1}| < t_{0,05/2;10}$, отсюда можно сделать вывод, что регрессоры X2 и X5 значимы на уровне α=0,05.

5. Выполним проверку модели на коллинеарность факторов, используя критерий VIF. В окне модели: Тесты / Мультиколлинеарность (рис.1.4).

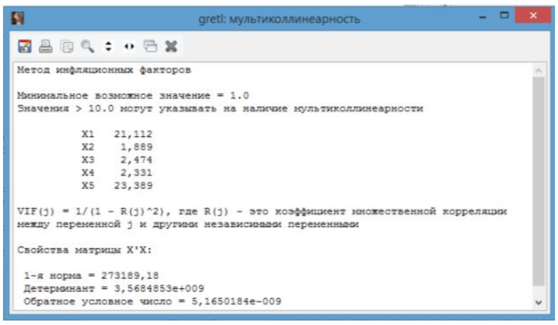


Рис. 1.4. Результаты теста на коллинеарность методом инфляционных факторов

Значения критерия VIF для факторов X1, X5 больше, чем значение 10, поэтому данные факторы признаются коллинеарными, включать их в модель регрессии одновременно нельзя.

6. Для отбора регрессоров выполним последовательное исключение избыточных переменных. В окне модели вызовем пункт меню: Тесты / Избыточные переменные (рис. 1.5)

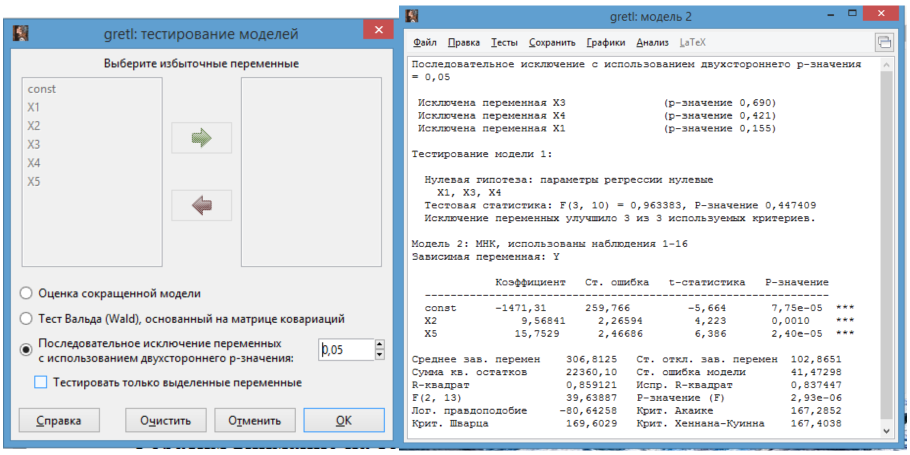


Рис. 1.5. Исключение избыточных переменных

Как видно на рисунке 1.5. модель после исключения избыточных переменных имеет больший скорректированный коэффициент детерминации (Исправленный R^2), меньшие значения информационных критериев Шварца, Акаике, Хеннана-Куинна. Регрессоры X2, X5 значимы на уровне α=0,01. Анализ свойств ряда остатков путем их тестирования на гетероскедастичность и наличие автокорреляции для подтверждения корректности МНК-оценок регрессии рассмотрен в теме 3.

Линейная двухфакторная модель регрессии имеет следующий вид:

$Y=-1471,31+9,568X_2+15,753X_5+e$, R^2скорр.=0,837.

С увеличением расходов на рекламу на 1 тыс. руб. объем продаж увеличивается в среднем на 9,568 тыс. руб. при неизменном индексе потребительских расходов, а при росте индекса потребительских расходов на 1 процентный пункт, объем продаж увеличивается в среднем на 15,753 тыс. руб. при неизменных расходах на рекламу. Модель объясняет 84% вариации объема продаж. Для измерения прогнозных качеств модели определим среднюю абсолютную процентную ошибку – MAPE. В окне модели: Анализ / Прогнозы (рис. 1.6).

```
Статистика для оценки прогноза

Средняя ошибка (ME)                                  -1,9895e-013
Средняя квадратичная ошибка (MSE)                     1397,5
Корень из средней квадратичной ошибки (RMSE)          37,383
Средняя абсолютная ошибка (MAE)                       31,368
Средняя процентная ошибка (MPE)                      -1,6628
Средняя абсолютная процентная ошибка (MAPE)           10,646
U-статистика Тейла (Theil's U)                        0,79754
Пропорция смещения, UM                                0
Пропорция регрессии, UR                               0
Пропорция возмущений, UD                              1
```

Рис. 1.6. Окно статистик для оценки прогноза

Средняя абсолютная процентная ошибка составила 10,646%. Снова выполним проверку модели на коллинеарность, используя критерий VIF (рис.1.7).

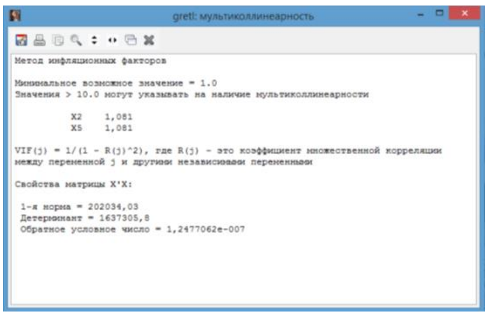

```
Метод инфляционных факторов

Минимальное возможное значение = 1.0
Значения > 10.0 могут указывать на наличие мультиколлинеарности

         X2     1,081
         X5     1,081

VIF(j) = 1/(1 - R(j)^2), где R(j) - это коэффициент множественной корреляции
между переменной j и другими независимыми переменными

Свойства матрицы X'X:

 1-я норма = 202034,03
 Детерминант = 1637305,8
 Обратное условное число = 1,2477062e-007
```

Рис. 1.7. Результаты теста на коллинеарность факторов X2, X5

Как видно из рис.1.7, значение критерия VIF для каждого из факторов X2, X5 меньше, чем значение 10, значит данные факторы независимы друг от друга, коллинеарности нет.

Для ранжирования факторов X2, X5 по силе влияния на объем продаж выполним расчет средних коэффициентов эластичности. Прежде всего, определим средние значения переменных: Переменная / Описательная статистика (рис. 1.8)

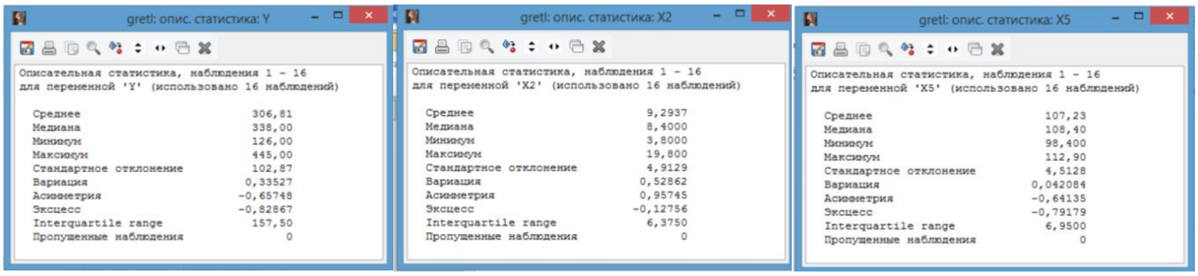


Рис. 1.8. Описательная статистика для переменных Y, X2, X5

Используя кнопку «Просмотр сессии» (четвертая слева) в нижней панели инструментов Gretl, введем скаляры MY=306,81; MX2=9,294; MX5=107,23; b2=9,568; b5=15,75 (рис. 1.9) и определим средние коэффициенты эластичности по формулам: $Э_2 = b_2 \cdot \frac{\bar{x}_2}{\bar{y}}$; $Э_5 = b_5 \cdot \frac{\bar{x}_5}{\bar{y}}$.

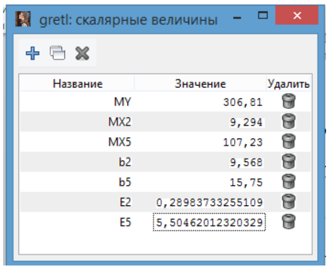


Рис. 1.9. Окно ввода скалярных величин

С увеличением расходов на рекламу на 1% относительно своего среднего уровня объем продаж возрастает на 0,29% среднего значения при среднем уровне индекса потребительских расходов. Увеличение индекса потребительских расходов на 1% относительно своего среднего уровня приводит к росту объема продаж на 5, 505% среднего значения при среднем уровне расходов на рекламу.

7. Для определения с вероятностью 1-α доверительных интервалов, в границах которых находятся истинные значения для коэффициентов регрессии используем встроенную функцию Gretl. В окне модели вызовем пункт меню Анализ / Доверительные интервалы для коэффициентов (рис. 1.7).

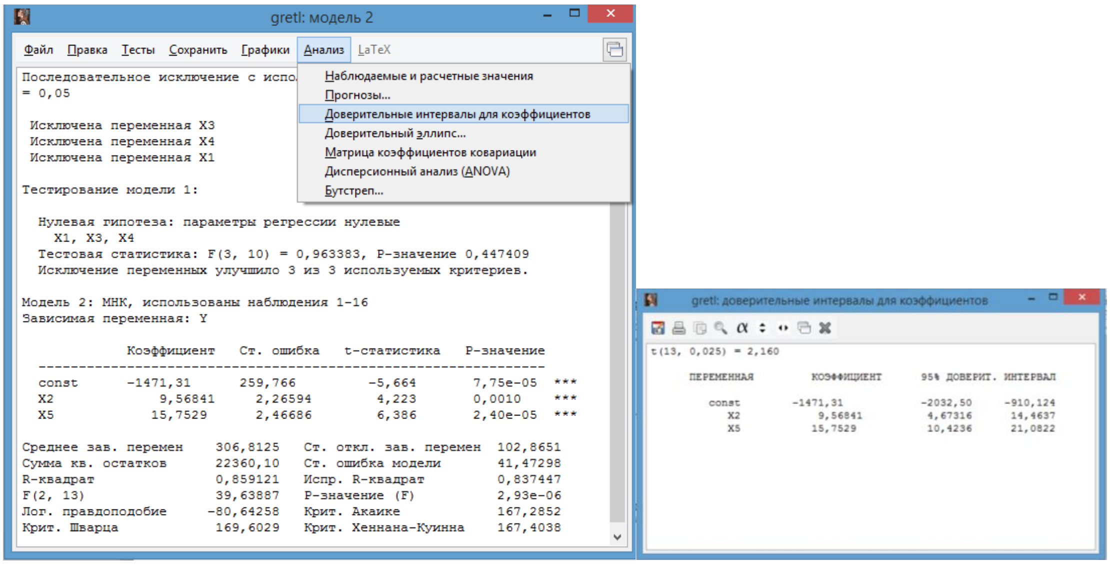


Рис. 1.7. Доверительные интервалы для коэффициентов регрессии

Истинное значение коэффициента при переменной X2 «Расходы на рекламу» с вероятностью 95% накрывается интервалом $[4{,}673;14{,}464]$. Истинное значение коэффициента при переменной X5 «Индекс потребительских расходов» с вероятностью 95% накрывается интервалом $[10{,}424;21{,}082]$.

Отметим, что если доверительный интервал накрывает 0, то можно сделать вывод о том, что коэффициент не значим. Еще при помощи доверительного интервала можно проверить предположение о том, может ли коэффициент при регрессоре равняться заданному значению. В нашем случае, на вопрос задания о том, приведет ли увеличение расходов на рекламу на 1 тыс. рублей к увеличению объема продаж на 12 тыс. рублей, с надежностью 95% можно ответить утвердительно.

8. Ответим на вопрос, верно ли, что влияние расходов на рекламу на объем продаж в три раза меньше, чем влияние индекса потребительских расходов. Данная гипотеза будет выглядеть представленным образом:

H0:$3\beta_2=\beta_5$,

H1: ограничение не выполняется.

Чтобы проверить гипотезу, нужно провести тест Фишера для линейных ограничений и ответить на вопросы о том, связаны ли оценки коэффициенты регрессии между собой некоторым линейным соотношением. Если наложить указанное в нулевой гипотезе ограничение на исходную модель:

$Y=\beta_0+\beta_2X_2+\beta_5X_5+\varepsilon$,

то в результате мы получим модель вида

$Y=\beta_0+\beta_2(X_2+3*X_5)+\varepsilon$, которая является моделью с ограничением.

Чтобы оценить модель с ограничением, создадим новую переменную X2_X5 = X2+3*X5. В основном меню Gretl выберем пункт: Добавить / Добавить новую переменную, оценим модель с вновь созданной переменной (рис. 1.8).

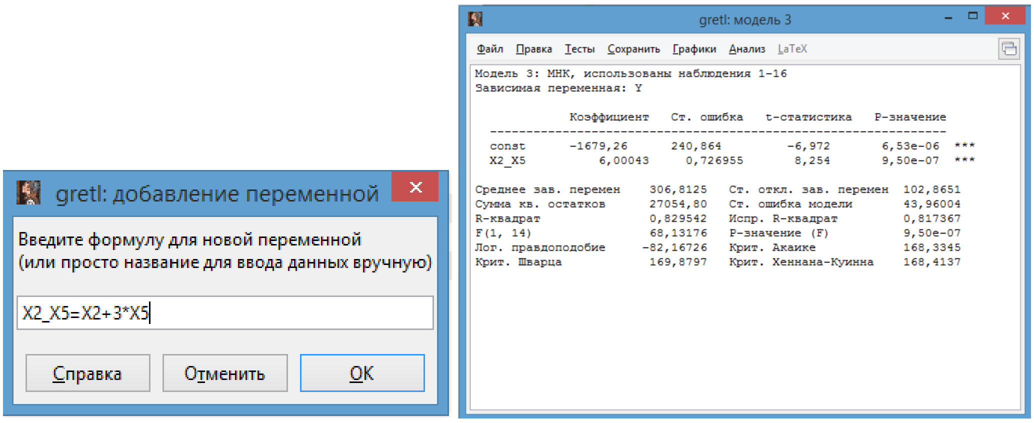


Рис. 1.8. МНК-оценки модели с ограничением

В полученной модели коэффициент при регрессоре X2_X5 значим, модель в целом значима. Рассчитаем значение F-статистики:

$$F=\frac{(RSS_R-RSS_{UR})/q}{RSS_{UR}/(n-m-1)}=\frac{(27054{,}80-22360{,}10)/1}{22360{,}10/(16-2-1)}=2{,}729$$

$F_{0,05;1;13}$=4,667. Расчетное значение F-статистики меньше, чем критическое, следовательно, утверждение о том, что влияние расходов на рекламу на объем продаж в три раза меньше, чем влияние индекса потребительских расходов, верно.

Чтобы провести этот тест в автоматическом режиме в меню модели без ограничений выберем пункт Тесты / Линейные ограничения и запишем ограничение, получим результаты (рис. 1.9), которые не противоречат ручному расчету (P-значение критерия >0,05).

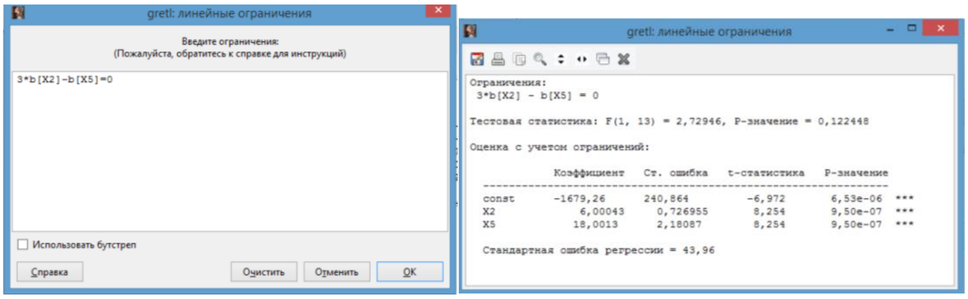


Рис. 1.9. Результаты проверки линейного ограничения

Получим прогноз объема продаж при заданных условием задания X2=6 и X5=113,2:

Y=-1471,31+9,568*6+15,753*113,2=369,338 тыс. руб.

Задание 1.2. Имеются данные о прибыли (Y, тыс. руб.), издержках обращения (X1, тыс. руб.) и обороте (X2, тыс. руб.) 20 организаций оптовой торговли. Оценить линейную модель многофакторной регрессии с использованием векторно-матричного способа МНК и встроенных средств Gretl, провести анализ ее качества по данным таблицы 1.2. Построить прогноз прибыли, если каждый из факторов увеличить на 5% от его среднего уровня.

Таблица 1.2

Кросс-секционные данные о кредитных организациях

№ пп	Y	X1	X2	№ пп	Y	X1	X2
1	21,324	11,67	42,012	11	114,123	63,33	227,988
2	27,765	15	54	12	123,759	68,33	245,988
3	30,954	16,67	60,012	13	138,355	76,68	276,048
4	45,125	25	90	14	138,124	76,68	276,048
5	58,123	32,22	115,992	15	141,786	78,33	281,988
6	60,473	33,33	119,988	16	135,668	75	270

7	69,965	38,33	137,988	17	146,771	81,12	292,032
8	75,456	41,67	150,012	18	149,768	82,78	298,008
9	92,667	51,11	183,996	19	158,234	87,78	316,008
10	105,457	58,33	209,988	20	167,463	92,77	333,972

Решение в среде Gretl

1. Создание листа 2 в файле исходных данных в Excel «ЗанятиеMP.xlsx» и импорт данных из таблицы Excel: Файл / Открыть / Импорт / Excel / файл «ЗанятиеMP.xlsx» /лист 2 (рис. 1.8).

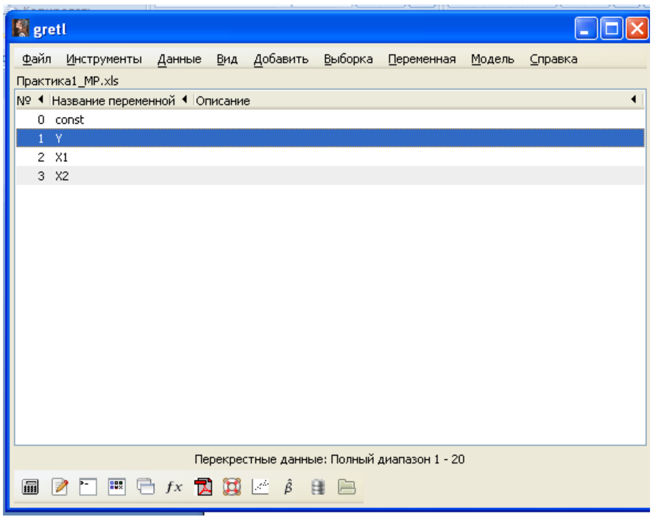


Рис. 1.8. Финальное окно импорта данных

2. Расчет линейных коэффициентов парной корреляции: Вид / Корреляционная матрица (рис. 1.2)

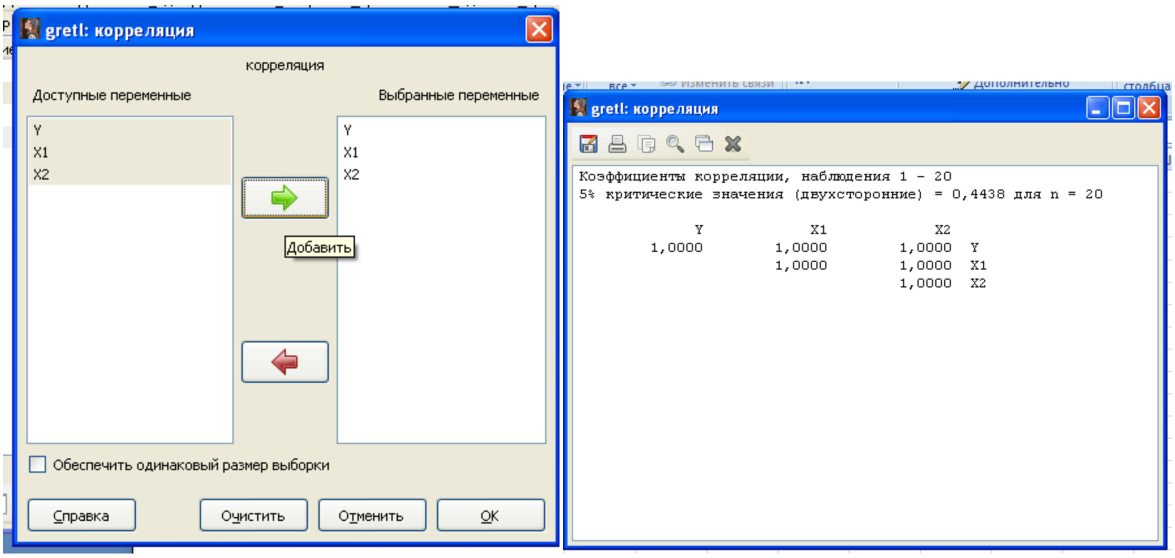


Рис. 1.9. Матрица линейных коэффициентов парной корреляции

Знак линейного коэффициента корреляции между зависимой переменной Y и факторами X1, X2 ($r_{yx1}>0$, $r_{yx2}>0$) показывает, с увеличением издержек обращения и товарооборота прибыль оптовой организации увеличивается, что не противоречит экономическому смыслу задачи. При этом, линейный коэффициент парной корреляции между факторами (r_{x1x2}) равен единице, значит факторы X1 и X2 являются совершенно коллинеарными. Это следует из того, что издержки обращения (X1) являются слагаемым товарооборота. Поэтому не следует в качестве переменных, в частности, регрессоров, выбирать показатели, связанные между собой через расчетные формулы.

Для проверки регрессоров на коллинеарность можно применить метод расчета определителя матрицы линейных коэффициентов парной корреляции. Для формирования матрицы регрессоров: Добавить / Матрицу (рис. 1.10).

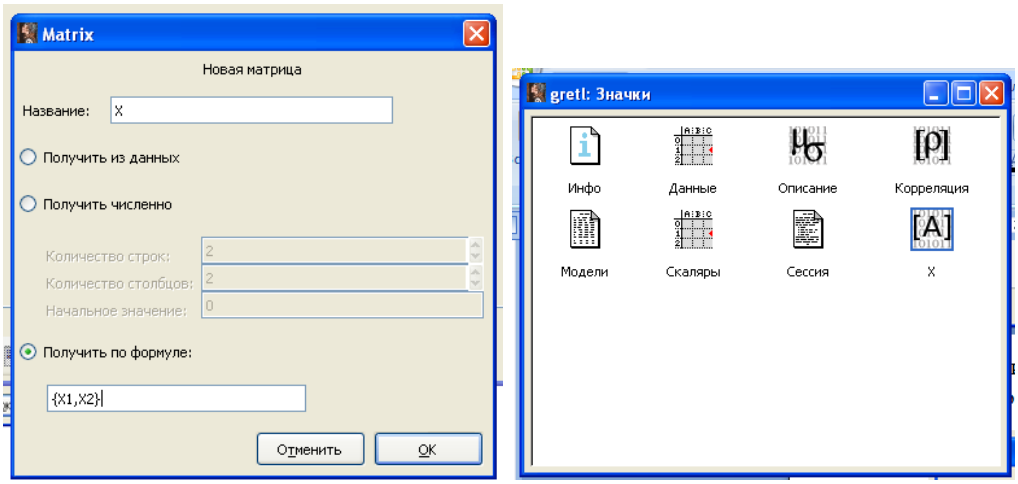


Рис. 1.10. Окно добавления матрицы регрессоров

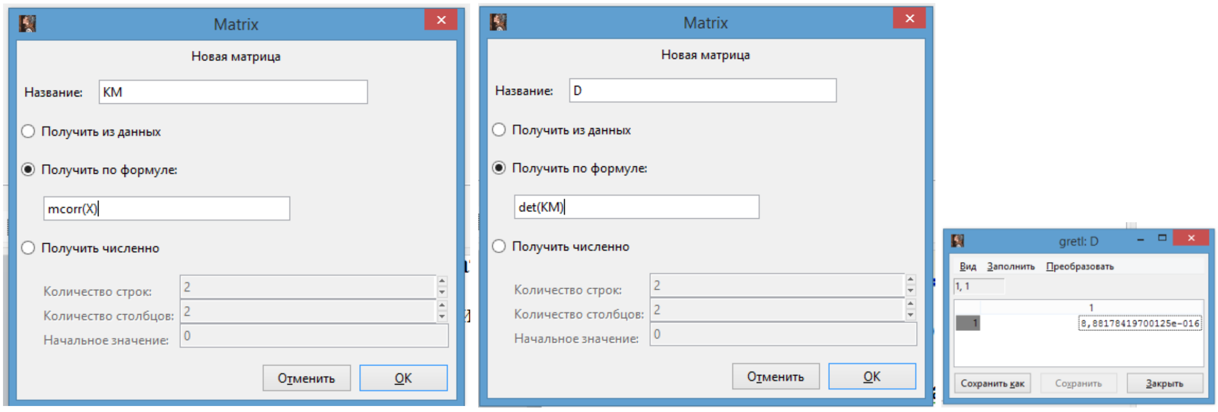


Рис. 1.11. Окна добавления матрицы линейных коэффициентов межфакторной парной корреляции и расчета ее определителя

Затем получаем матрицу линейных коэффициентов межфакторной парной корреляции: Добавить / Матрицу, и выполняем расчет определителя матрицы линейных коэффициентов парной корреляции (рис. 1.11).

Определитель матрицы det=8,882e-016, практически равен нулю, что подтверждает наличие совершенной коллинеарности регрессоров.

В таких условиях одновременно включать два коллинеарных регрессора X1, X2 в многофакторную модель нельзя, поэтому построим две линейные модели парной регрессии отдельно с каждым из них. Обычно предпочтение отдается тому фактору, который имеет наиболее тесную взаимосвязь с зависимой переменной Y.

3. Применим матричный способ для оценивания параметров линейной модели парной регрессии с фактором X2 классическим методом наименьших квадратов. Добавим единичный вектор X0: Добавить / Добавить новую переменную. Затем сформируем структурную матрицу XC из единичного вектора X0 и регрессора X2: Добавить / Матрицу… (рис. 1.12)

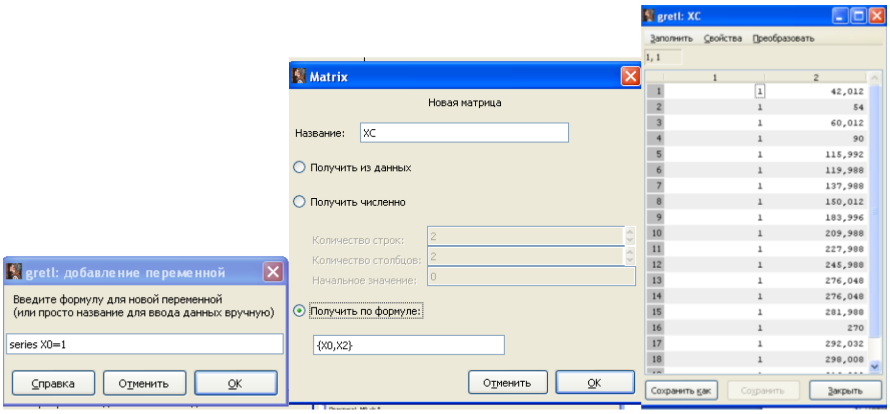


Рис. 1.12. Добавление единичного вектора и структурной матрицы

Для получения вектора оценок параметров регрессии B используем возможности Gretl для добавления матрицы:

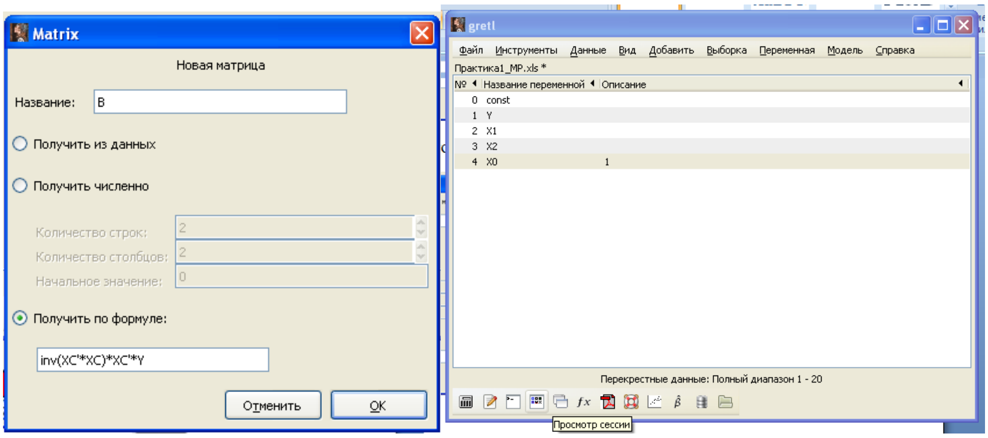


Рис. 1.13. Получение вектора оценок параметров регрессии B

Для просмотра значений вектора оценок параметров регрессии B используем кнопку просмотра сессии (четвертая слева) на нижней панели инструментов Gretl (рис. 1.14).

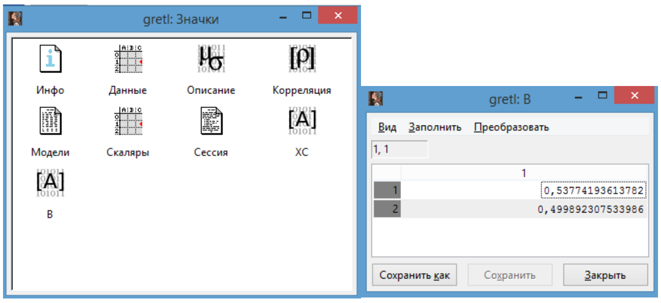


Рис. 1.14. Вектор оценок параметров регрессии B

3. Применим скалярный способ для оценивания параметров линейной модели парной регрессии с фактором X1 классическим методом наименьших квадратов: Модель / Метод наименьших квадратов (рис. 1.15)

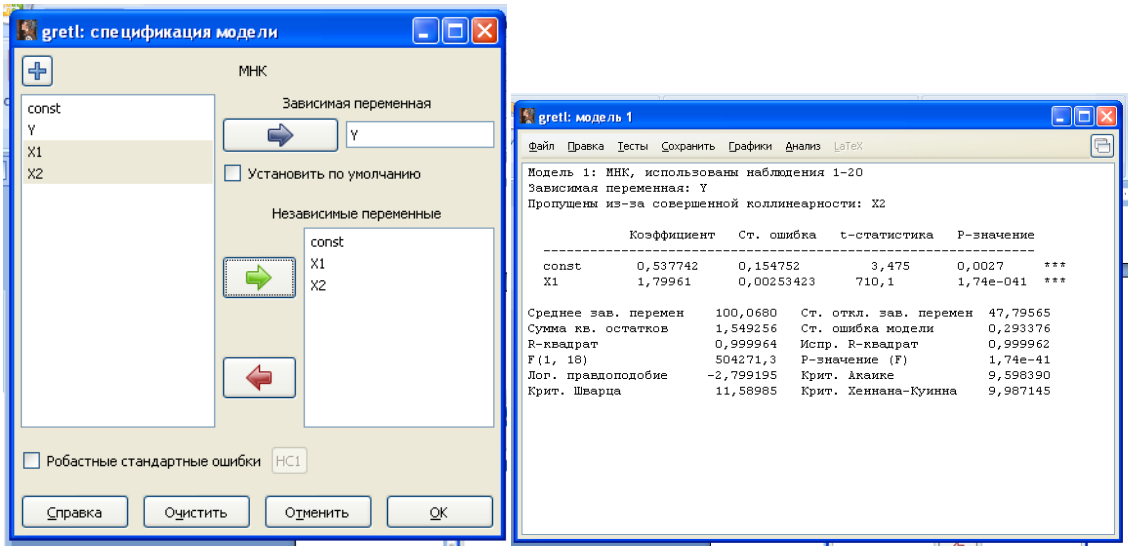


Рис. 1.15. Окно спецификации модели и МНК-оценки с регрессором X1

Запишем модель: Y=0,538+1,800X1+e, R2=0,999. Коэффициент регрессии показывает, что с увеличением издержек обращения на 1 тыс. руб. прибыль торговых организаций увеличивается в среднем на 1,8 тыс. руб. Модель объясняет 99% вариации зависимой переменной.

4. Изменение спецификации модели для включения регрессора X2: Правка / Изменить модель (рис. 1.16)

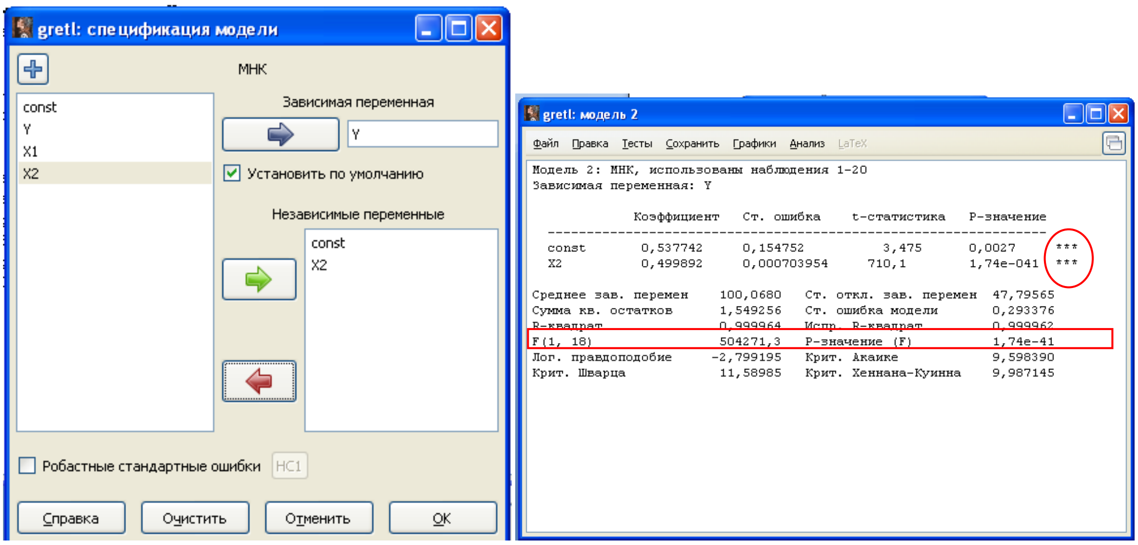


Рис. 1.16. Окно спецификации модели и МНК-оценки с регрессором X2

Запишем модель: Y=0,538+0,500X2+e, R2=0,999. Коэффициент регрессии показывает, что с увеличением товарооборота на 1 тыс. руб. прибыль торговых организаций увеличивается в среднем на 0,5 тыс. руб. Модель объясняет 99% вариации зависимой переменной.

5. Применение теста Фишера (F-теста) для проверки незначимости регрессии в целом. Самый быстрый и простой способ: если p-значение (F) < 0,01, то модель значима (рис. 1.9). Также можно воспользоваться другими возможностями Gretl. В основном меню выберем Инструменты / Критические значения / Фишера и введем нужные параметры распределения (рис.1.17).

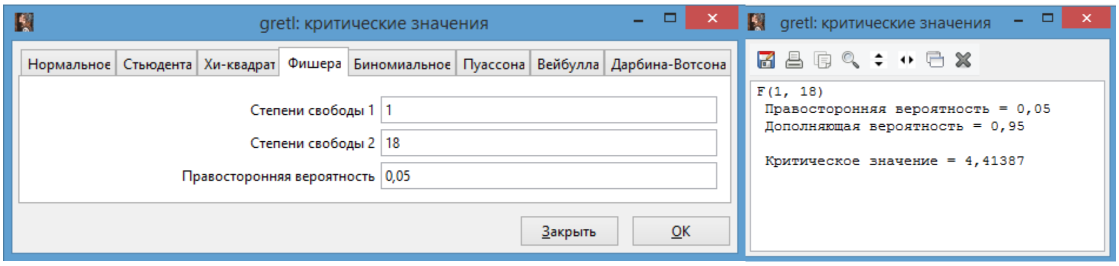


Рис. 1.17. Окно Инструмента «Критические значения» (Фишера)

Сравниваем расчетное значение F-статистики с критическим Fрасч > F(α;m;n-m-1) при α=0,05: 504271,3>4,414. Можно сделать вывод, что нулевая гипотеза о незначимости регрессии в целом отвергается при α=0,05.

8. Применение теста Стьюдента (t-теста) для проверки незначимости отдельного коэффициента при регрессоре. Статистически значимые параметры модели обозначены в конце строки звездочками (рис. 1.16). 1 звездочка – значимость при уровне 10%, 2 звездочки – значимость при уровне 5 %, 3 звездочки – значимость при уровне 1 %. В сообщении, выдаваемом после оценивания модели, предлагается последовательность исключения переменных (если она необходима). Также можно воспользоваться другими возможностями Gretl. В основном меню выберем Инструменты / Критические значения / Стьюдента и введем нужные параметры распределения. Заметим, что для распределения Стьюдента надо вводить не двухстороннюю вероятность α=0,05, а только правостороннюю, то есть в нашем случае 0,025 (рис. 1.18).

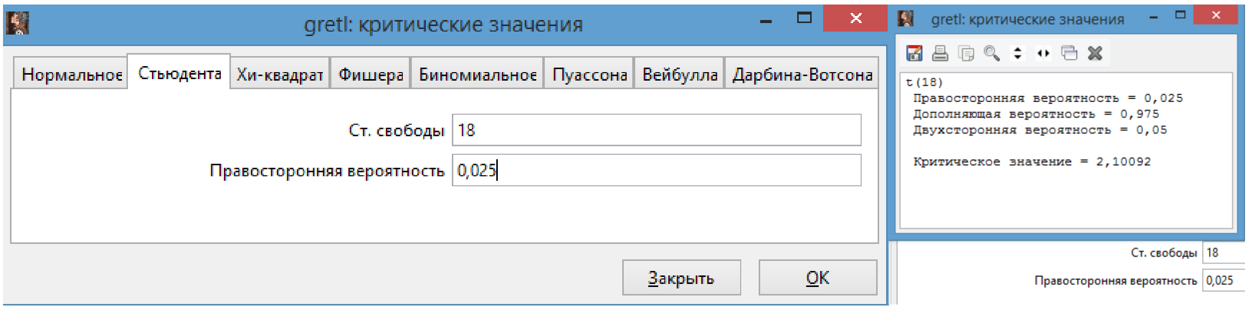


Рис. 1.18. Окно Инструмента «Критические значения» (Стьюдента)

Сравниваем расчетное и критическое значения t-статистик для переменной X2. В нашем случае $|t_b| > t_{0,05/2;18}$, то есть 710,1>2,101, отсюда можно сделать вывод, что регрессор X2 значим.

Обе линейные модели парной регрессии имеют хорошее качество подгонки и согласованную с экономической теорией экономическую интерпретацию. Для получения прогноза прибыли определим прогнозные значения каждого из факторов, увеличив средние значения на 5%, введем необходимые скаляры (рис. 1.19). Используем в скалярах следующие формулы:

PX1=MX1*procent;

PX2=MX2*procent;

PY1=a+b1*PX1;

PY2=a+b2*PX2.

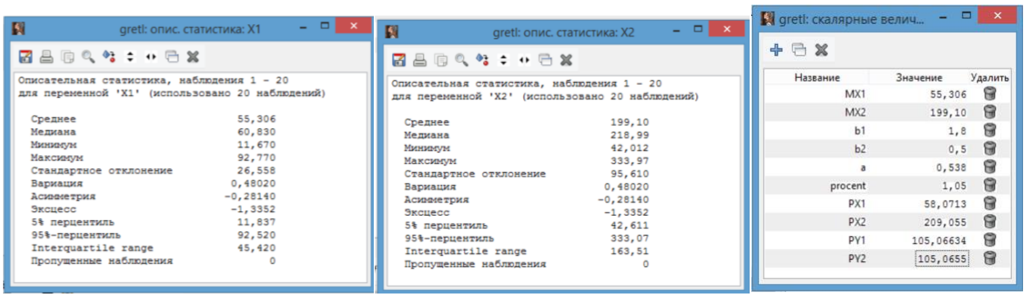


Рис.1.19. Окно ввода скалярных величин

Прогноз прибыли составляет 105,066 тыс. руб.

2. Оценивание параметров линейной модели множественной регрессии в условиях мультиколлинеарности. Гребневая регрессия

Расчетные формулы

Вектор МНК-оценок регрессии:

$$B = (X'X)^{-1} \cdot X'Y \qquad (2.1)$$

Ридж-регрессия или гребневая регрессия предполагает оценку параметров по следующей формуле:

$$B = (X^T X + \lambda I)^{-1} X^T Y \qquad (2.2)$$

где I – единичная матрица, λ – параметр ридж-регрессии («гребень»).

Стандартная ошибка S_{bk} ридж-оценки *k*-го коэффициента регрессии, равная корню квадратному из соответствующего диагонального элемента ковариационной матрицы векторной оценки:

$$S_b = \hat{\sigma}\sqrt{(X'X + \lambda I)^{-1}} \ , \qquad (2.3)$$

где $\hat{\sigma}^2 = \dfrac{E'E}{n-m-1}; E = Y - XB$.

Добавление параметра λ решает проблему плохой обусловленности матрицы $X^T X$. Эти оценки смещены, в отличие от МНК-оценок. Однако доказано, что существует такое λ, при котором эти оценки более эффективны, чем оценки МНК (МНК наиболее эффективны среди несмещенных оценок). Тем не менее, четких правил выбора этого параметра нет.

Гребневые оценки параметров регрессии хоть и смещены, но имеют лучшие характеристики точности. В сущности, гребневая регрессия искусственно занижает коэффициенты корреляции, чтобы могли быть вычислены более устойчивые оценки коэффициентов регрессии. Проблема использования данного метода на практике сводится к выбору подходящего значения λ, обычно применяется перебор значений дискретной сетки в интервале от 0,1 доя 0,4.

Задание 2.1. Для 20 городов России были собраны данные о годовых расходах на печать (Y, млн. руб.) наиболее популярных газет, объемах розничной продажи газет в городе (X1, млн. руб.) и количестве семей в городе (X2, тыс. ед.). Для факторов были взяты логарифмы (X1, X2) с целью уменьшения разброса данных и упрощения их обработки. Построить модель множественной регрессии и провести анализ ее качества по данным таблицы 2.1.

Таблица 2.1

Данные о годовых расходах на печать наиболее популярных газет

Y	2,8	2,3	2,5	2,7	2,6	2,4	3,5	2,1	1,9	3,8
X1	4,4	3,68	3,92	4,32	4,24	3,76	5,52	3,36	3,04	5,92
X2	3,00	2,51	2,67	2,94	2,89	2,56	3,76	2,29	2,07	4,03
Y	2,7	2,9	3,3	2,7	1,8	1,9	2,2	2,3	3,6	2,2
X1	4,24	4,72	5,28	4,24	3,04	3,12	3,6	3,68	5,76	3,6
X2	2,89	3,22	3,6	2,89	2,07	2,13	2,46	2,51	3,92	2,45

Решение в среде Gretl

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

2. Построение линейной модели множественной регрессии и проверка ее на коллинеарность факторов методом инфляционных факторов (критерий VIF): Модель / Метод наименьших квадратов. Затем в окне модели: Тесты / Мультиколлинеарность (рис. 2.1).

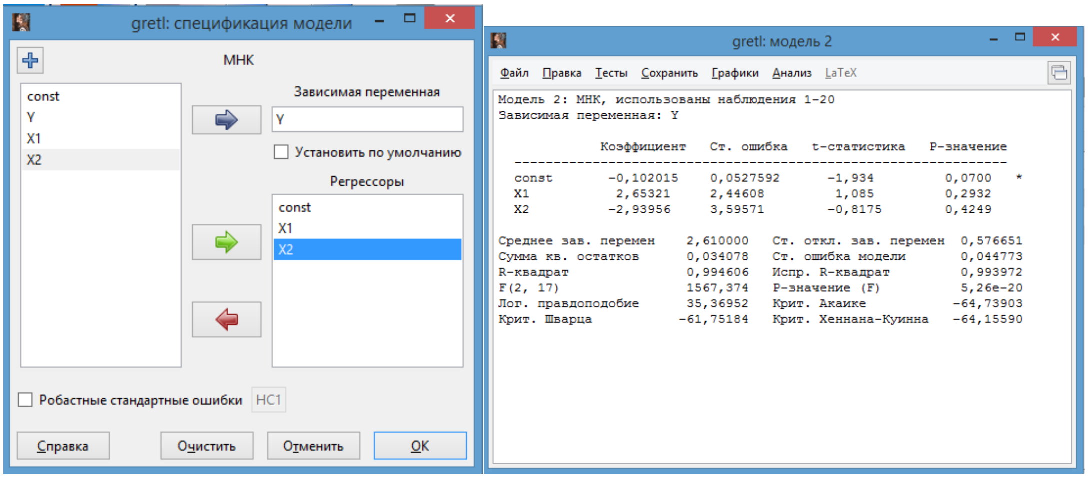


Рис. 2.1. Построение линейной модели регрессии с двумя факторами

Высокое значение R^2 при одновременно незначимых коэффициентах регрессии является признаком наличия коллинеарных факторов, (рис. 2.2).

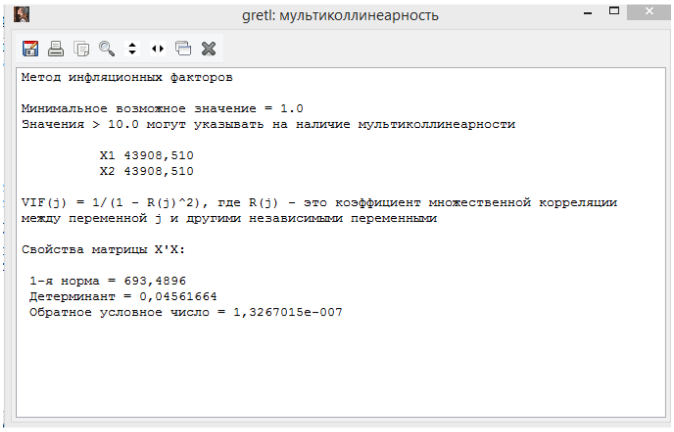


Рис. 2.2. Проверка на коллинеарность по критерию VIF

Метод инфляционных факторов подтверждает наличие мультиколлинеарности в линейной модели множественной регрессии.

3. Добавление новой переменной X0: Добавить /Добавить новую переменную… (рис. 1.3).

4. Формирование структурной матрицы факторов: Добавить / Матрицу… (рис. 1.3).

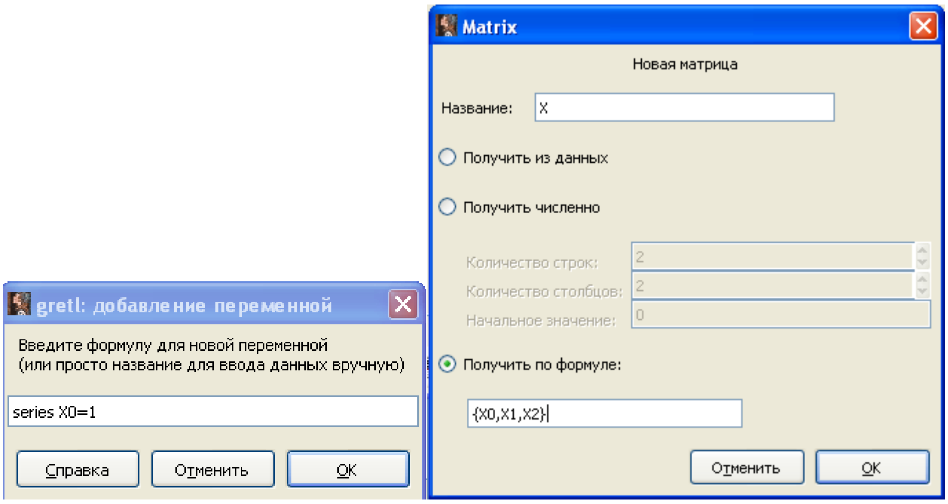


Рис. 2.3. Окна добавления новой переменной и новой матрицы X

5. Формирование единичной матрицы порядка 3: Добавить / Матрицу… (рис. 2.4, рис. 2.5).

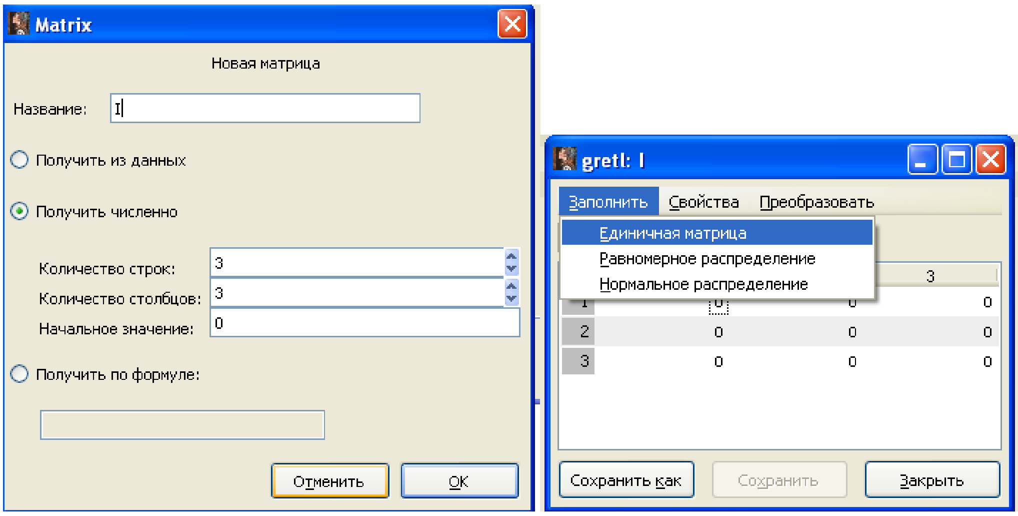


Рис. 2.4. Окно добавления единичной матрицы I

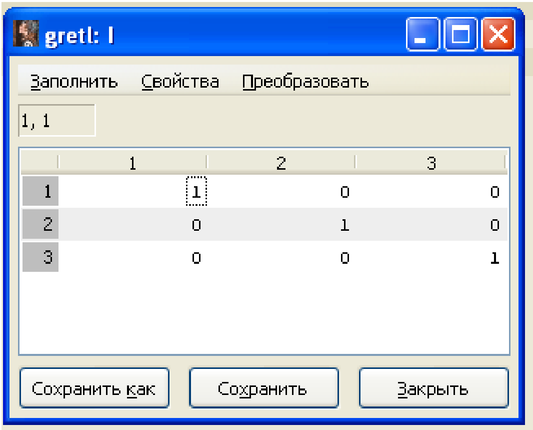


Рис. 2.5. Единичная матрица I

6. Ввод значений «гребня» λ для построения моделей: Вид / Сессия / Скалярные величины {кнопка «Просмотр сессии», в ней кнопка «+»}, (рис. 2.6).

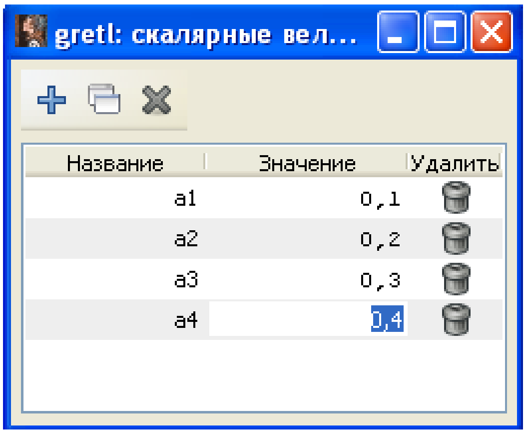


Рис. 2.6. Окно ввода скалярных величин

7. Нахождение обратной информационной матрицы для ридж-регрессии: Добавить / Матрицу…(рис. 2.7).

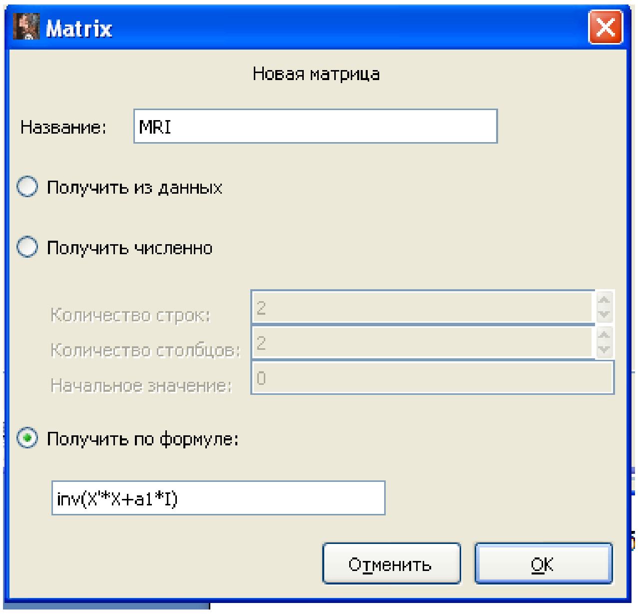


Рис. 2.7. Окно добавления единичной матрицы MRI

8. Нахождение вектора коэффициентов: Добавить / Матрицу…(рис. 2.8).

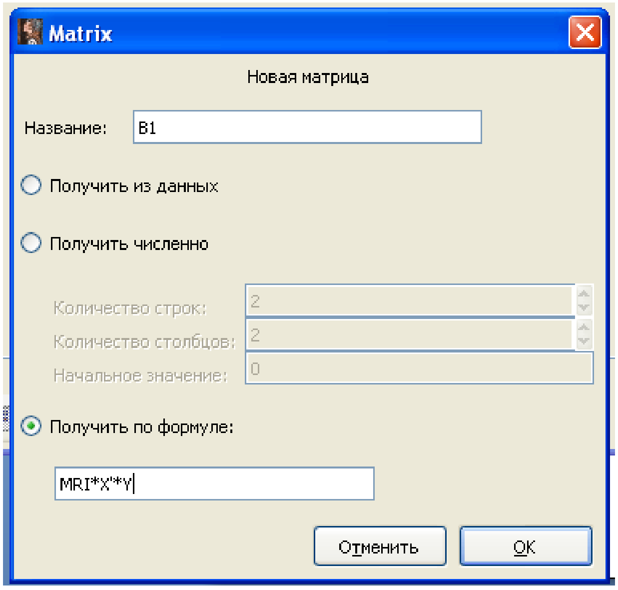


Рис. 2.8. Окно добавления вектора коэффициентов B1

9.Нахождение вектора расчетных значений: Добавить / Матрицу, (рис. 2.9).

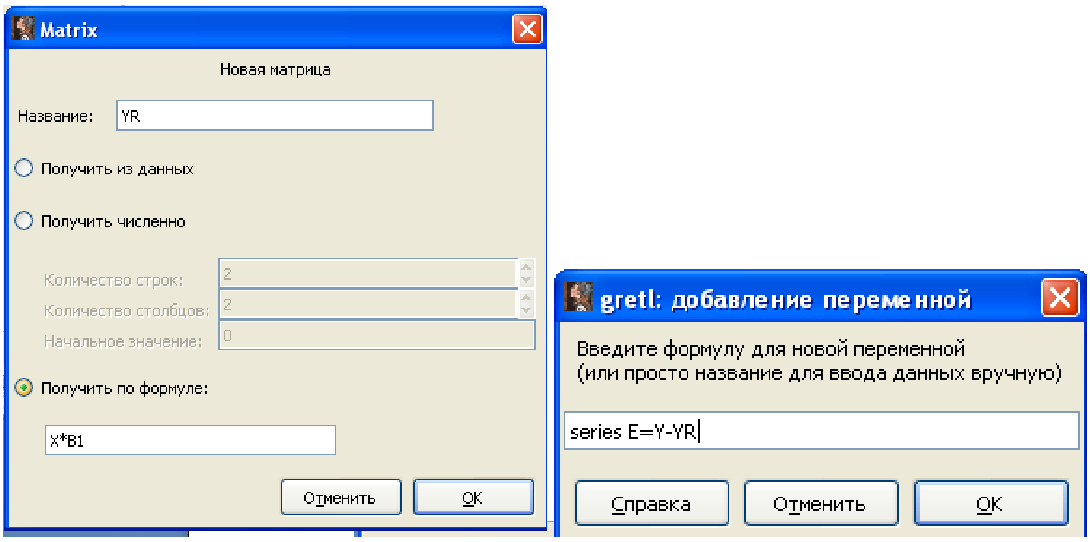


Рис. 2.9. Окно добавления вектора расчетных значений YR

10. Нахождение вектора остатков E: Добавить / Добавить новую переменную E…

11. Ввод скалярных величин «объем выборки», «количество факторов»: Вид / Сессия / Скалярные величины {кнопка «Просмотр сессии», в ней кнопка «+»}. Для нахождения стандартной ошибки регрессии *Se*: Вид / Сессия / Скалярные величины в столбец «Значение» надо ввести формулу: =sqrt(sum(E^2)/(n-p-1)), (рис. 2.11).

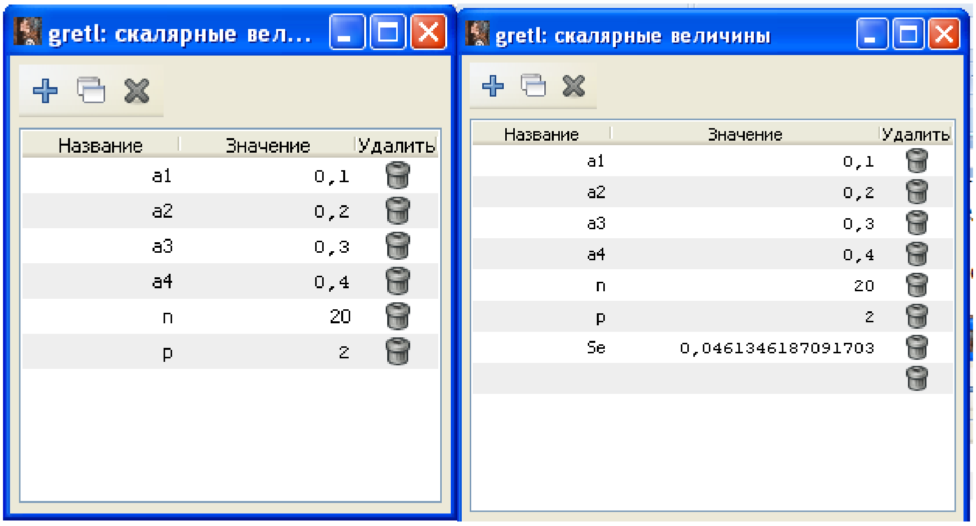


Рис. 2.11. Окно ввода скалярных величин

12. Определение единичного вектора-столбца VE размерностью 3X1 для нахождения вектора-столбца стандартных ошибок коэффициентов регрессии (рис. 2.12).

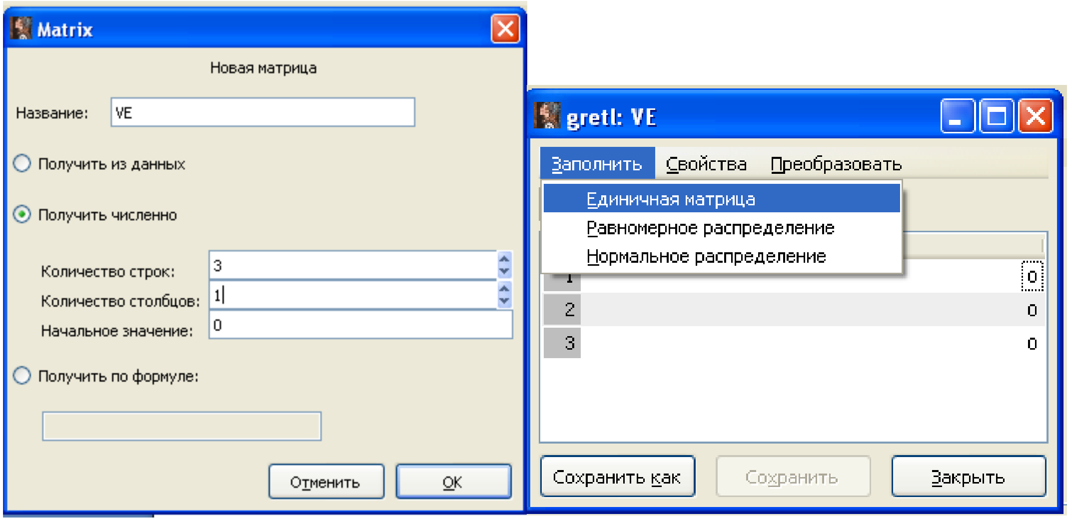


Рис. 2.12. Окно добавления единичного вектора-столбца VE

13. Нахождение вектора-столбца стандартных ошибок коэффициентов регрессии: Добавить / Матрицу…, (рис. 2.13).

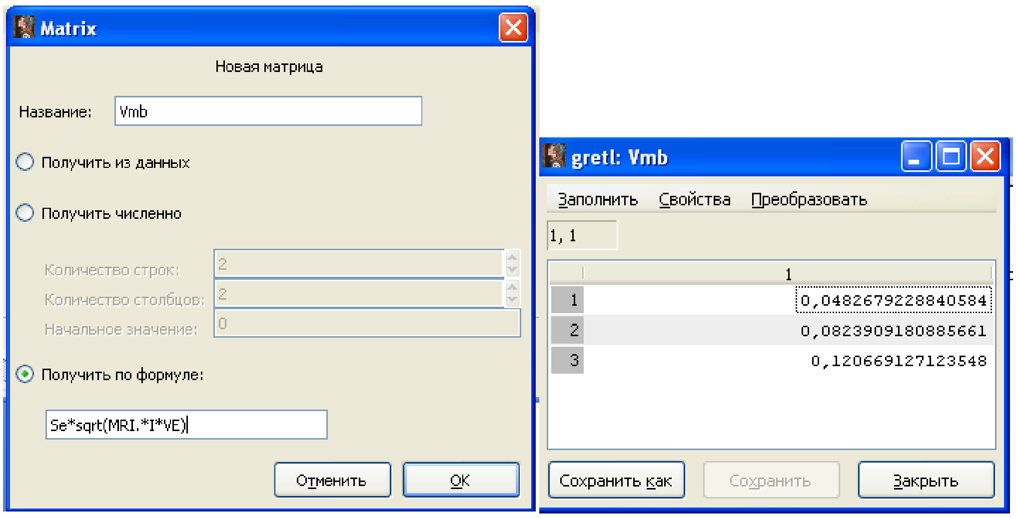


Рис. 2.13. Окно добавления вектора-столбца стандартных ошибок коэффициентов регрессии Vmb

14. Нахождение вектора наблюдаемых t-статистик коэффициентов регрессии: Добавить / Матрицу…, (рис. 2.14).

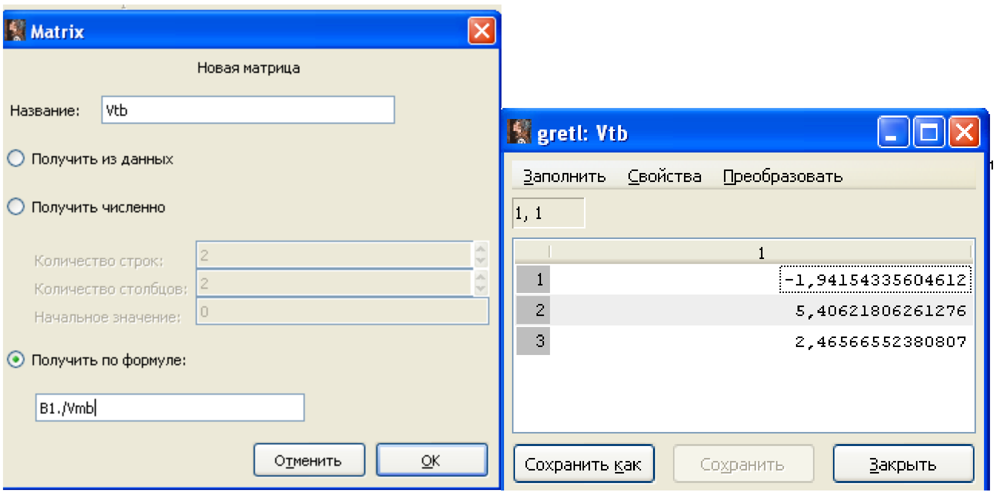


Рис. 2.14. Окно добавления вектора наблюдаемых t-статистик коэффициентов регрессии

15. Нахождение критического значения распределения Стьюдента: Инструменты / Критические значения…(рис. 2.15).

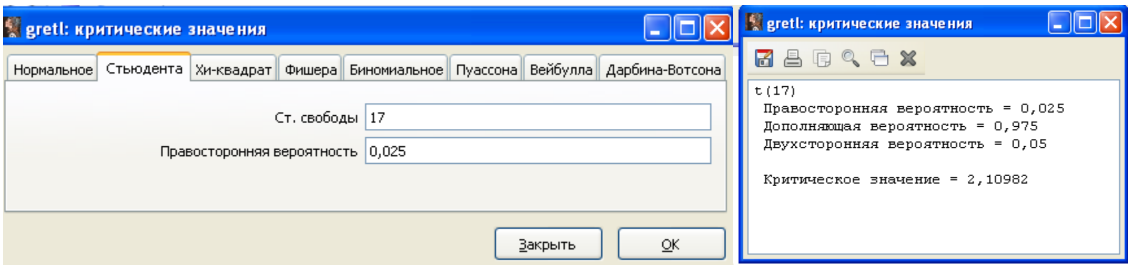


Рис. 2.15. Окно критических значений распределения Стьюдента

5,4>2,109 ; 2,46>2,109 – оба коэффициента регрессии статистически значимы.

16. Ввод скаляров «общая сумма квадратов отклонений», «регрессионная сумма квадратов отклонений», «коэффициент детерминации» : Вид / Сессия / Скалярные величины {«Просмотр сессии», кнопка «+»}, (рис. 2.16).

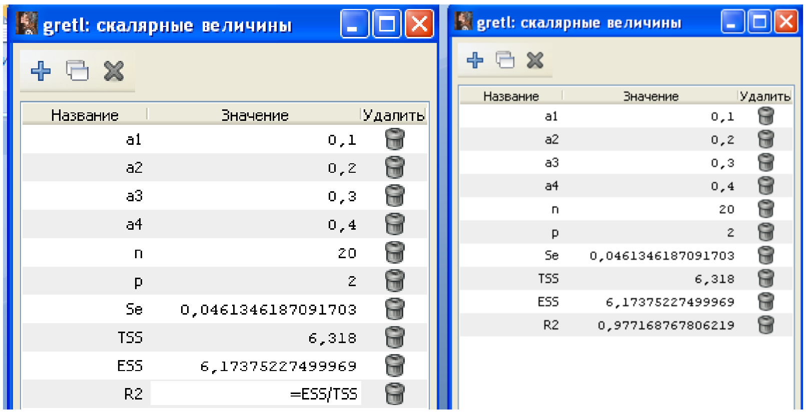


Рис. 2.16. Окно ввода скаляров

17. Итерации 7-16 необходимо повторить трижды для a2=0,2; a3=0,3; a4=0,4.

Затем результаты сравнить с таблицей 1.2 и выбрать наиболее качественную итерацию по критериям t-статистики, стандартной ошибки и R^2.

Таблица 2.2

Результаты гребневой регрессии

«Гребень»	Гребневые оценки коэффициентов регрессии	t-статистика	Se	R^2 скор-ректированный
λ_1=0,1	a= -0,09371426498569	-1,94154335604 61	0,046	0,977
	b1=0,445423269565654	5,40621806261276		
	b2=0,297529706536547	2,46566552380807		
λ_2=0,2	a=-0,074505753444208	-1,6052862624934	0,0468	0,96
	b1=0,441110265706663	7,43298857853048		
	b2=0,297199712434271	3,42940891064206		
λ_3=0,3	a=-0,058847639120987	-1,3504411531499	0,047	0,95
	b1=0,438142945837271	9,14657947108013		
	b2=0,296094964359754	4,24295330221462		
λ_4=0,4	a=-0,045841415839257	-1,0458283618239	0,048	0,943
	b1=0,435793886762127	9,96651875655972		
	b2=0,29497929953729	4,63998858321583		

Для выбора подходящего значения λ и наилучших гребневых оценок регрессии рекомендуется ориентироваться на минимум стандартной ошибки модели Se, максимум скорректированного коэффициента детерминации и абсолютных значений t-статистик. В данном задании исходя из минимума Se=0,046 наиболее подходящим является «гребень» λ_1=0,1.

3. Оценивание параметров линейной модели множественной регрессии в условиях мультиколлинеарности. Регрессия на главных компонентах

Расчетные формулы

Основная идея метода главных компонент заключается в замене сильно коррелированных переменных совокупностью новых переменных, между которыми корреляция отсутствует. При этом новые переменные являются линейными комбинациями исходных переменных:

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m, \\ z_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ z_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mm}x_m. \end{cases}$$

Переменные $z_1, z_2, \ldots, z_m$ называют главными компонентами. Максимально возможное число выделяемых компонент равно числу переменных. Метод главных компонент выстраивает линейные комбинации переменных в порядке убывания их влияния на совокупную дисперсию первоначальных данных. Иными словами, первая главная компонента является линейной комбинацией переменных, обладающей наиболее высокой дисперсией, вторая компонента – второй по величине дисперсией и так далее. Для каждой следующей компоненты дисперсия убывает, а последняя компонента будет иметь наименьшую дисперсию. Дисперсии, соответствующие компонентам, называются собственными значениями. Можно предполагать, что исходные переменные $x_1, x_2, \ldots, x_m$ уже стандартизированы, так что все переменные имеют нулевое математическое ожидание и единичную дисперсию. При этом матрица $V = X^T X$ является корреляционной матрицей для исходных данных.

Для первой главной компоненты $z_1 = \sum_{i=1}^{m} a_{1i} x_i = a_1^{\ T} \cdot x^T = x \cdot a_1$, где $a_1 = (a_{11}, a_{12}, ..., a_{1m})$, справедливы равенства $M(z_1)$ = $a_1^{\ T} \cdot M(x^T) = 0$; $D(z_1) = M(z_1^{\ 2}) = M(a_1^{\ T} x^T x a_1) = a_1^{\ T} M(x^T x) a_1 = a_1^{\ T} V a_1$.

Хорошо известно, что невырожденная корреляционная матрица $V = X^T X$ имеет m положительных собственных значений и m соответствующих им собственных векторов.

Пусть a_1 собственный вектор матрицы $V = X^T X$, а λ_1 соответствующее этому собственному вектору собственное значение, то есть $Va_1 = \lambda_1 a_1$. Умножая последнее равенство слева на $a_1^{\ T}$, получаем $a_1^{\ T} V a_1 = \lambda_1 a_1^{\ T} a_1$. Чтобы вектор a_1 однозначно определить, дополнительно потребуем, чтобы $a_1^{\ T} a_1 = 1$. Тогда $D(z_1) = a_1^{\ T} V a_1 = \lambda_1$ и проблема нахождения первой главной компоненты с максимальной дисперсией решается путем нахождения наибольшего собственного значения λ_1 и соответствующего ему собственного вектора a_1 корреляционной матрицы $V = X^T X$.

Рассуждая аналогично, находим вторую главную компоненту $z_2 = \sum_{i=1}^{m} a_{2i} x_i = a_2^{\ T} \cdot x^T = x \cdot a_2$, при условиях нормировки $a_2^{\ T} a_2 = 1$ и линейной независимости (ортогональности векторов) $a_1^{\ T} a_2 = 0$. Дисперсия второй главной компоненты z_2 будет равна второму по величине собственному значению λ_2 матрицы $V = X^T X$. Проверим, что главные компоненты z_1 и z_2 не коррелируют между собой. В самом деле

$$\begin{aligned} &\mathrm{cov}(z_1, z_2) = M(z_1 - M(z_1), z_2 - M(z_2)) = M(z_1, z_2) = M(a_1^{\ T} x x^T a_2) = \\ &= a_1^{\ T} M(x x^T) a_2 == a_1^{\ T} V a_2 = a_1^{\ T} \lambda_2 a_2 = \lambda_2 a_1^{\ T} a_2 = 0. \end{aligned}$$

Продолжая процесс построения, получаем систему главных компонент, не коррелирующих друг с другом, с дисперсиями равными собственным числам корреляционной матрицы V. Так как исходные переменные были сильно коррелированны, то матрица $V = X^T X$ плохо обусловлена, то есть ее определитель близок к нулю. С другой стороны можно показать, что определитель $V = \lambda_1 \cdot \lambda_2 \cdots \lambda_m$. Следовательно, можно ожидать, что одно или несколько последних собственных значений матрицы достаточно малы. Тогда, отбросив соответствующие главные компоненты, мы получаем возможность сократить размерность задачи, уменьшить число факторов в модели.

Задание 3.1. Компания «Альбион» осуществляет торговлю товарами бытовой химии и косметики на российском рынке. Ее успех определяется, в частности, человеческим фактором. С целью изучения его влияния на среднеквартальный объем продаж (млн. руб., y) через такие показатели, как затраты на поощрение персонала (млн. руб., x_1), численность работников фирмы, имеющих поощрения (тыс. чел, x_2), количество совершаемых покупок, (x_3, млн. шт.) была сформирована таблица исходных данных за 10 кварталов. Постройте регрессионную модель, отражающую зависимость объема продаж от указанных факторов.

Таблица 3.1

Объем продаж компании «Альбион»

Y	X1	X2	X3	Y	X1	X2	X3
26,2	1,1	1,1	1,2	33,6	1,8	1,8	1,9
25,9	1,4	1,5	1,1	34,2	1,9	1,8	2
32,5	1,7	1,8	2	34,4	2	2,1	2,1
30,3	1,7	1,7	1,8	35,5	2,3	2,4	2,5
31,7	1,8	1,9	1,8	36,5	2,5	2,5	2,4

Решение в среде Gretl

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

2. Построение регрессии по исходным данным: Модель / Метод наименьших квадратов… (рис. 3.1).

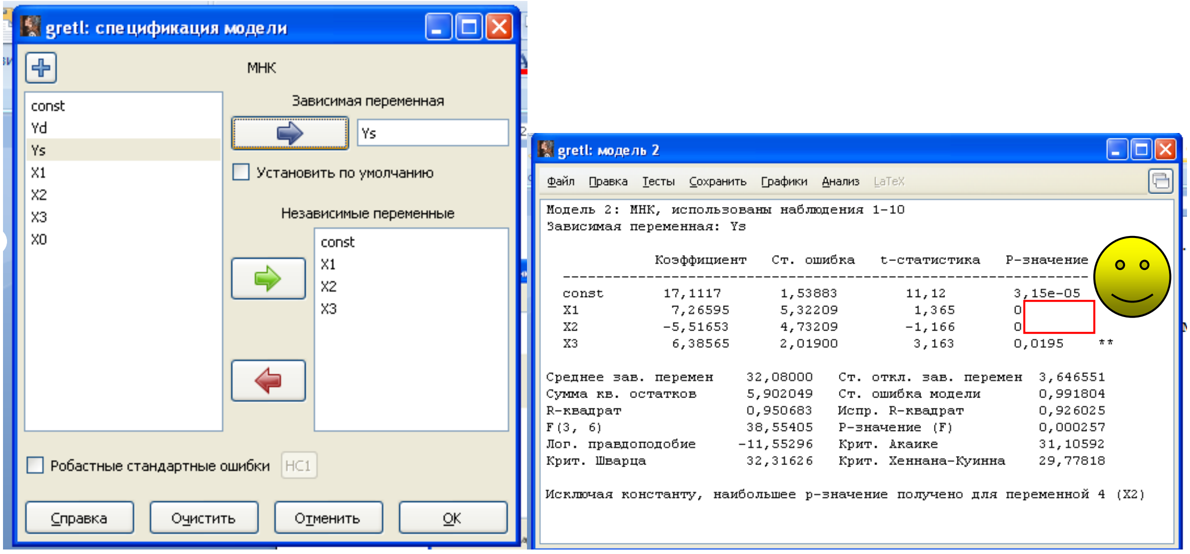


Рис. 3.1. Модель регрессии по исходным данным

Получаем: Y=17,11+7,27X1-5,52X2+6,39X3. R^2скорр.= 0,926. Модель значима по критерию Фишера, но коэффициенты регрессии при X_1,X_2 не значимы при, тем не менее, высоком коэффициенте детерминации. Такая ситуация весьма вероятна при мультиколлинеарности, и модель с коллинеарными факторами применять нельзя.

3. Проверка модели на мультиколлинеарность. Для выявления коллинеарных факторов построим матрицу линейных коэффициентов парной корреляции: Вид/ Корреляционная матрица…(рис. 3.2).

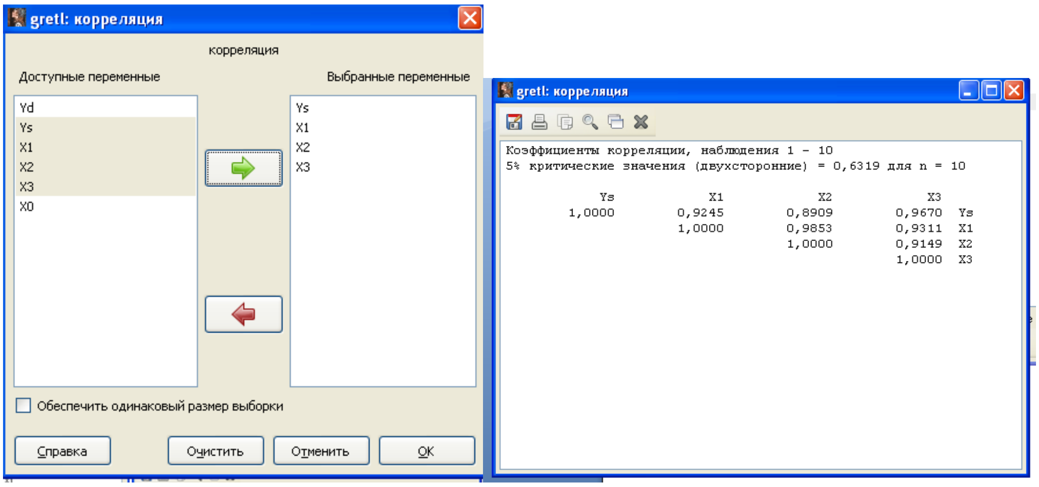


Рис. 3.2. Матрица линейных коэффициентов парной корреляции

Визуально обнаружены три пары коллинеарных факторов: X1 и X_3; X_1 и X_2; X_2 и X_3, линейные коэффициенты парной корреляции для каждой пары больше, чем 0,7.

Проведем тест на мультиколлинеарность методом инфляционных факторов: Тесты/Мультиколлинеарность… (рис. 3.3).

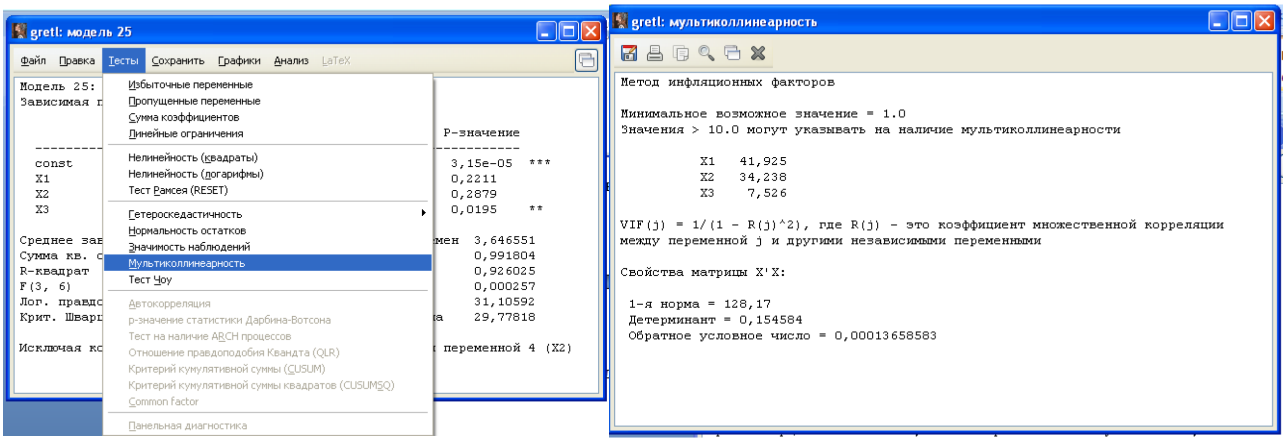


Рис. 3.3. Тест на мультиколлинеарность факторов

4. Попытаемся улучшить модель методом последовательного исключения избыточных переменных: Тесты/Избыточные переменные…(рис. 3.4).

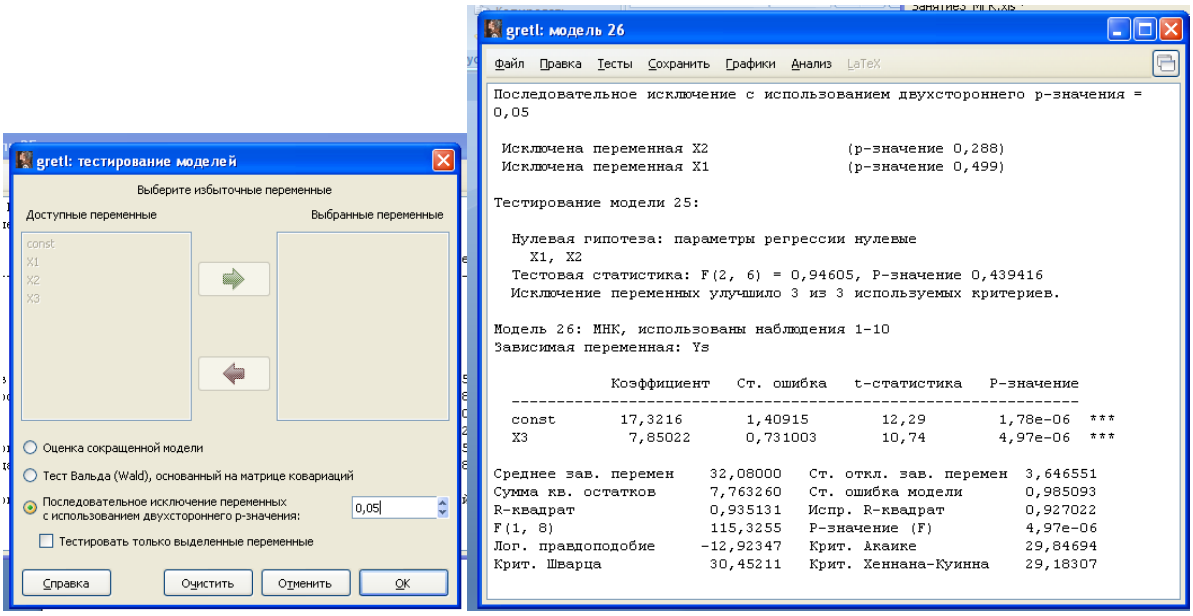


Рис. 3.4. Результаты теста на мультиколлинеарность факторов

Получаем: Y= 17,32+7,85*X_3, R^2=0,935, по тесту Фишера модель значима, по тесту Стьюдента параметры значимы на уровне α=0,01.

5. Допустим, что исследователь исходя из цели и проверяемой гипотезы принял решение, что в модели должны остаться все три фактора. Метод наименьших квадратов не позволяет строить такую модель. Применим регрессию на главных компонентах. Проведем z- стандартизацию переменных: $z_{ij} = \dfrac{x_{ij} - m_i}{\sigma_i}$. Введем скаляры (рис. 2.5): Вид / Сессия / Скаляры / Добавить… :

mX1=mean(X1); mX2=mean(X2); mX3=mean(X3); SX1=sd(X1); SX2=sd(X2); SX3=sd(X3).

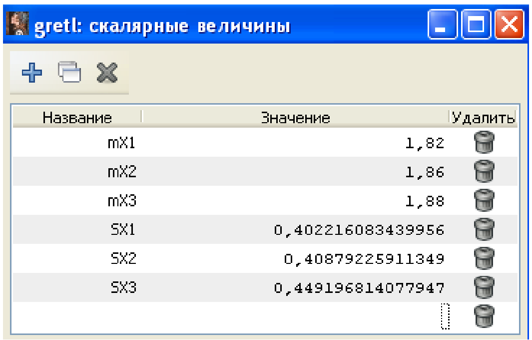


Рис. 3.5. Ввод скалярных величин

6. Находим стандартизованные переменные Z1,Z2,Z3: Добавить / Добавить новую переменную…(рис. 3.6).

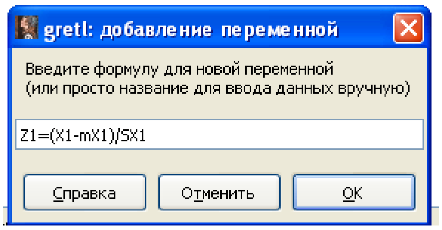


Рис. 3.6. Окно добавления новой переменной

7. Вычисляем главные компоненты: Вид / Главные компоненты / Использовать матрицу корреляций…(рис. 3.7).

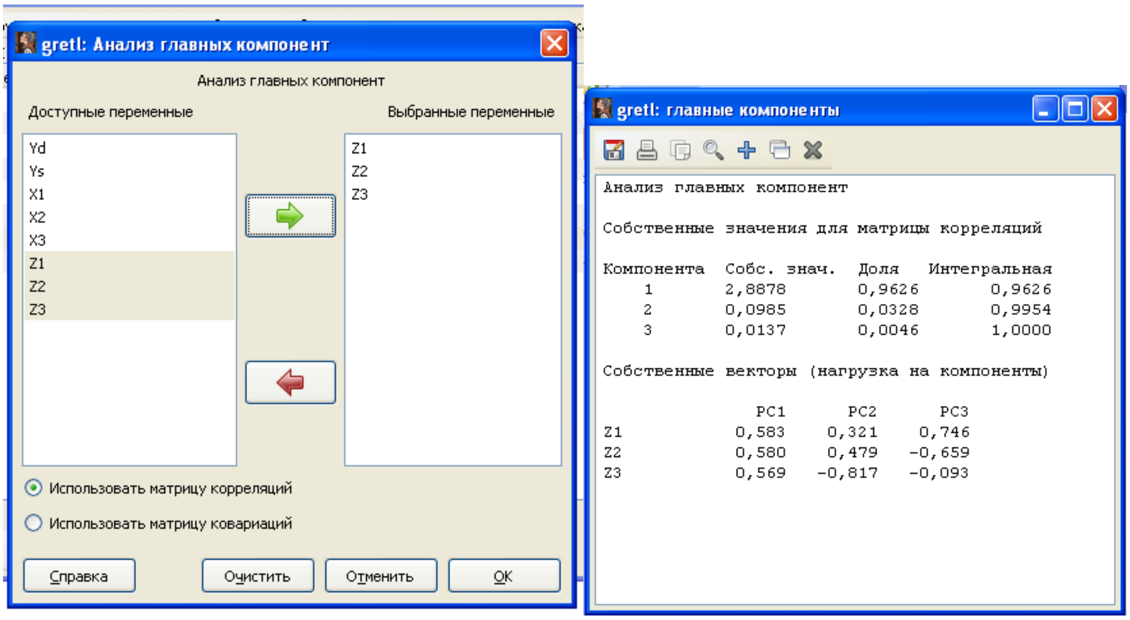


Рис. 3.7. Вычисление главных компонент

Выделено три компоненты, только первая имеет собственное значение (λ) больше единицы, объясняет 96,3% дисперсии зависимой переменной. Вторая и третья компоненты имеют собственные значения меньше единицы и объясняют вместе 3,7% дисперсии зависимой переменной. Сумма собственных значений составила 3: (λ1+λ2+λ3=3), представлены три стандартизованные переменные. Факторные нагрузки на компоненты – коэффициенты парной корреляции, для первой компоненты меньше 0,7 для всех переменных, для второй компоненты больше 0,7 (модуль) для переменной Z3, для третьей компоненты больше 0,7 для переменной Z1.

Запишем уравнение регрессии для первой компоненты:

$$PC1=0,583Z1+0,580Z2+0,569Z3.$$

Для сохранения главных компонент: кнопка «плюс» / Добавить в набор данных / Все компоненты / Ок.

8. Проводим корреляционный анализ главных компонент: Вид / Корреляционная матрица…(рис. 3.8).

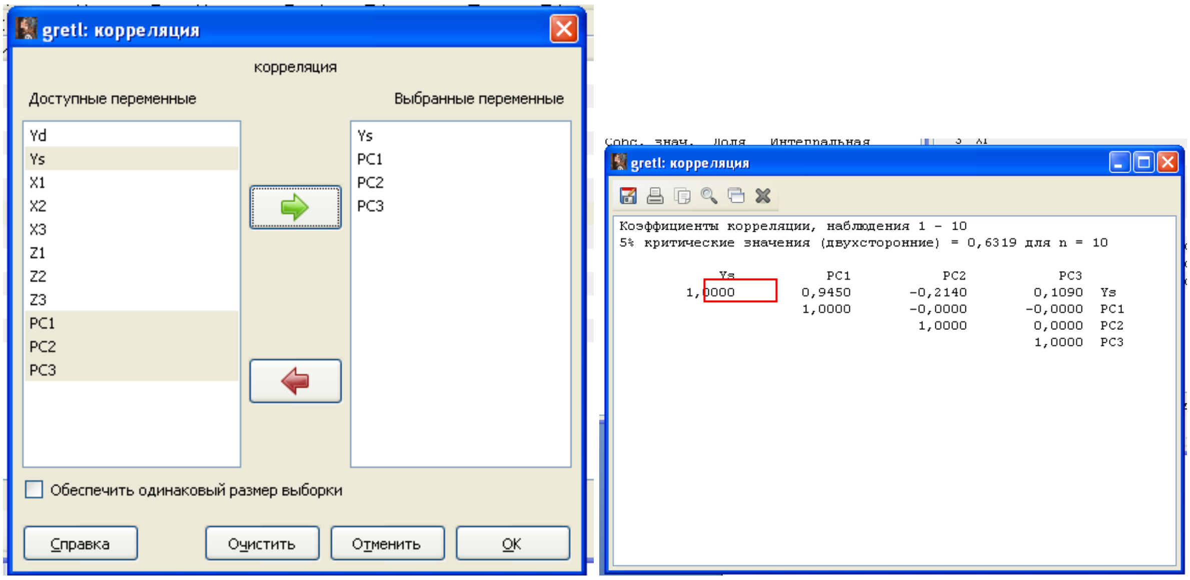


Рис. 3.8. Окно корреляционного анализа главных компонент

Главные компоненты между собой не коррелированы. Взаимосвязь с зависимой переменной высокая у первой главной компоненты, у второй и третьей – слабая. Значит, целесообразно строить регрессию на первую главную компоненту.

9. Построение регрессии на все главные компоненты и исключение избыточных переменных: Модель / Метод наименьших квадратов…(рис. 3.9).

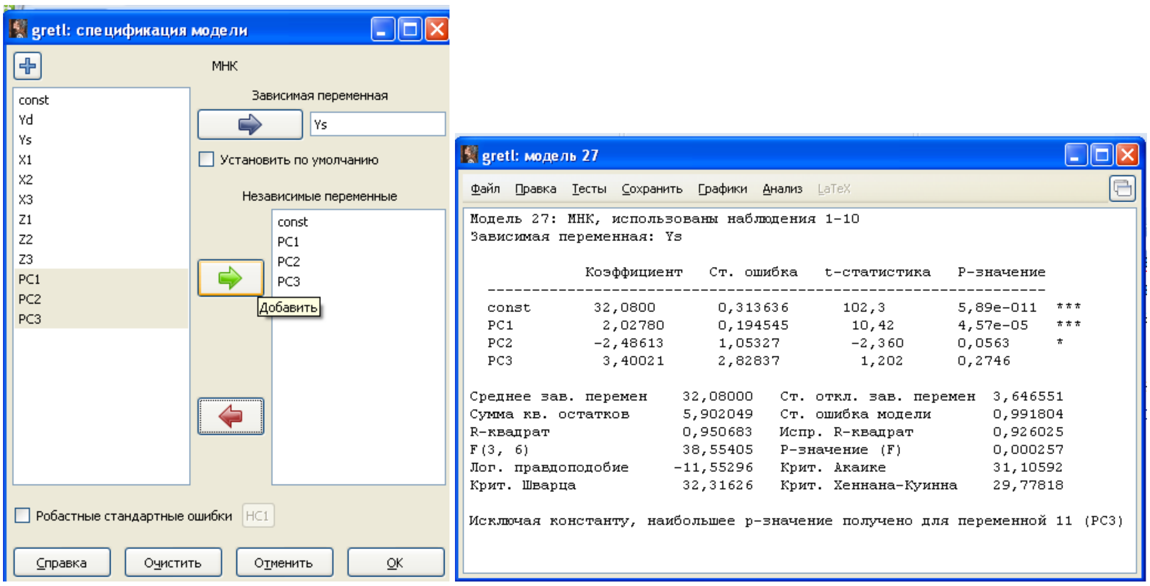
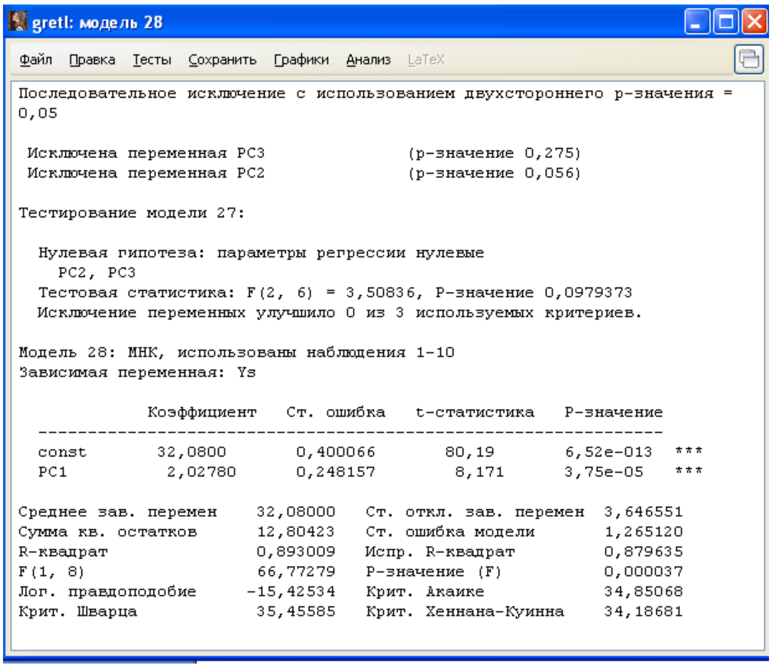


Рис. 3.9. Регрессия на главные компоненты

Получим: Y=32,08+2,03PC1-2,48PC2+3,4PC3. R^2скорр.=0,926. Модель по критерию Фишера значима, по критерию Стьюдента с надежностью 99% значим коэффициент регрессии при PC1, с надежностью 90% - коэффициент регрессии при PC2, коэффициент регрессии при PC3 не значим.

10. Усовершенствуем модель, исключая избыточные переменные: Тесты / Избыточные переменные…/ Последовательное исключение переменных (рис. 3.10).




```
Последовательное исключение с использованием двухстороннего p-значения =
0,05

 Исключена переменная PC3              (p-значение 0,275)
 Исключена переменная PC2              (p-значение 0,056)

Тестирование модели 27:

  Нулевая гипотеза: параметры регрессии нулевые
    PC2, PC3
  Тестовая статистика: F(2, 6) = 3,50836, P-значение 0,0979373
  Исключение переменных улучшило 0 из 3 используемых критериев.

Модель 28: МНК, использованы наблюдения 1-10
Зависимая переменная: Ys

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const       32,0800       0,400066       80,19       6,52e-013  ***
  PC1          2,02780      0,248157        8,171      3,75e-05   ***

Среднее зав. перемен   32,08000   Ст. откл. зав. перемен  3,646551
Сумма кв. остатков     12,80423   Ст. ошибка модели       1,265120
R-квадрат              0,893009   Испр. R-квадрат         0,879635
F(1, 8)                66,77279   P-значение (F)          0,000037
Лог. правдоподобие    -15,42534   Крит. Акаике            34,85068
Крит. Шварца           35,45585   Крит. Хеннана-Куинна    34,18681
```

Рис. 3.10. Регрессия после исключения избыточных переменных

Получаем: Y=32,08+2,028PC1. Однако затруднительна содержательная интерпретация модели применительно к главным компонентам. Поэтому целесообразно перейти к модели, содержащей исходные факторы, которая легко поддается экономической интерпретации.

11. Подставим в уравнение регрессии в пункте 2.10 уравнение для первой главной компоненты (пункт 2.7)

Ys=32,08+2,028*(0,583Z1+0,580Z2+0,569Z3).

В буквенном выражении:

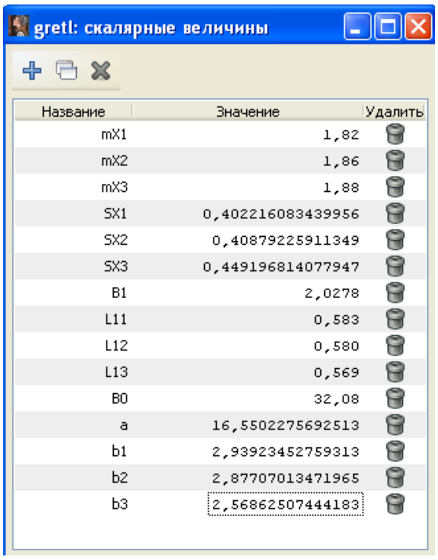
$$Ys = \beta_0 + \beta_1 l_{11} Z_1 + \beta_1 l_{12} Z_2 + \beta_1 l_{13} Z_3 = \beta_0 + \beta_1 l_{11} \frac{x_1 - \bar{x}_1}{\sigma_{x1}} + \beta_1 l_{12} \frac{x_2 - \bar{x}_2}{\sigma_{x2}} + \beta_1 l_{13} \frac{x_3 - \bar{x}_3}{\sigma_{x3}} =$$

$$= \beta_0 - \beta_1 \left(\frac{l_{11}\bar{x}_1}{\sigma_{x1}} + \frac{l_{12}\bar{x}_2}{\sigma_{x2}} + \frac{l_{13}\bar{x}_3}{\sigma_{x3}}\right) + \frac{\beta_1 l_{11}}{\sigma_{x1}} x_1 + \frac{\beta_1 l_{12}}{\sigma_{x2}} x_2 + \frac{\beta_1 l_{13}}{\sigma_{x3}} x_3 = a^* + b_1^* \cdot x_1 + b_2^* \cdot x_2 + b_3^* \cdot x_3.$$

Введем скаляры: Вид/Сессия/Скаляры/Добавить… : B0=32,08; B1=2,0278; L11=0,583; L12=0,580; L13=0,569.

12.Определим коэффициенты исходного уравнения (a, $b^*{}_1$,$b^*{}_2$,$b^*{}_3$) согласно формулам через ввод скаляров: Вид / Сессия / Скаляры / Добавить…(рис. 3.12)

$$a* = \beta_0 - \beta_1 \left(\frac{l_{11}\bar{x}_1}{\sigma_{x1}} + \frac{l_{12}\bar{x}_2}{\sigma_{x2}} + \frac{l_{13}\bar{x}_3}{\sigma_{x3}}\right);\ b^*{}_1 = \frac{\beta_1 l_{11}}{\sigma_{x1}}; b^*{}_2 = \frac{\beta_1 l_{12}}{\sigma_{x2}}; b^*{}_3 = \frac{\beta_1 l_{13}}{\sigma_{x3}}.$$



gretl: скалярные величины

Название	Значение	Удалить
mX1	1,82	
mX2	1,86	
mX3	1,88	
SX1	0,402216083439956	
SX2	0,40879225911349	
SX3	0,449196814077947	
B1	2,0278	
L11	0,583	
L12	0,580	
L13	0,569	
B0	32,08	
a	16,5502275692513	
b1	2,93923452759313	
b2	2,87707013471965	
b3	2,56862507444183	

Рис. 3.12. Окно ввода скалярных величин

Получим: Y=16,550+2,94X1+2,88X2+2,57X3.

С увеличением затрат на поощрение персонала на 1 млн. руб. объем продаж увеличивается в среднем на 2,94 млн. руб. при неизменных прочих факторах. С увеличением численности сотрудников, имеющих поощрения, на 1 тыс. человек, объем продаж увеличивается в среднем на 2,88 млн. руб. при неизменных прочих факторах. С увеличением количества покупок на 1 млн. штук объем продаж увеличивается в среднем на 2,57 млн. руб. при неизменных прочих факторах.

4. Выбор спецификации регрессионной модели. Нелинейный МНК

Расчетные формулы

При эконометрическом моделировании достаточно часто приходится принимать решение в пользу нелинейных моделей. В этом случае система нормальных уравнений для оценивания параметров модели методом наименьших квадратов оказывается нелинейной. Эта система решается нелинейными итерационными процедурами нелинейной оптимизации. В итоге модель оценивается нелинейным МНК. При нелинейном МНК имеем систему n нелинейных уравнений

$$\mathbf{J}' \cdot \mathbf{d} = 0 \qquad (4.1)$$

где J – прямоугольная (m×n) матрица Якоби с элементами

$$J_{ij} = \frac{\partial F(x_i; p_1, p_2, \ldots, p_n)}{\partial p_j} \qquad (4.2)$$

Представим метод численного решения системы (4.1), основанный на той же идее, что и в методе Ньютона. Предположим, что имеется начальное приближение $p_1^0, p_2^0, \ldots, p_n^0$ для искомых параметров. Разложим элементы вектора ***d*** в ряд Тейлора в окрестности начального приближения и оставим в этом разложении только линейные члены. Результат можно записать в матричной форме следующим образом:

$$\mathbf{d}(\mathbf{p}) = \mathbf{d}(\mathbf{p}_0) + \mathbf{J}(\mathbf{p}_0) \cdot (\mathbf{p} - \mathbf{p}_0) \qquad (4.3)$$

Здесь ***p*** обозначает совокупность n параметров $\boldsymbol{p_1,..,p_n}$; мы можем рассматривать ***p*** как n-мерный вектор. $\boldsymbol{p_0}$ – вектор, соответствующий начальному приближению. Запись ***d(p)*** означает, что элементы вектора d вычислены при значениях параметров, представленных вектором p, а $\boldsymbol{J(p_0)}$ – матрица Якоби, вычисленная при начальном приближении $\boldsymbol{p_0}$. Подставляя (3.3) в уравнение (3.1), получим:

$$\mathbf{J}'(\mathbf{p})\cdot\left[\mathbf{d}(\mathbf{p}_0)+\mathbf{J}(\mathbf{p}_0)(\mathbf{p}-\mathbf{p}_0)\right]=0 \quad (4.4)$$

В уравнении (6) фигурирует матрица Якоби, вычисленная в точке минимума, то есть $\boldsymbol{J(p)}$ и $\boldsymbol{J(p_0)}$, вообще говоря, различаются. Если бы не это обстоятельство, из (3.4) можно было бы найти вектор поправок $\boldsymbol{Dp=p\text{-}p_0}$ к начальному приближению. Заметим, что уравнение (4.4), в отличие от (4.1), не является точным, поскольку в нем использовано приближенное соотношение (4.3), полученное путем отбрасывания членов высших порядков в формуле Тейлора. Вероятно, степень точности уравнения (4.4) не ухудшится от введения еще одного приближения, если только дополнительная погрешность имеет величину того же порядка, что и погрешность (4.3). Исходя из этих соображений, будем считать, что $\boldsymbol{J(p)}$ и $\boldsymbol{J(p_0)}$ совпадают. Можно показать, что вносимая при этом в уравнение (4.4) ошибка соответствует членам второго и более высоких порядков относительно разности $\boldsymbol{p\text{-}p_0}$, то есть такое приближение действительно находится на уровне точности уравнения (4.3).

Для проверки правильности спецификации модели используется тест Рамсея (RESET-тест). Он позволяет определить, помогает ли нелинейная комбинация оцененного (расчетного) значения зависимой переменной лучше объяснить изменения самой зависимой переменной. Если качество объяснения при этом улучшается, значит, модель специфицирована неправильно, в ней есть системные недочеты, нужна корректировка спецификации. Нулевая гипотеза о совместной незначимости коэффициентов при вновь добавленных регрессорах в нелинейной комбинации, то есть качество модели не улучшилось, спецификация корректна:

$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_l = 0$

где $Y = a + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \alpha_1 \hat{y}^2 + \alpha_2 \hat{y}^3 + \alpha_l \hat{y}^{l+1} + \varepsilon$

Для проверки нулевой гипотезы применяется статистика Фишера:

$$F = \frac{(RSS_R - RSS_{UR})/l}{RSS_{UR}/(n-k)}$$

где RSS_R – сумма квадратов остатков для модели без оцененных зависимых переменных $\hat{y}^{l+1}$;

RSS_{UR} – сумма квадратов остатков для модели с оцененными зависимыми переменными в качестве регрессоров;

l – количество регрессоров, которые являются степенями прогнозного значения зависимой переменной.

Задание 4.1. Провести регрессионный анализ данных по заработной плате (Y, долл.) и возрасту (X, лет) 20 работников совместного предприятия с целью бюджетирования затрат на оплату труда.

Таблица 4.1

Данные о заработной плате и возрасте работников предприятия

Зар. плата	300	400	300	320	200	350	350	400	380	400
Возраст	29	40	36	31	23	45	38	40	50	47
Зар.плата	250	350	200	400	220	320	390	360	260	250
Возраст	28	30	25	48	30	40	40	38	29	25

Решение в среде Gretl

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

2. Визуальный анализ поля корреляции: Вид / График / Разброс X-Y (рис. 4.1).

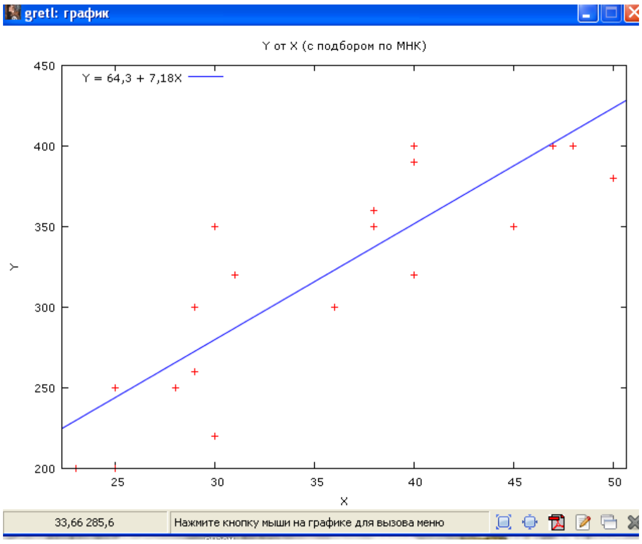
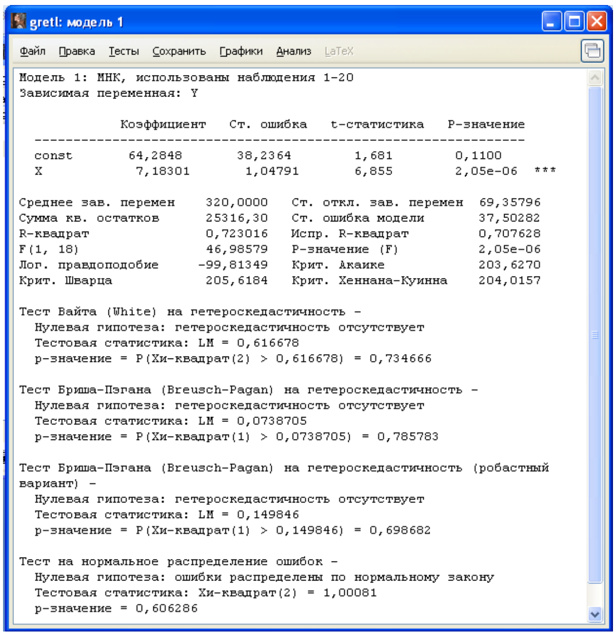


Рис. 4.1. Поле корреляции заработной платы и возраста работников

Можно предположить наличие нелинейной (степенной) зависимости между переменными. Проверим это соответствующими тестами, реализованными в Gretl. Реализация тестов требует сравнения с линейной моделью регрессии. Поэтому предварительно построим линейную модель регрессии.

3. Построение линейной модели регрессии: Модель / Метод наименьших квадратов…(рис. 4.2).



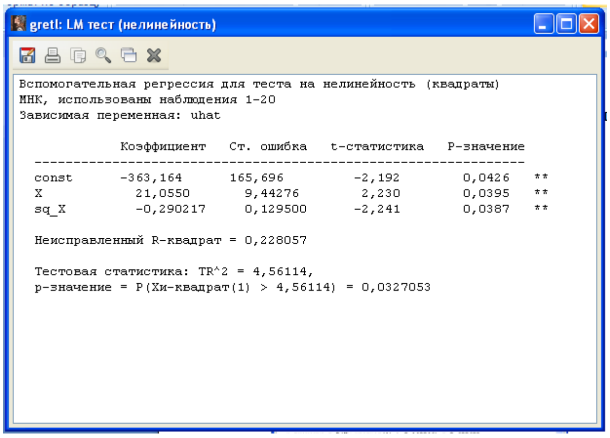

```
Модель 1: МНК, использованы наблюдения 1-20
Зависимая переменная: Y

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        64,2848      38,2364        1,681        0,1100
  X             7,18301      1,04791       6,855        2,05e-06  ***

Среднее зав. перемен    320,0000   Ст. откл. зав. перемен  69,35796
Сумма кв. остатков      25316,30   Ст. ошибка модели       37,50282
R-квадрат               0,723016   Испр. R-квадрат         0,707628
F(1, 18)                46,98579   P-значение (F)          2,05e-06
Лог. правдоподобие     -99,81349   Крит. Акаике            203,6270
Крит. Шварца            205,6184   Крит. Хеннана-Куинна    204,0157

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 0,616678
  p-значение = P(Хи-квадрат(2) > 0,616678) = 0,734666

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 0,0738705
  p-значение = P(Хи-квадрат(1) > 0,0738705) = 0,785783

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 0,149846
  p-значение = P(Хи-квадрат(1) > 0,149846) = 0,698682

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 1,00081
  p-значение = 0,606286
```

Рис. 4.2. Линейная модель парной регрессии

Y=64,28+7,18X+ε, R^2 = 0,723, Se=37,51, тесты Фишера и Стьюдента подтверждают статистическую значимость модели и параметров с вероятностью 99%. Во всех тестах подтверждена гомоскедастичность остатков (p>0,05). Остатки имеют нормальный закон распределения (p>0,05).

4. Для проверки обоснованности применения нелинейных моделей выполним тесты на нелинейность. Проверим обоснованность применения степенной модели. Используем линейную модель регрессии (рис. 4.2): Y=64,28+7,18X+ε.

Проведем тесты: В окне модели: Тесты / Нелинейность (квадраты), (рис. 4.3):

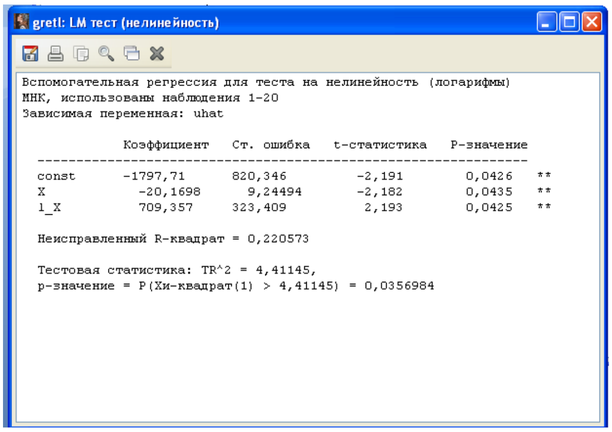


```
gretl: LM тест (нелинейность)

Вспомогательная регрессия для теста на нелинейность (квадраты)
МНК, использованы наблюдения 1-20
Зависимая переменная: uhat

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const      -363,164      165,696        -2,192        0,0426    **
  X            21,0550       9,44276       2,230        0,0395    **
  sq_X         -0,290217     0,129500     -2,241        0,0387    **

  Неисправленный R-квадрат = 0,228057

  Тестовая статистика: TR^2 = 4,56114,
  p-значение = P(Хи-квадрат(1) > 4,56114) = 0,0327053
```

Рис. 4.3. Результат теста на нелинейность (квадраты)

P- значение <0,05, значит нулевая гипотеза о корректности линейной модели отклоняется.

В окне модели: Тесты / Нелинейность (логарифмы), (рис. 4.4):



```
gretl: LM тест (нелинейность)

Вспомогательная регрессия для теста на нелинейность (логарифмы)
МНК, использованы наблюдения 1-20
Зависимая переменная: uhat

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const      -1797,71      820,346        -2,191        0,0426    **
  X            -20,1698      9,24494      -2,182        0,0435    **
  l_X          709,357     323,409         2,193        0,0425    **

  Неисправленный R-квадрат = 0,220573

  Тестовая статистика: TR^2 = 4,41145,
  p-значение = P(Хи-квадрат(1) > 4,41145) = 0,0356984
```

Рис. 4.4. Результат теста на нелинейность (логарифмы)

Р-значение <0,05, значит нулевая гипотеза о корректности линейной модели отклоняется.

В окне модели: Тесты / Тест Рамсея (Reset) (рис. 4.5):

$$Y_i = a_0 + a_1 X + a_2 \hat{Y}^2 + a_3 \hat{Y}^3$$

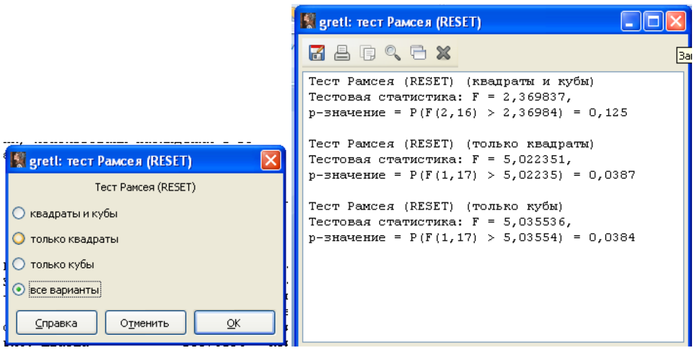


Рис. 4.5. Результат теста Рамсея на нелинейность

По тесту Рамсея для случая «квадраты и кубы» нулевая гипотеза о линейности модели принимается (p>0,05), для случаев «только квадраты», «только кубы» нулевая гипотеза о линейности модели отклоняется (p<0,05).

Итоговое решение в пользу нелинейной модели. Однако тесты непосредственно не указывают на наиболее подходящую функциональную форму модели. Поэтому найдем ее перебором ряда нелинейных моделей.

5. Построение степенной модели

5.1. Первый способ: линеаризация модели $Y=a*x^b$ через логарифмирование: lnY=a+blnX. Добавим переменные: Добавить / Добавить новые переменные (рис. 4.6).

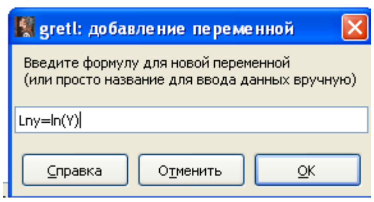
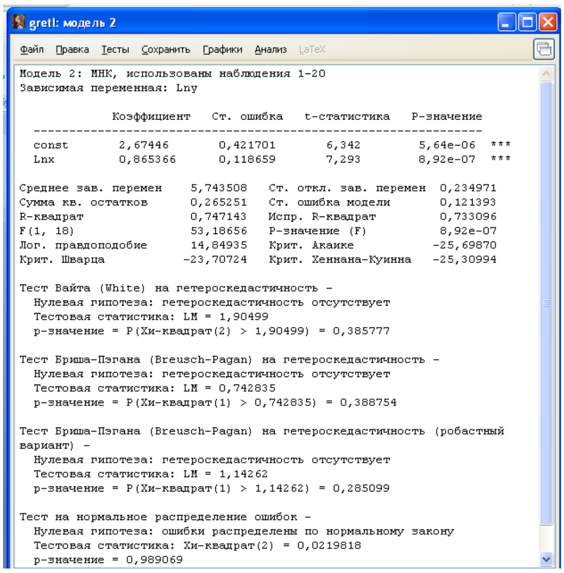


Рис. 4.6. Окно добавления новой переменной

Построение линейной модели регрессии по переменным lnY,lnX: Модель/Метод наименьших квадратов…(рис. 4.7).



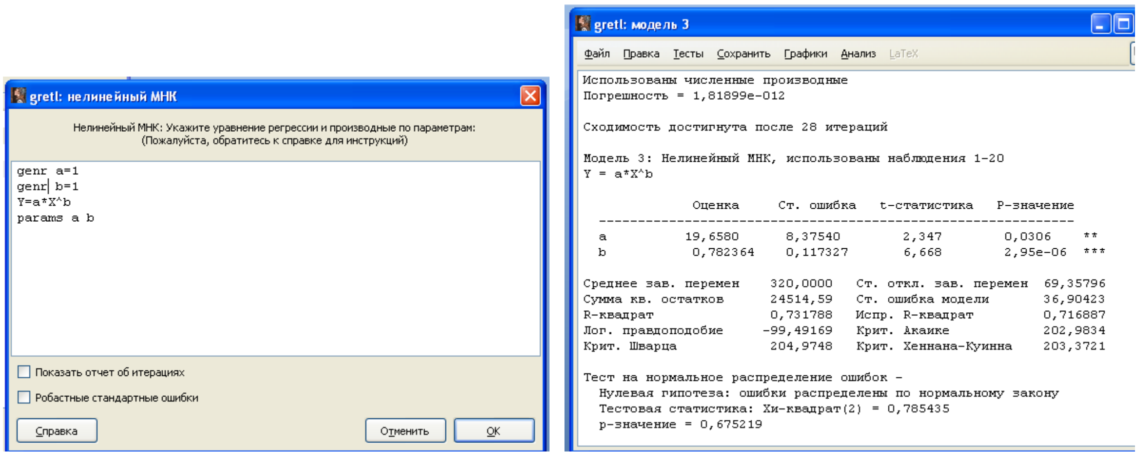

```
Модель 2: МНК, использованы наблюдения 1-20
Зависимая переменная: Lny

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const       2,67446      0,421701       6,342       5,64e-06  ***
  Lnx         0,865366     0,118659       7,293       8,92e-07  ***

Среднее зав. перемен    5,743508   Ст. откл. зав. перемен  0,234971
Сумма кв. остатков      0,265251   Ст. ошибка модели       0,121393
R-квадрат               0,747143   Испр. R-квадрат         0,733096
F(1, 18)                53,18656   Р-значение (F)          8,92e-07
Лог. правдоподобие      14,84935   Крит. Акаике           -25,69870
Крит. Шварца           -23,70724   Крит. Хеннана-Куинна   -25,30994

Тест Вайта (White) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 1,90499
  p-значение = P(Хи-квадрат(2) > 1,90499) = 0,385777

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 0,742835
  p-значение = P(Хи-квадрат(1) > 0,742835) = 0,388754

Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность (робастный
вариант) -
  Нулевая гипотеза: гетероскедастичность отсутствует
  Тестовая статистика: LM = 1,14262
  p-значение = P(Хи-квадрат(1) > 1,14262) = 0,285099

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 0,0219818
  p-значение = 0,989069
```

Рис. 4.7. Регрессия по переменным lnY, lnX

LnY=2,67+0,865lnX+ε, R^2 = 0,733, Se=0,12, тесты Фишера и Стьюдента подтверждают статистическую значимость модели и параметров с вероятностью 99%. Во всех тестах подтверждена гомоскедастичность остатков (p>0,05). (p<0,05). В окне ввода скаляров определим свободный коэффициент a=exp(2,67)). Y=14,44*x0,865*ε

5.2. Второй способ: путем использования нелинейного МНК. Построение модели регрессии: Модели / Нелинейные модели / Нелинейный МНК

Вначале применим численное дифференцирование (рис. 4.8):




```
genr a=1
genr b=1
Y=a*X^b
params a b
```

```
Использованы численные производные
Погрешность = 1,81899e-012

Сходимость достигнута после 28 итераций

Модель 3: Нелинейный МНК, использованы наблюдения 1-20
Y = a*X^b

             Оценка     Ст. ошибка   t-статистика   P-значение
  ------------------------------------------------------------
  a         19,6580      8,37540        2,347        0,0306    **
  b          0,782364    0,117327       6,668        2,95e-06  ***

Среднее зав. перемен   320,0000   Ст. откл. зав. перемен  69,35796
Сумма кв. остатков     24514,59   Ст. ошибка модели       36,90423
R-квадрат              0,731788   Испр. R-квадрат         0,716887
Лог. правдоподобие    -99,49169   Крит. Акаике            202,9834
Крит. Шварца           204,9748   Крит. Хеннана-Куинна    203,3721

Тест на нормальное распределение ошибок -
  Нулевая гипотеза: ошибки распределены по нормальному закону
  Тестовая статистика: Хи-квадрат(2) = 0,785435
  p-значение = 0,675219
```

Рис. 4.8. Нелинейный МНК путем численного дифференцирования

Запишем модель нелинейной регрессии: Y=19,66*x0,782*ε, R^2=0,72, Se=36,9. Остатки имеют нормальный закон распределения (p>0,05). Зададим начальные параметры более точно, применим аналитическое дифференцирование, пользуясь контекстным меню - добавить производную (рис. 4.9).

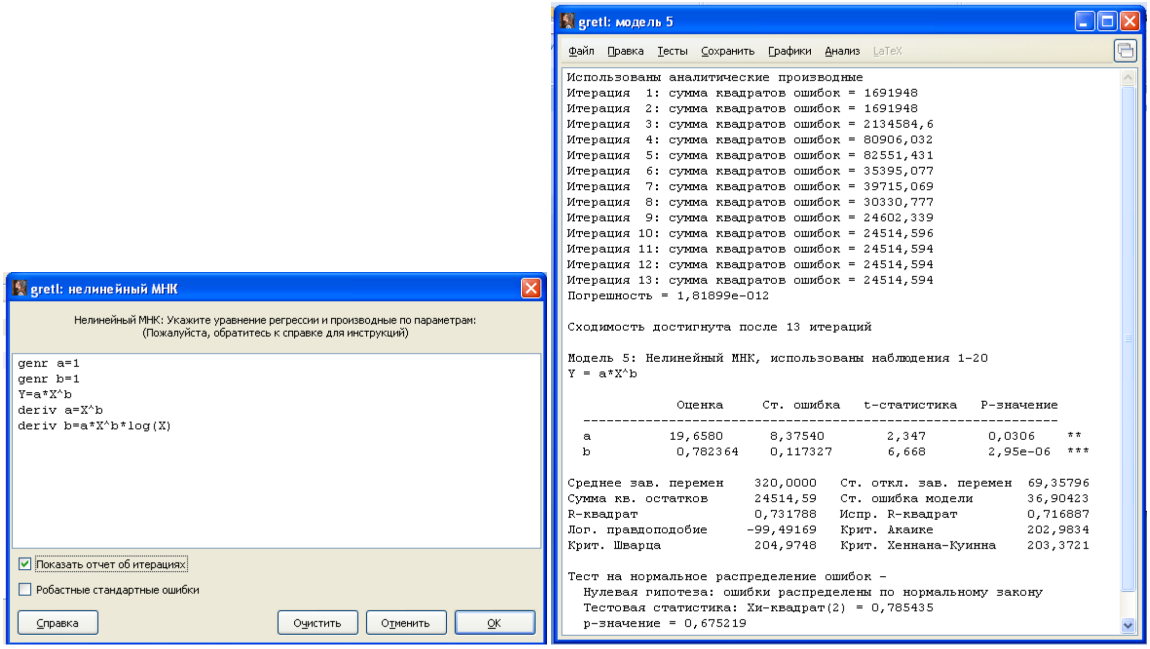


Рис. 4.9. Нелинейный МНК путем аналитического дифференцирования

Запишем степенную модель регрессии: Y=19,66*x^{0,782}*ε, R^2=0,72, Se=36,9. Остатки имеют нормальный закон распределения (p>0,05). Применим как начальные параметры МНК-оценки a=15,504; b=0,865 (рис. 4.10).

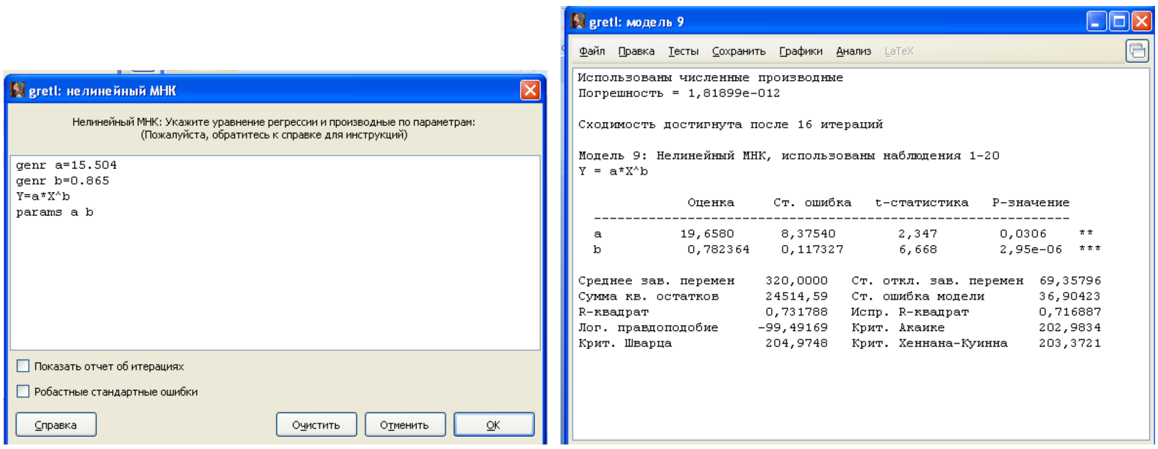


Рис. 4.10. Нелинейный МНК путем аналитического дифференцирования по МНК-оценкам.

$Y=19,66*x^{0,782}*\varepsilon$, R^2=0,72, Se=36,9. Остатки имеют нормальный закон распределения (p>0,05).

5.3. Построение экспоненциальной модели $Y=a*e^{bx}$

Путем численного дифференцирования (рис. 4.11):

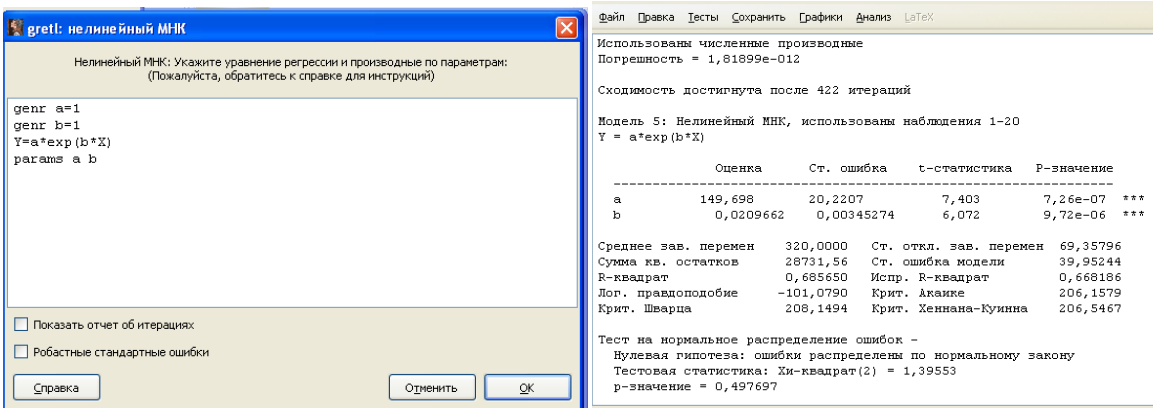


Рис. 4.11. Нелинейный МНК путем численного дифференцирования (экспонента)

Запишем степенную модель регрессии: $Y=149,698*e^{0,021*x}*\varepsilon$, R^2=0,67, Se=39,95. Остатки имеют нормальный закон распределения (p>0,05).

Путем грубых оценок начальных параметров: a=100, b=0 (рис. 4.12):

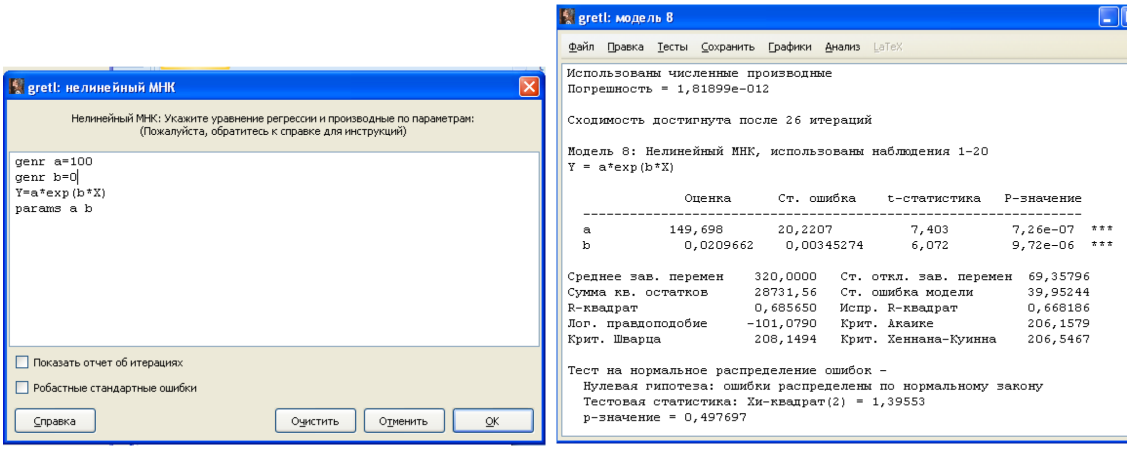


Рис. 4.12. Нелинейный МНК путем численного дифференцирования (экспонента, a=100, b=0)

Получим степенную модель регрессии: $Y=149,698*e^{0,021*x}*\varepsilon$, R^2=0,67, Se=39,95. Остатки имеют нормальный закон распределения (p>0,05).

5.4. Построение полулогарифмической модели Y=a+b*lnX (рис. 4.13).

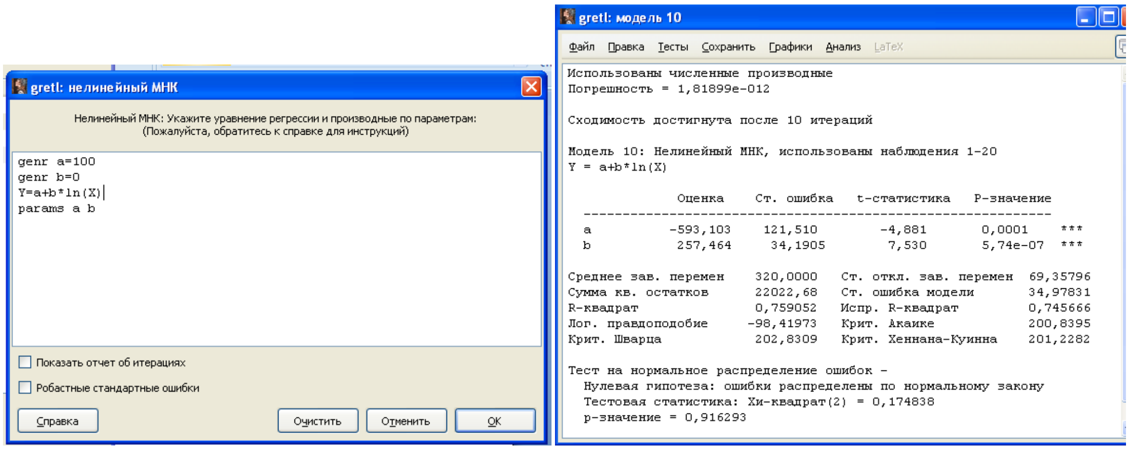


Рис. 4.13. Нелинейный МНК путем численного дифференцирования (полулогарифмическая модель)

Получим полулогарифмическую модель регрессии: Y=-593,103+257,464 lnX+ε, R^2=0,75; Se=14,98. Остатки имеют нормальный закон распределения (p>0,05).

5.5. Построение обратной функции Y=1/(a+b*X) (рис. 4.14):

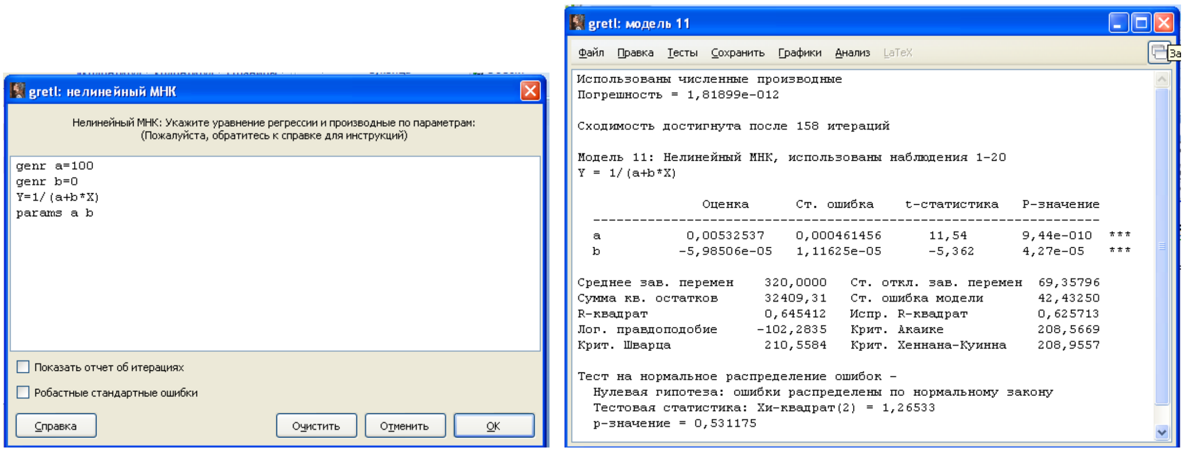


Рис. 4.14. Нелинейный МНК путем численного дифференцирования (обратная модель)

Получим обратную модель регрессии: Y=1/(0,0053-5,985*10-5*X)+ε, R^2=0,63, Se=42,43. Остатки имеют нормальный закон распределения (p>0,05).

Применим грубые оценки начальных параметров: a=0, b=0 (рис. 4.15).

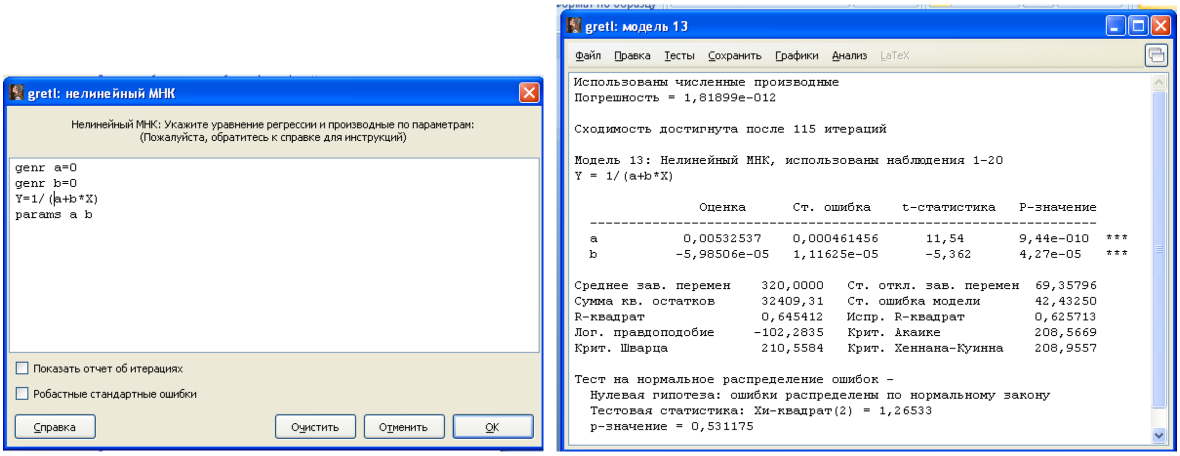


Рис. 4.15. Нелинейный МНК путем численного дифференцирования (обратная модель, a=1, b=0)

Получим обратную модель регрессии: Y=1/(0,0053-5,985*10-5*X)+ε, R^2=0,63, Se=42,43. Остатки имеют нормальный закон распределения (p>0,05).

5.6. Построение гиперболической модели по равносторонней гиперболе Y=a+b/x (рис. 4.16).

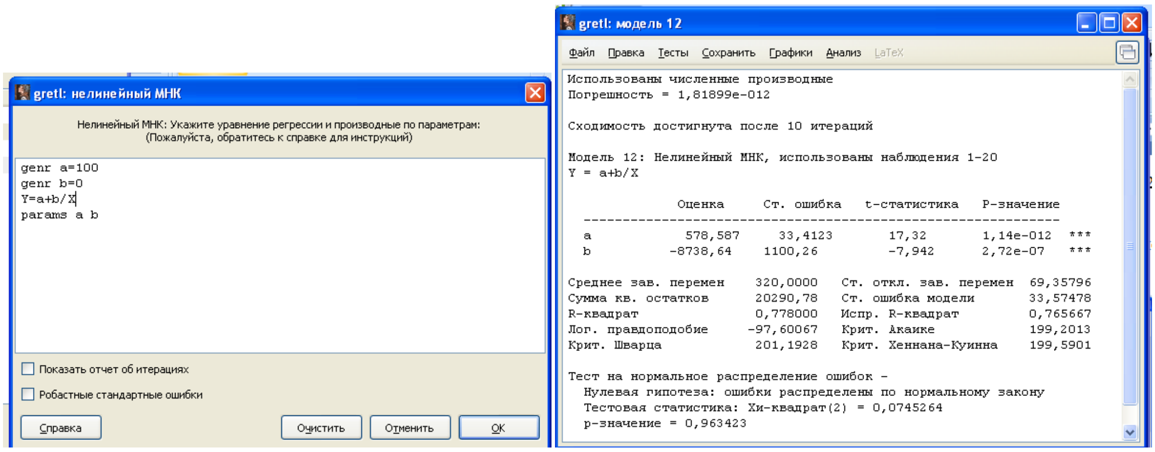


Рис. 4.16. Нелинейный МНК путем численного дифференцирования (гипербола)

Получим обратную модель регрессии: Y=578,587-8738,64/X+ε, R^2=0,766, Se=33,574. Остатки имеют нормальный закон распределения (p>0,05).

Таблица 4.2

Сводная таблица результатов моделирования

Тип модели	Вид модели	R^2	Se
Линейная	Y=64,28+7,18X+ε	0,723	37,51
Степенная	$Y=19{,}66*x^{0{,}782}*\varepsilon$	0,717	36,90
Экспоненциальная	$Y=149{,}698*e^{0{,}021*x}*\varepsilon$	0,668	39,95
Полулогарифмическая	Y=-593,103+257,464 lnX+ε	0,746	14,98
Обратная	Y=1/(0,0053-5,985*10-5*X)+ε	0,626	42,43
Гиперболическая	Y=578,587-8738,64/X+ε	0,766	33,574

Таким образом, исходя из коэффициента детерминации и стандартной ошибки модели наиболее приемлемы полулогарифмическая и гиперболическая модели.

5. Гетероскедастичность и автокорреляция в остатках. Обобщенный МНК

Расчетные формулы

Гетероскедастичностью называется нарушение условия постоянства дисперсии ошибок в модели регрессии. Эффективным методом предварительного анализа однородности ошибок в модели регрессии является визуальный анализ графиков остатков. Как правило, рассматриваются следующие графики:

- остатки в зависимости от расчетных значений Y_x;
- остатки в зависимости от отдельных объясняющих переменных: (e_i, x_k);
- остатки в зависимости от номера наблюдений: (e_i, i) (только в случае временных рядов).

Приведем несколько статистических критериев, позволяющих выявить отклонения от однородности ошибок, а именно выявить явление гетероскедастичности. Во многих критериях проверяется нулевая гипотеза:

$Ho: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$.

Критерий Уайта. Первый вариант. Оцениваем модель

$$e_i^2 = a_0 + \sum_k a_k x_{ik} + \sum_k b_k x_{ik}^2, i = 1,\ldots,n$$

Пусть R^2 - коэффициент детерминации в этой модели. Здесь $e_1,\ldots,e_n$ – остатки, полученные при оценивании основной модели регрессии. Проверяется гипотеза $H_0: a_k = b_k = 0$.

Для проверки гипотезы используется статистика nR^2. При больших объемах выборки и при заданном уровне значимости нулевая гипотеза отвергается, если $nR^2 > \chi^2_{\alpha/2;2p-2}$.

Критерий Уайта. Второй вариант. Оцениваем модель

$$e_i^2 = a_0 + \sum_k a_k x_{ik} + \sum_k b_k x_{ik}^2 + \sum_{j,k} c_{jk} x_{ij} x_{ik},$$

$$i = 1,\ldots,n$$

Пусть R^2 - коэффициент детерминации в этой модели. Здесь $e_1,\ldots,e_n$ – остатки, полученные при оценивании основной модели регрессии. Проверяется гипотеза $H_0: a_k = b_k = c_{jk} = 0,\ k, j = 1,\ldots,m$

Для проверки гипотезы используется статистика nR^2. При больших объемах выборки и при заданном уровне значимости нулевая гипотеза отвергается, если $nR^2 > \chi^2_{\alpha/2;p-1}$.

Тест Бройша-Пагана основан на следующей двухшаговой процедуре:

1) в исходной модели регрессии находим МНК-остатки e_i и обозначаем

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 .$$

2) вычисляем объясненную сумму квадратов ESS во вспомогательной регрессии

$$(\frac{e_i^2}{\hat{\sigma}^2}) = \gamma_0 + z_{i1}\gamma_1 + ... + z_{ip}\gamma_p + error .$$

При справедливости нулевой гипотезы $H_0 : \sigma_i^2 = \sigma^2 = f(\gamma_0)$ (т.е. при гомоскедастичности ошибок регрессии) LM-статистика (статистика множителей Лагранжа) LM=ESS/2 асимптотически имеет распределение хи-квадрат. Следовательно, нулевая гипотеза отвергается при $LM > \chi^2_{kp}$, $\chi^2_{kp} = \chi^2(\alpha; p)$.

Другая форма теста Бройша-Пагана основана на статистике nR_0^2, где коэффициент R_0^2 вычисляется по вспомогательной регрессии

$$e_i^2 = \gamma_0 + z_{i1}\gamma_1 + ... + z_{ip}\gamma_p + error .$$

При справедливости нулевой гипотезы (H_0) статистика nR_0^2 асимптотически имеет распределение хи-квадрат:

$$nR_0^2 \approx \chi_p^2 .$$

Для корректировки модели на гетероскедастичность часто используется частный случай обобщенного метода наименьших квадратов (GLS) - взвешенный МНК. Метод основан на минимизации взвешенной суммы квадратов отклонений $\sum_{i=1}^{n} w_i^2 \cdot (y_i - \hat{y}_i)^2 \to \min$. В некоторых численных реализациях взвешенного МНК предполагается, что стандартные отклонения ошибок прямо пропорциональны значению одной из переменных: $\sigma_i = \sigma x_i$. Тогда в качестве весов следует брать $w_i = 1/x_i$. Разделив регрессионное уравнение на x_i, получим

$$\frac{y_i}{x_i} = \frac{a}{x_i} + b + \frac{\varepsilon_i}{x_i} .$$

Введем новые переменные $y_i^* = y_i/x_i$, $x_i^* = 1/x_i$, $\varepsilon^* = \varepsilon_i/x_i$. Ошибки в новой модели некоррелированы и гомоскедастичны. Далее оценивается преобразованная форма модели и ее коэффициенты используются для записи исходной модели с коррекцией на гетероскедастичность.

Задание 5.1. Провести регрессионный анализ данных об урожайности зерновых культур в некоторой стране за период 1945- 2005 гг. и проверить остатки регрессии на гетерокедастичность графически, а также используя тесты Уайта, Бреуша-Пагана, Коенкера, тест ранговой корреляции Спирмена.

Таблица 5.1

Данные об урожайности зерновых культур

X	1945	1946	1947	1948	1949	1950	1951	1952	1953
Y	5,6	4,6	7,3	6,7	6,9	7,9	7,4	8,6	7,8
X	1954	1955	1956	1957	1958	1959	1960	1961	1962
Y	7,7	8,4	9,9	8,4	11,1	10,4	10,9	10,7	10,9
X	1963	1964	1965	1966	1967	1968	1969	1970	1971
Y	8,3	11,4	9,5	13,7	12,1	14	13,2	15,6	15,4
X	1972	1973	1974	1975	1976	1977	1978	1979	1980
Y	14	17,6	15,4	10,9	17,5	15	18,5	14,2	14,9
X	1981	1982	1983	1984	1985	1986	1987	1988	1989
Y	12,6	15,2	15,9	14,4	16,2	18	18,3	17	18,8
X	1993	1998	2001	2002	2003	2004	2005		
Y	15,7	15,1	19,4	19,6	17,8	18,8	18,5		

Решение в среде Gretl

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

2. Построение регрессии урожайности в зависимости от времени: Модель/Метод наименьших квадратов… (рис. 5.1).

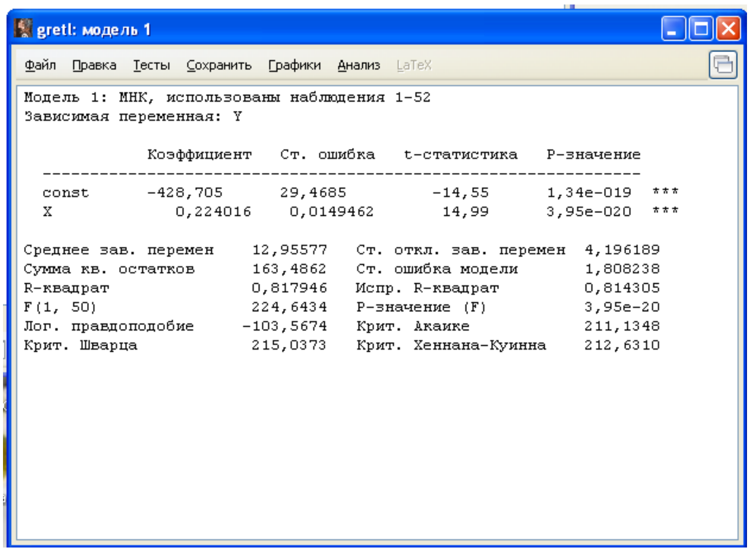


Рис. 5.1. Линейная модель регрессии урожайности в зависимости от времени

Получим линейную модель регрессии: Y=-428,705+0,224*X+ε.

3. Построение графика остатков: Графики / График остатков / В зависимости от X…

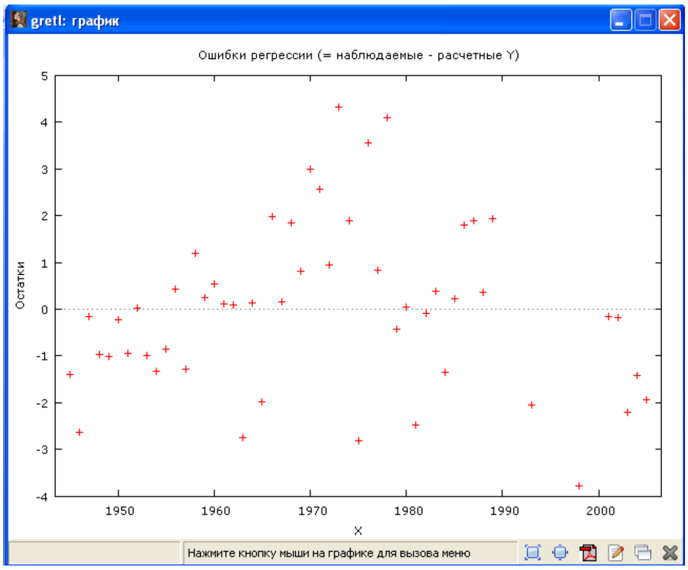


Рис. 5.2. График остатков в зависимости от переменной X

Можно предположить зависимость остатков от изменения переменной X.

4. Для более ясной визуализации найдем модули остатков: Сохранить / Остатки. Затем: Добавить / Добавить новую переменную (рис. 5.3).

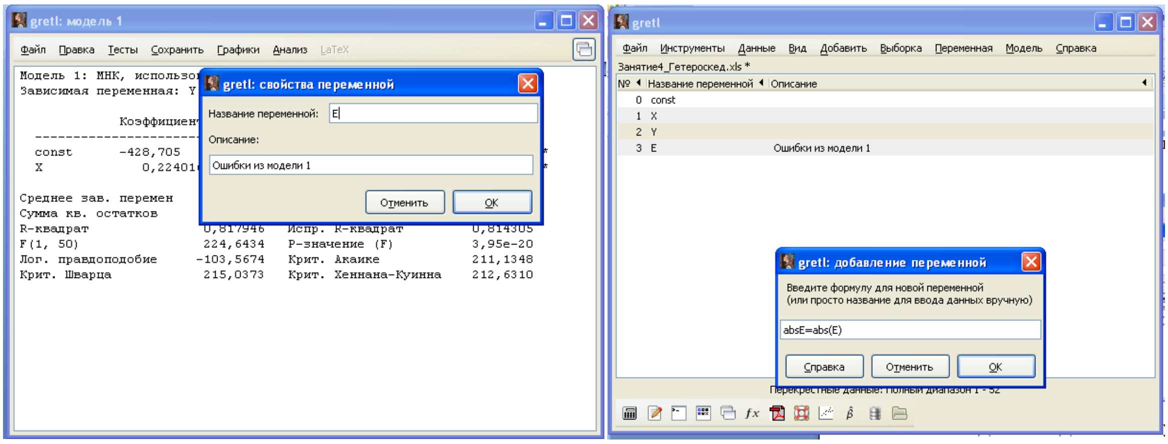


Рис. 5.3. Окно добавления модулей остатков

5. В главном меню строим график модулей остатков: Вид / График / График X-Y(рис. 5.4).

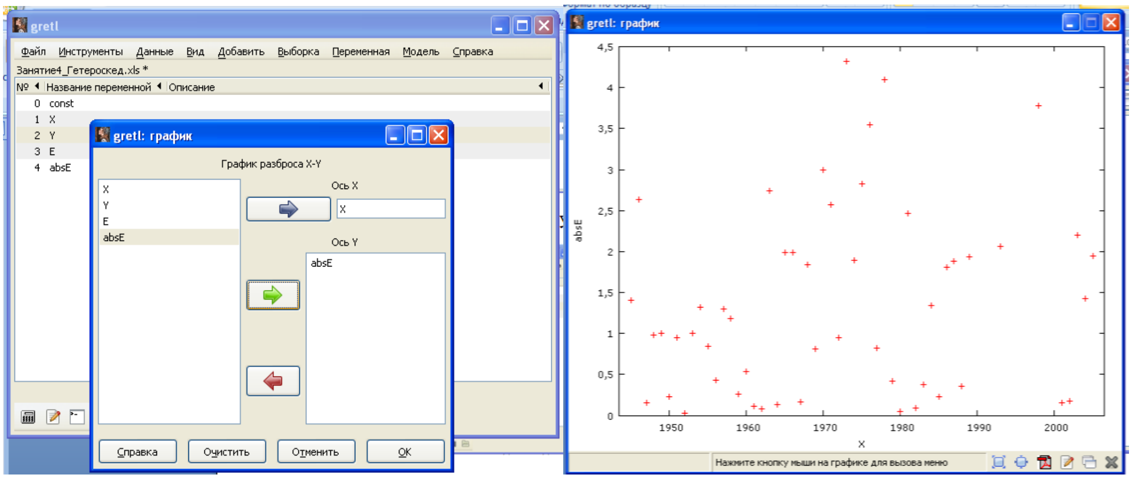


Рис. 5.4. График модулей остатков в зависимости от переменной X

Визуально можно предположить гомоскедастичность остатков.

6. Выполним проверку при помощи статистических тестов: В окне модели: Тесты / Гетероскедастичность…(рис. 5.5).

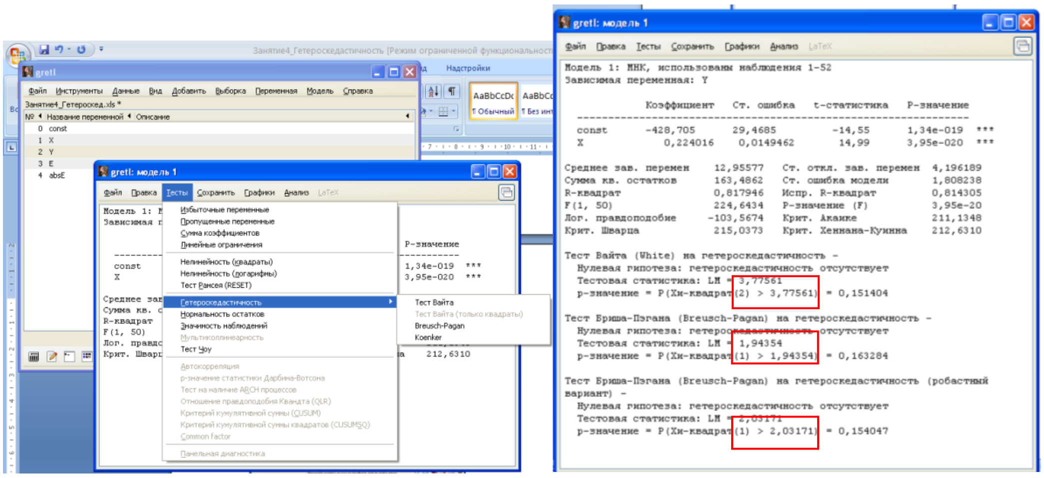


Рис. 5.5. Тестирование остатков на гетероскедастичность

В тестах Уайта, Бреуша-Пагана, Коенкера p-значение для «хи-квадрат» статистики больше, чем, 0,05. Значит, подтверждается нулевая гипотеза о гомоскедастичности остатков регрессии урожайности в зависимости от времени.

С целью измерения тесноты взаимосвязи между остатками регрессии и переменной X выполним тест Спирмена. Для теста Спирмена: Модель / Робастные оценки / Ранговая корреляция / Корреляция Спирмена (рис. 5.6):

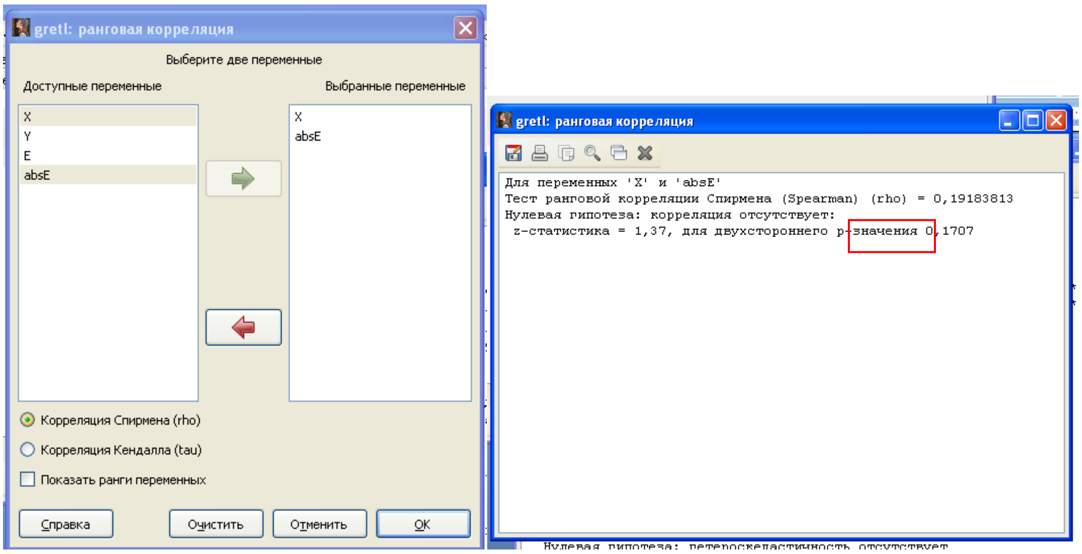


Рис. 5.6. Тест ранговой корреляции Спирмена

Коэффициент ранговой корреляции Спирмена составил 0,192, связь между остатками регрессии и переменной X слабая, p-значение для t-статистики больше, чем 0,05. Значит, нулевая гипотеза о незначимости коэффициента ранговой корреляции Спирмена подтверждается, остатки регрессии гомоскедастичны.

Задание 5.2. Провести регрессионный анализ данных о динамике среднедушевых сбережений населения (Y, млрд. руб.) и проверить остатки регрессии на гетерокедастичность графически, а также используя тесты Уайта, Бреуша-Пагана, Коенкера. Выполнить коррекцию регрессии на гетероскедастичность.

Таблица 5.2

Динамика среднедушевых сбережений населения

	,3	,1	,2	,9	,0	,7	,8	,5	,5	,0
	,0	,0	,0	,0	,0	,0	,0	,0	,0	0,0

Решение с помощью пакета Gretl

1. Создание рабочего листа 2 с исходными данными в Excel, его сохранение в файле «ЗанятиеГТС.xlsx» и импорт данных из таблицы Excel. В ос-

новном меню выберем пункт: Файл / Открыть / Импорт / Excel / ЗанятиеГТС.xlsx / лист 2.

2. Построение регрессии среднедушевых сбережений населения в зависимости от времени: Модель / Метод наименьших квадратов…(рис. 5.7).

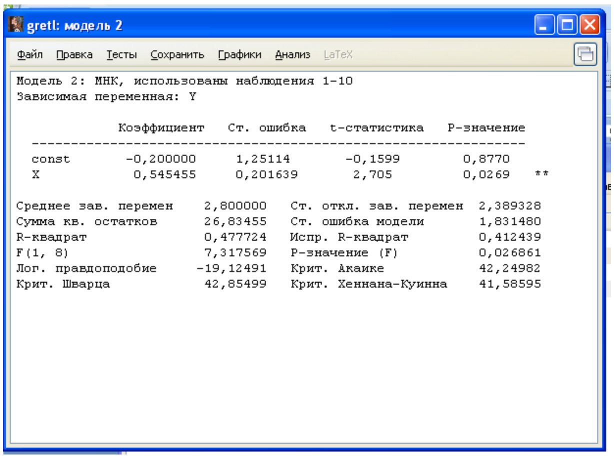


Рис. 5.7. Линейная модель регрессии среднедушевых сбережений населения в зависимости от времени

Получим модель регрессии: $Y=-0{,}2+0{,}545X+\varepsilon$.

3. Найдем модули остатков: Сохранить / Остатки. Затем: Добавить / Добавить новую переменную.

4.График модулей остатков: Вид / График / График X-Y(рис. 5.8).

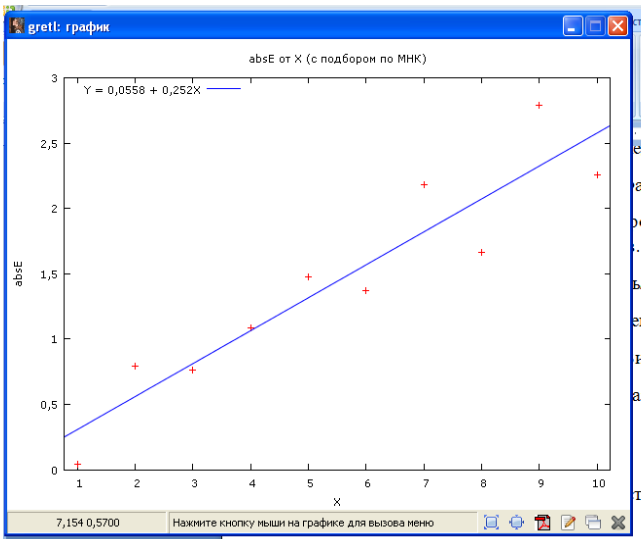


Рис. 5.8. График модулей остатков в зависимости от переменной X

Визуально можно предположить гетероскедастичность остатков.

5. Выполним проверку при помощи статистических тестов: Тесты / Гетероскедастичность…(рис. 5.9).

Отчет после проведения тестов:

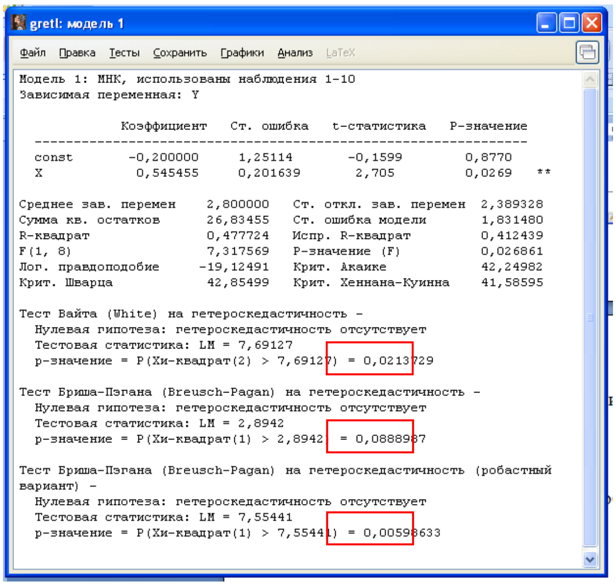


Рис. 5.9. Тестирование остатков на гетероскедастичность

По тесту Уайта, Бреуша-Пагана (робастный вариант) гетероскедастичность обнаружена (p<0,05).

6. Для теста Спирмена: Модель / Робастные оценки / Ранговая корреляция/ Корреляция Спирмена (рис. 5.10).

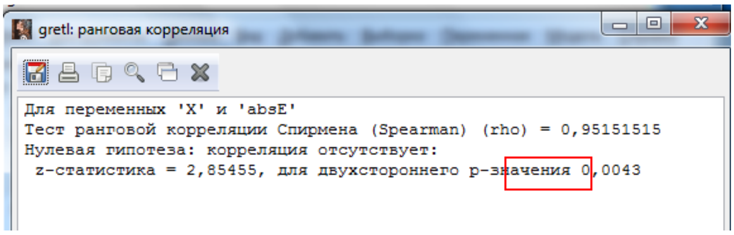


Рис. 4.10. Тест ранговой корреляции Спирмена

Коэффициент ранговой корреляции Спирмена составил 0,952, связь между остатками регрессии и переменной X тесная, p-значение для t-статистики меньше, чем 0,05. Значит, нулевая гипотеза о незначимости коэффициента ранговой корреляции Спирмена отклоняется, остатки регрессии гетероскедастичны. Выполним коррекцию модели с гетероскедастичными остатками путем использования частного случая обобщенного метода наименьших квадратов взвешенного МНК. Предположим, что $\sigma_e^2=\sigma_0*x_i^2$.

4.7. Формируем структурную матрицу XC: Добавить / Матрицу / Получить из данных (рис. 5.11).

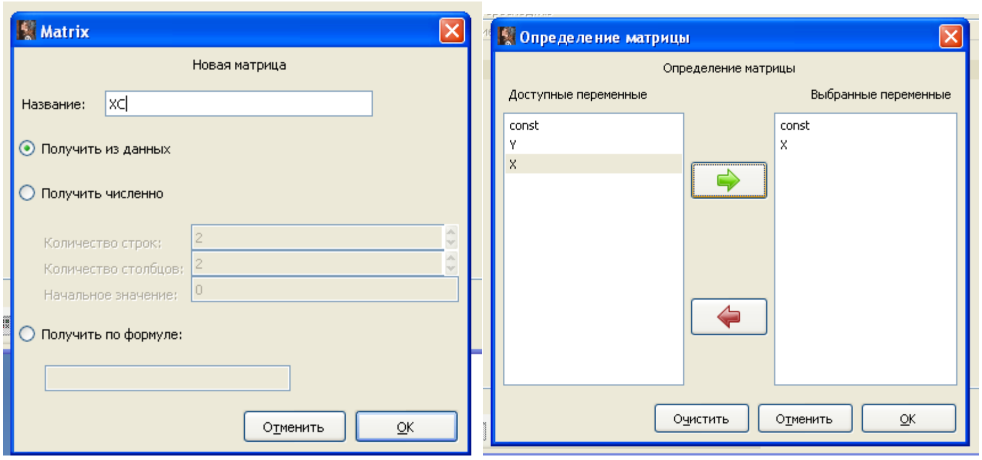


Рис. 5.11. Окно добавления структурной матрицы

8. Формируем ковариационную матрицу V. Вначале формируем единичную матрицу I порядка 10: Добавить / Матрицу / Получить численно (10x10) / Заполнить / Единичную матрицу (рис. 5.12).

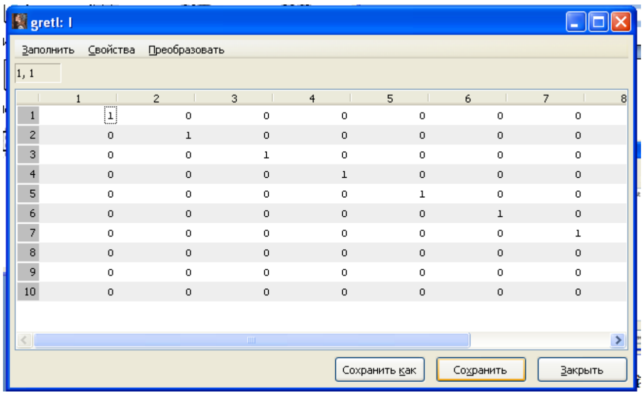


Рис. 5.12. Окно добавления единичной матрицы

9. Формируем матрицу XP: Добавить / Матрицу / Получить по формуле. Затем формируем ковариационную матрицу V: Добавить / Матрицу / Получить по формуле (рис. 5.13).

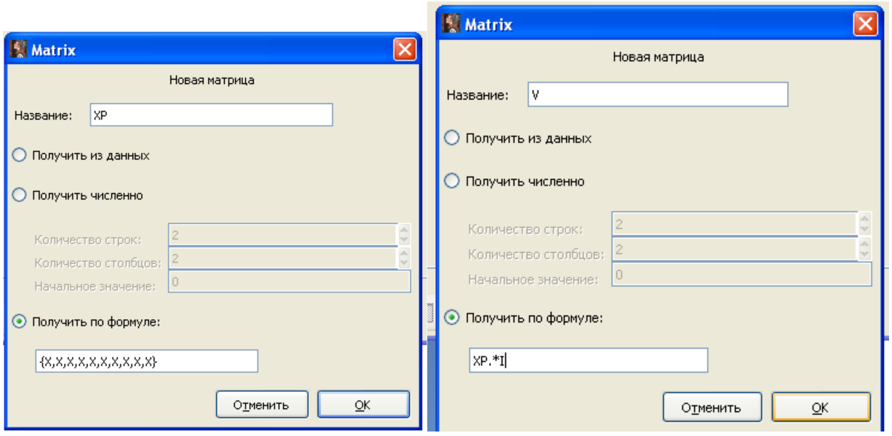


Рис. 5.13. Окно добавления ковариационной матрицы

10. Расчет вектор-столбца коэффициентов B. Формируем обратную матрицу VI: Добавить / Матрицу / по формуле (рис. 5.14):

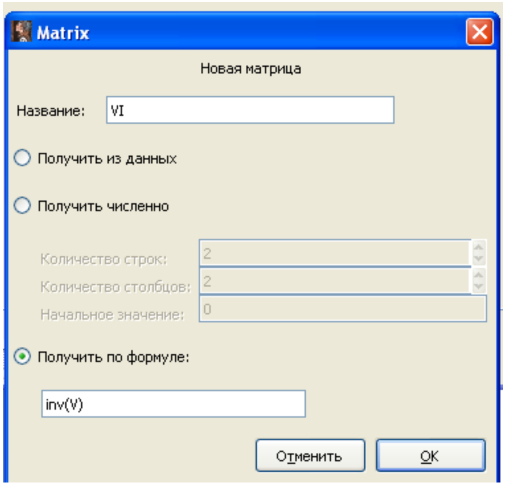


Рис. 5.14. Окно добавления обратной матрицы VI

11. Формируем вектор-столбец коэффициентов B (рис. 5.15):

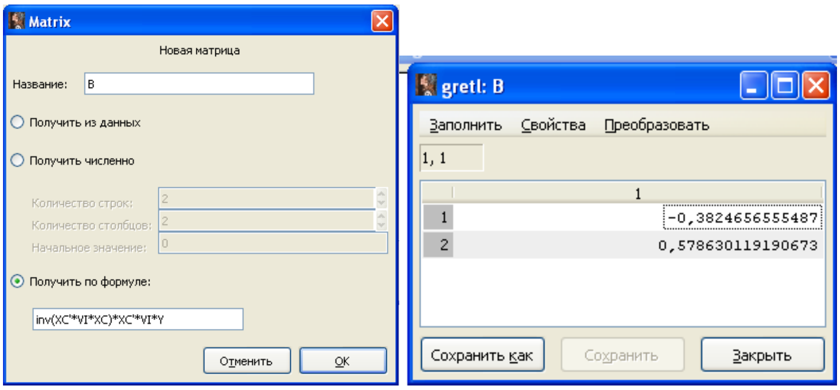


Рис. 5.15. Окно добавления вектор-столбца коэффициентов B

Получаем модель: Y=-0,382+0,578X+ε.

12. Выполним преобразование исходной модели, разделив обе части уравнения на переменную X. Введем новые переменные: Добавить / Добавить новую переменную (рис. 5.16).

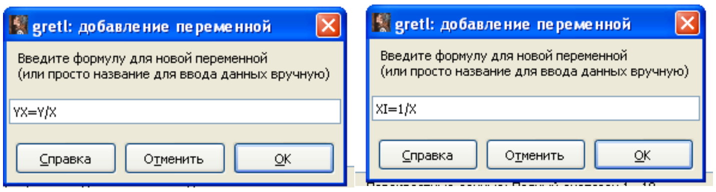


Рис. 5.16. Окно добавления новой переменной

13. Оцениваем регрессию YX=a*XI+b: Модель / Метод наименьших квадратов (рис. 5.17).

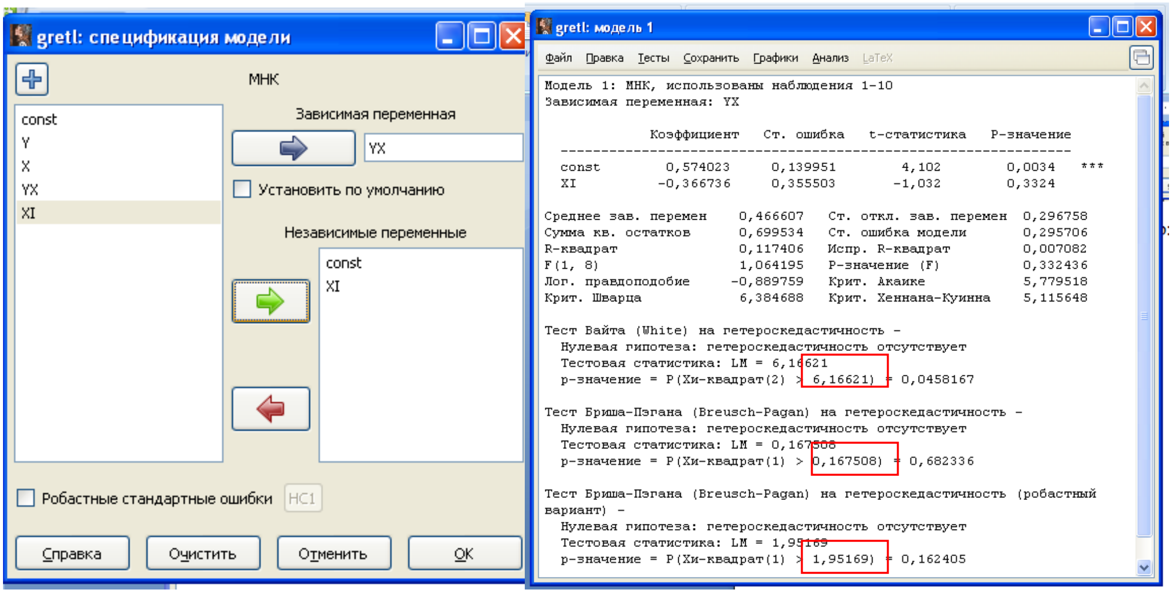


Рис. 5.17. Оценки регрессии после преобразования переменных

По тесту Уайта гетероскедастичность сохранилась ($p<0,05$). По тестам Бреуша-Пагана и Бреуша-Пагана (робастный вариант) остатки гомоскедастичны ($p>0,05$).

Модель регрессии с коррекцией на гетероскедастичность:

Y=-0,366+0,574X+ε.

Модель регрессии до коррекции (пункт 5.2): Y=-0,2+0,545X+ε.

Стандартные ошибки коэффициентов до коррекции (пункт 5.2) составили: m_a=1,25; m_b=0,201. После коррекции: m_a=0,355; m_b=0,139. Ошибки уменьшились.

При таком способе оценивания модели с использованием взвешенного метода наименьших квадратов необходимо провести достаточно большой объем вычислений для анализа качества модели. Поэтому воспользуемся одной из численных реализаций взвешенного МНК для построения модели в Gretl.

Теперь для коррекции модели регрессии используем встроенные инструменты Gretl. Первый способ коррекции на гетероскедастичность: Модель /Другие линейные модели / С коррекцией на гетероскедастичность (рис. 5.18).

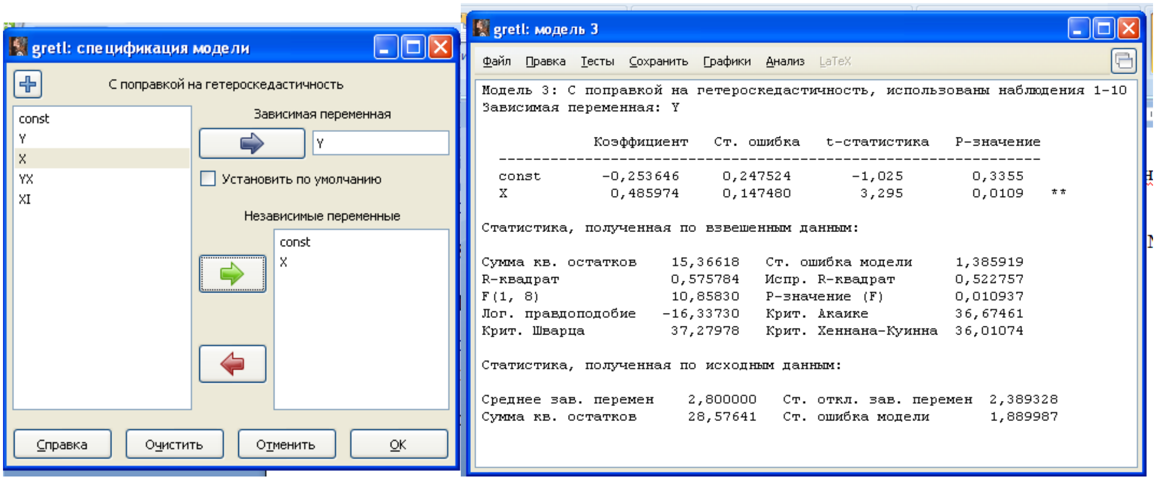


Рис. 5.18. Окно коррекции модели на гетероскедастичность (первый способ)

Получим модель: Y= -0,254+0,486X+ε.

Сравним точность оценивания, используя стандартные ошибки коэффициентов. До коррекции (пункт 5.2): m_a=1,25 m_b=0,201, после коррекции: m_a=0,247; m_b=0,147. Ошибки уменьшились.

Второй способ коррекции на гетероскедастичность: Модель / Другие линейные модели / Взвешенный МНК (рис. 5.19). В строке «Весовая переменная» указывается переменная, полученная после преобразования, в нашем случае переменная XI.

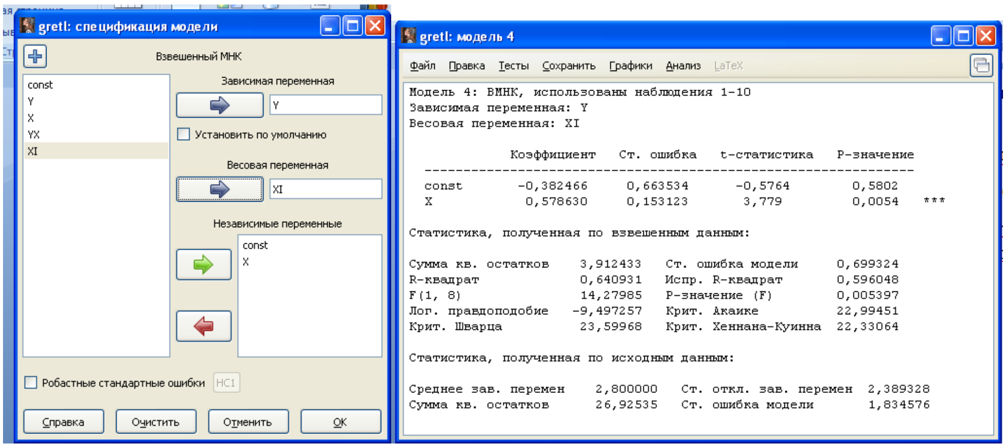


Рис. 5.19. Окно коррекции модели на гетероскедастичность (второй способ)

Получим модель: Y=-0,382+0,578X+ε, m_a=0,663, m_b=0,153.

Таблица 5.3

Сводная таблица результатов коррекции на гетероскедастичность

Тип уравнения	Вид уравнения	Стандартная ошибка коэф.а	Стандартная ошибка коэф.b	Стандартная ошибка модели
Исходное	Y=-0,2+0,545X+ε	1,25	0,201	1,831
Взвешенный МНК (вручную)	Y=-0,366+0,574X+ε	0,355	0,139	0,296
С поправкой на гетероскедастичность (Gretl)	Y= -0,254+0,486X+ε	0,247	0,147	1,386
Взвешенный МНК (Gretl)	Y=-0,382+0,578X+ε	0,663	0,153	0,699

Таким образом, исходя из стандартной ошибки модели и стандартных ошибок параметров регрессии *a* и *b*, наиболее приемлема модель регрессии, полученная с использованием «ручного» способа взвешенного МНК.

Задание 5.3. Провести регрессионный анализ динамики золотовалютных резервов РФ за период с 26.12.2003 г. по 07.01.2005 г., уровни еженедельные, по состоянию на пятницу каждой недели.

Таблица 5.4

Динамика золотовалютных резервов

t	Yt	t	Yt	t	Yt	t	Yt
0	77,8	98	83,6	196	89,2	294	100,1
7	77,1	105	83,5	203	89,2	301	105,2
14	78,9	112	83,2	210	88,9	308	107,3
21	79,1	119	82,8	217	88,7	315	112,8
28	82,7	126	82,7	224	89	322	113,1
35	84,1	133	83,4	231	89,6	329	113,9
42	84,3	140	82,7	238	88,3	336	117,1
49	88	147	83,2	245	88,8	343	121,6
56	86,7	154	85,4	252	89,1	350	120,3
63	86,4	161	85,6	259	90	357	119,8
70	84,6	168	86,2	266	92,6	364	120,7
77	84,6	175	87,4	273	94,3	371	124,5
84	84,8	182	87,9	280	95,3	378	124,6
91	83,7	189	88,3	287	98,3		

Решение в среде Gretl

1. Создание рабочего листа 3 с исходными данными в Excel, его сохранение в файле «ЗанятиеГТС.xlsx» и импорт данных из таблицы Excel. В основном меню выберем пункт: Файл / Открыть / Импорт / Excel / ЗанятиеГТС.xlsx / лист 3 (рис. 5.20).

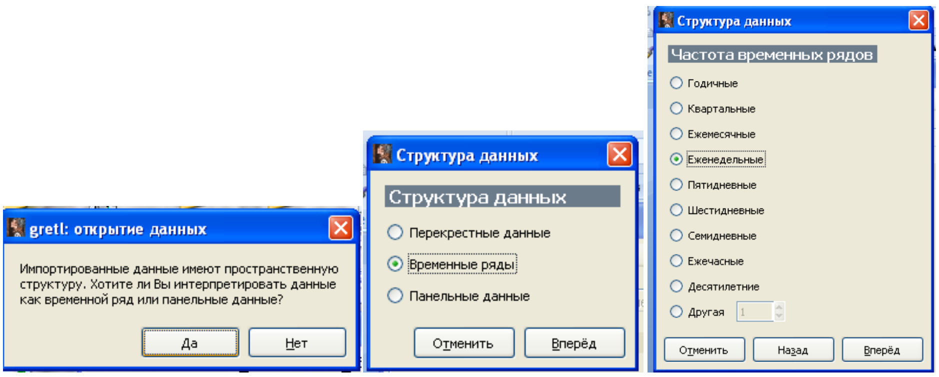


Рис. 5.20 (а). Окно импорта данных

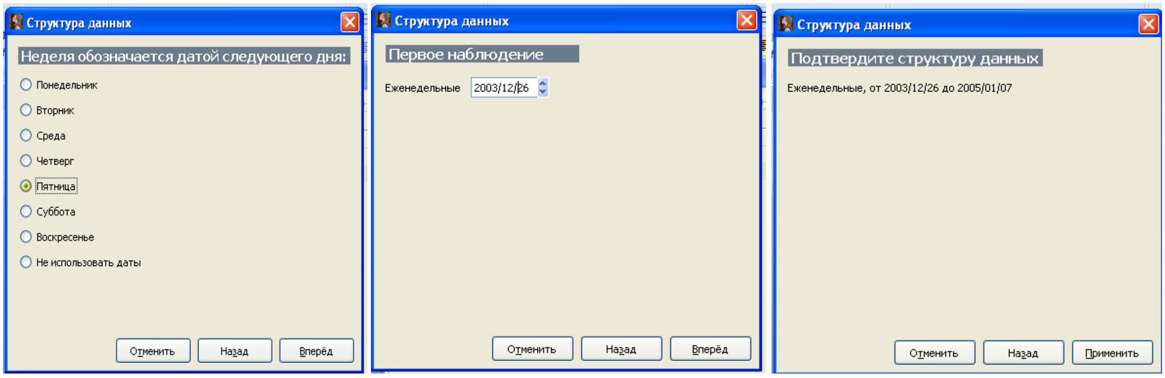
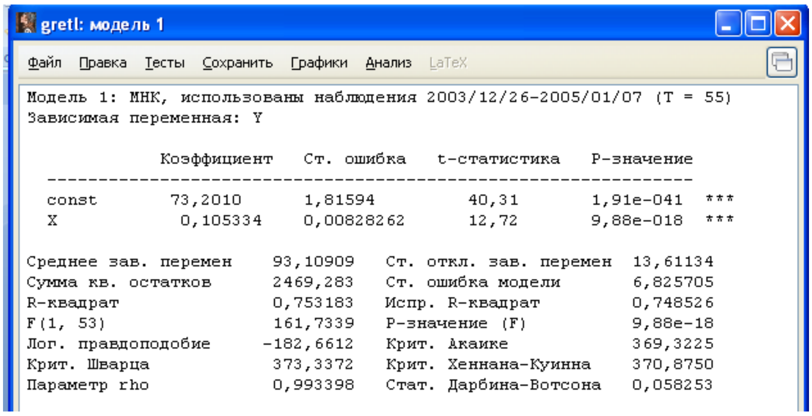


Рис. 5.20 (б). Окно импорта данных

2. Построение уравнения регрессии золотовалютных резервов РФ в зависимости от времени: Модель / Метод наименьших квадратов…(рис. 5.21).




```
Модель 1: МНК, использованы наблюдения 2003/12/26-2005/01/07 (T = 55)
Зависимая переменная: Y

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        73,2010      1,81594        40,31       1,91e-041  ***
  X             0,105334    0,00828262     12,72       9,88e-018  ***

Среднее зав. перемен     93,10909   Ст. откл. зав. перемен   13,61134
Сумма кв. остатков       2469,283   Ст. ошибка модели        6,825705
R-квадрат                0,753183   Испр. R-квадрат          0,748526
F(1, 53)                 161,7339   Р-значение (F)           9,88e-18
Лог. правдоподобие      -182,6612   Крит. Акаике             369,3225
Крит. Шварца             373,3372   Крит. Хеннана-Куинна     370,8750
Параметр rho             0,993398   Стат. Дарбина-Вотсона    0,058253
```

Рис. 5.21. Модель регрессии по исходным данным

Получим модель: $Yt=73{,}2+0{,}105*t+\varepsilon$.

3. Тестирование остатков на гетероскедастичность: Тесты / Гетероскедастичность…(рис. 5.22).

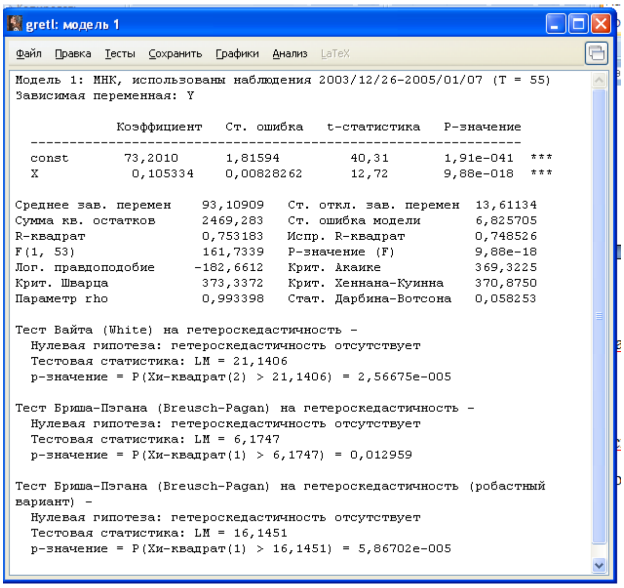


Рис. 5.22. Окно тестирования остатков на гетероскедастичность

Р-значение <0,05 во всех тестах, значит в остатках регрессии присутствует гетероскедастичность.

4. Визуальный анализ остатков на автокорреляцию: Графики / График остатков / В зависимости от X…(рис. 5.23).

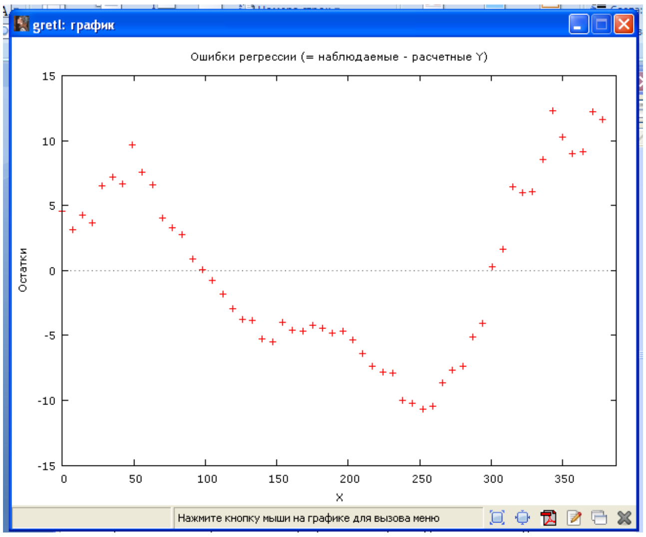


Рис. 5.23. График остатков регрессии

Чередование зон положительных и отрицательных знаков дает предположение о положительной автокорреляции в остатках.

5. Тестирование остатков на автокорреляцию: Тесты / Автокорреляция… Порядок лага=0 (рис. 5.24).

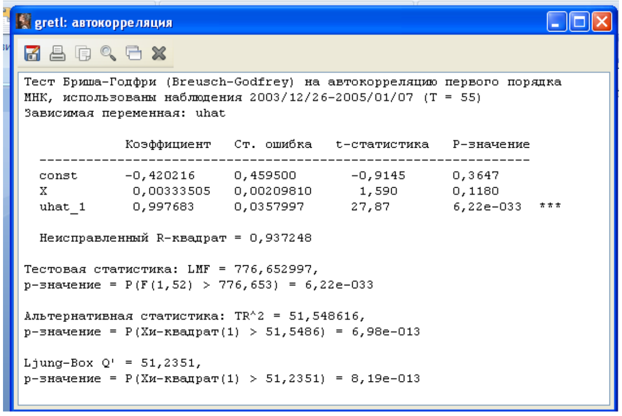


Рис. 5.24. Окно тестов на автокорреляцию остатков регрессии

Р-значение <0,05 в тесте Бриша-Годфри и в тесте Льюнга-Бокса, обнаружена автокорреляция.

6. Выполним визуальный анализ автокорреляционных функций - ACF и PACF. В окне модели: Графики / Коррелограмма остатков. Максимальный лаг=11(рис. 5.25).

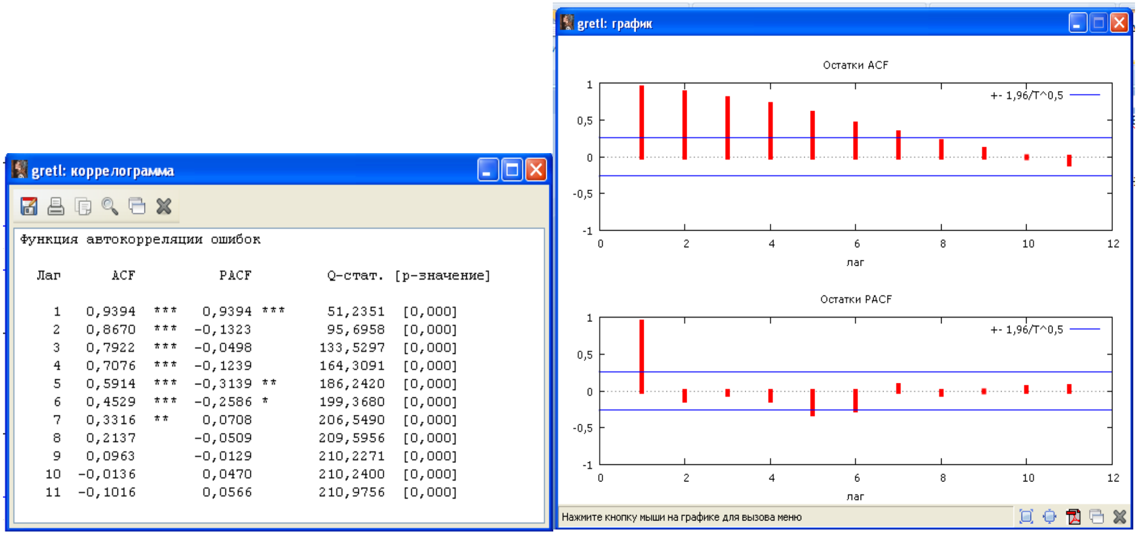


Рис. 5.25. График автокорреляционных функций - ACF и PACF

Синие линии на графике – критические значения распределения Стьюдента при α=0,05. Остатки ACF монотонно убывают. В остатках PACF максимальные значения имеют первый и пятый коэффициент. Доминирует автокорреляция первого порядка.

7. Тест Дарбина-Уотсона: В окне модели Тесты/ p-значение статистики Дарбина –Уотсона: DW=0,058. Определим критические значения: Инструменты/ Критические значения (рис. 5.26):

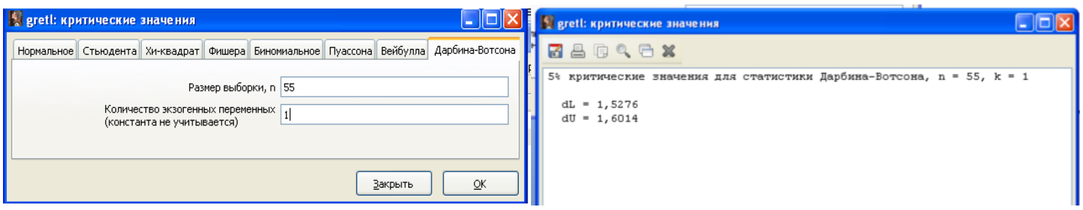


Рис. 5.26. Критические значения статистики Дарбина –Уотсона

DW<dL, P-значение <0,01, значит, обнаружена положительная автокорреляция. Для устранения автокорреляции в остатках регрессии выполним авторегрессионное преобразование.

8. Первый способ коррекции на автокорреляцию первого порядка путем авторегрессионного преобразования: Модель / Временные ряды / Кохрана-Оркатта (рис. 5.27):

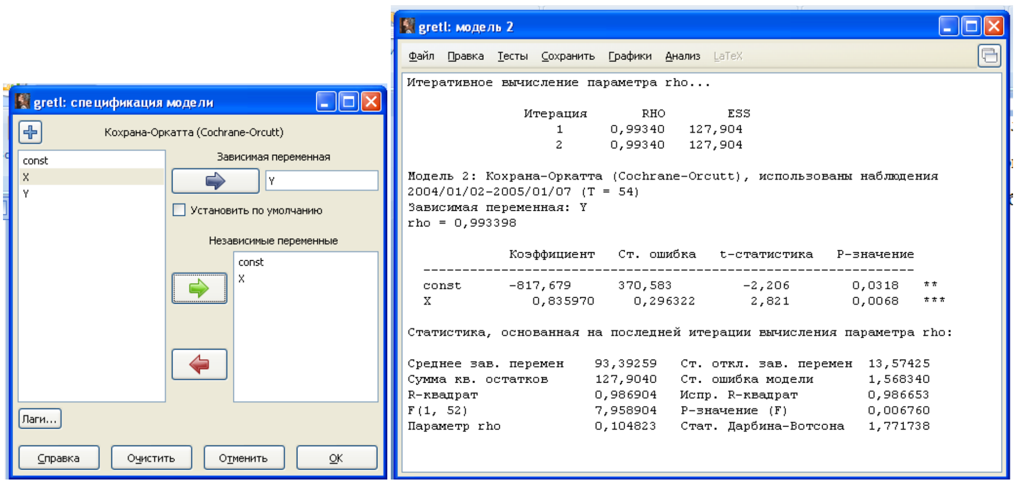


Рис. 5.27. Авторегрессионное преобразование методом Кохрана-Оркатта

Проведено две итерации. Первое наблюдение утеряно, n=54. Оценка коэффициента автокорреляции первого порядка равна: ρ=0,993. Получим модель: Yt=-817,68+0,836*t, статистика Дарбина-Уотсона = 1,77.

Определим критические значения: Инструменты / Критические значения (рис. 5.28):

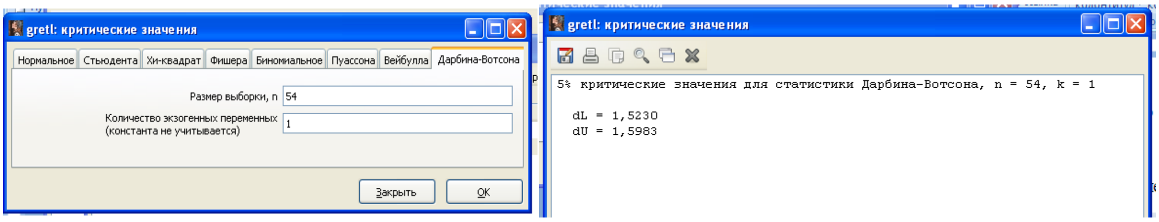


Рис. 5.28. Критические значения статистики Дарбина –Уотсона

DW=1,77 >d_U , автокорреляция в остатках отсутствует. Также коррелограмма подтверждает отсутствие автокорреляции малых порядков (рис. 5.29):

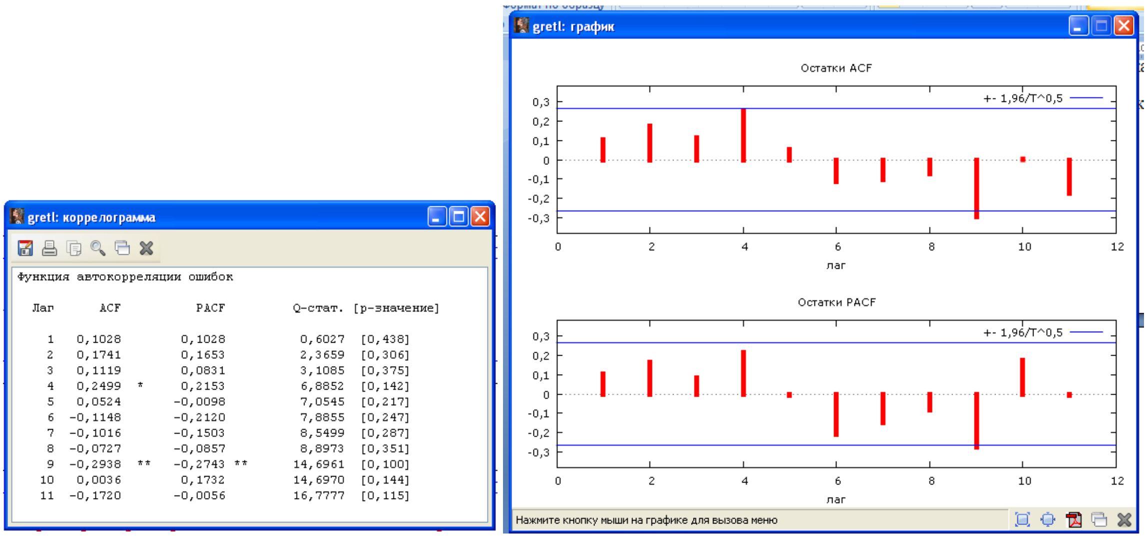


Рис. 5.29. График автокорреляционных функций - ACF и PACF (метод Кохрана-Оркатта)

9. Второй способ коррекции на автокорреляцию первого порядка. Проводим авторегрессионное преобразование: Модель / Временные ряды / Хилдрета-Лу (рис. 5.30):

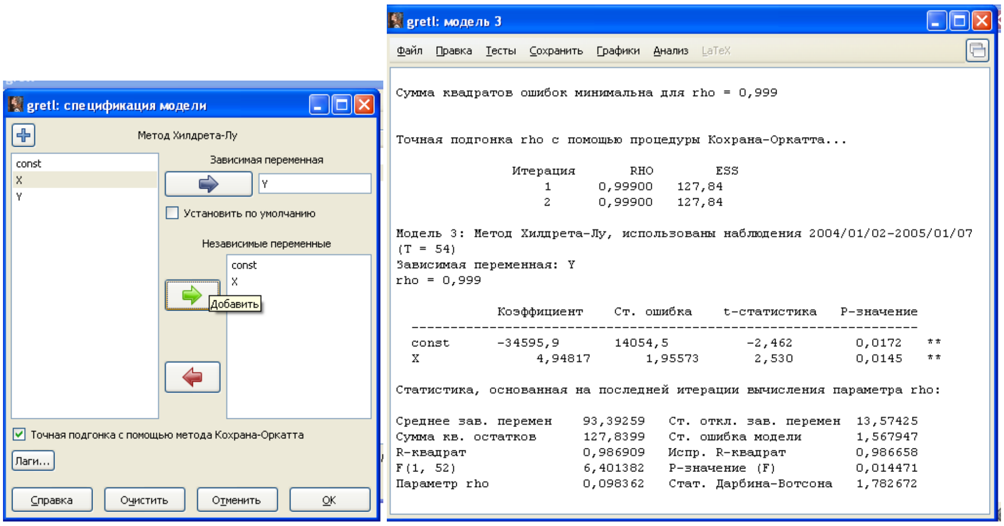


Рис. 5.30. Авторегрессионное преобразование методом Хилдрета-Лу

Проведено две итерации. Первое наблюдение утеряно, n=54. Оценка коэффициента автокорреляции первого порядка равна: ρ=0,999. Получим модель: Y=-34595,9+4,948X. Статистика Дарбина-Уотсона = 1,78. DW=1,78 >d_U , значит, автокорреляция в остатках отсутствует. Коррелограмма подтверждает отсутствие автокорреляции малых порядков (рис. 5.31):

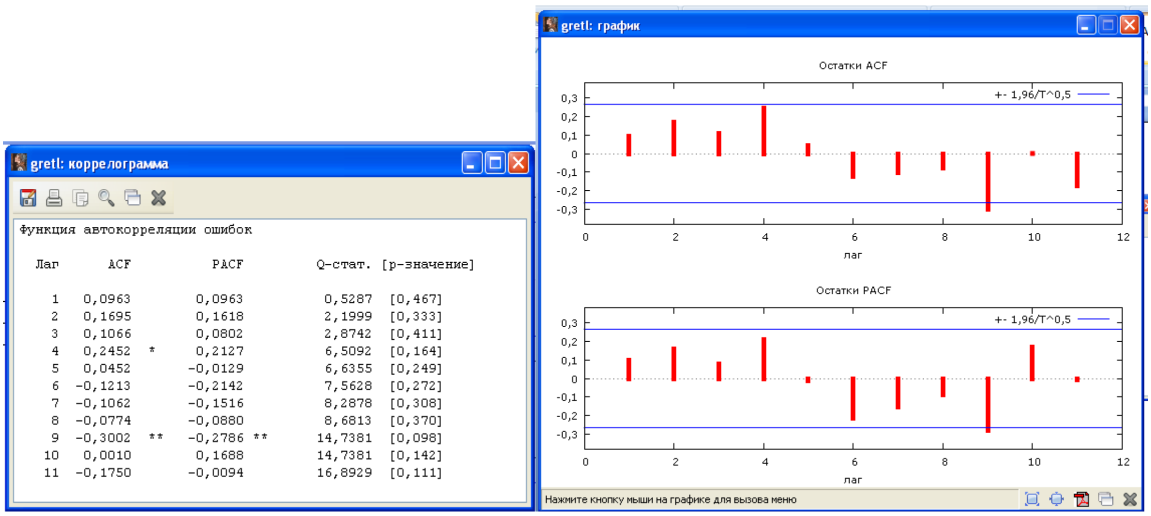
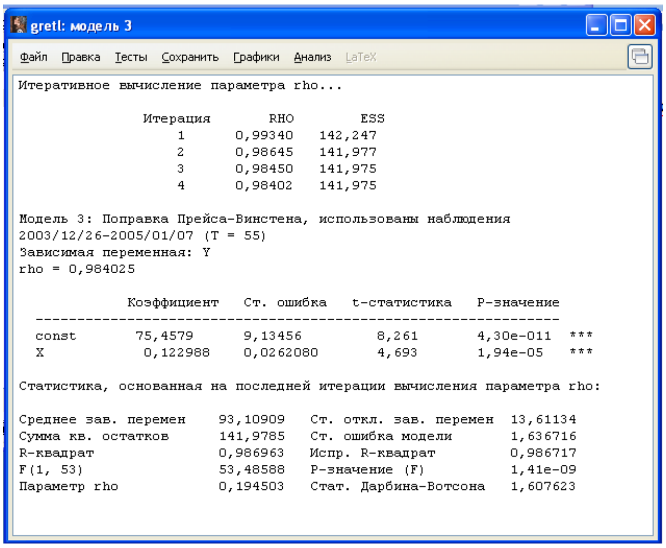


Рис. 5.31. График автокорреляционных функций - ACF и PACF

(метод Хилдрета-Лу)

10. Третий способ коррекции на автокорреляцию первого порядка. Проводим авторегрессионное преобразование: Модель / Временные ряды / Прайса-Винстена (рис. 5.32):


gretl: модель 3

Файл Правка Тесты Сохранить Графики Анализ LaTeX

```
Итеративное вычисление параметра rho...

              Итерация        RHO         ESS
                 1         0,99340    142,247
                 2         0,98645    141,977
                 3         0,98450    141,975
                 4         0,98402    141,975

Модель 3: Поправка Прейса-Винстена, использованы наблюдения
2003/12/26-2005/01/07 (T = 55)
Зависимая переменная: Y
rho = 0,984025

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ------------------------------------------------------------
  const       75,4579      9,13456          8,261       4,30e-011  ***
  X            0,122988    0,0262080        4,693       1,94e-05   ***

Статистика, основанная на последней итерации вычисления параметра rho:

Среднее зав. перемен     93,10909   Ст. откл. зав. перемен   13,61134
Сумма кв. остатков       141,9785   Ст. ошибка модели        1,636716
R-квадрат                0,986963   Испр. R-квадрат          0,986717
F(1, 53)                 53,48588   Р-значение (F)           1,41e-09
Параметр rho             0,194503   Стат. Дарбина-Вотсона    1,607623
```

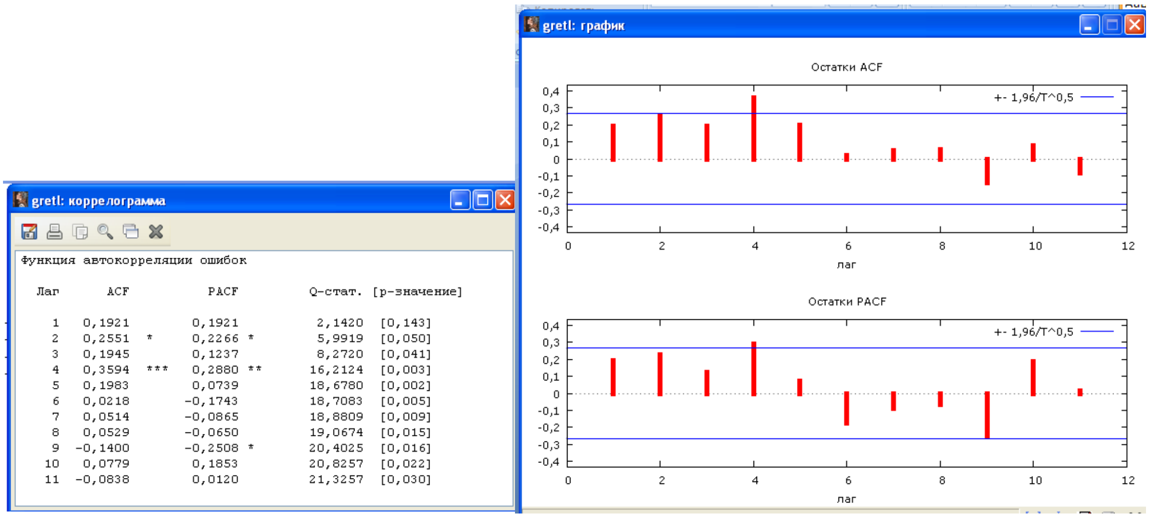
Рис. 5.32. Авторегрессионное преобразование методом Прайса-Винстена

Получим модель: Y=75,458+0,123X+ε.

Проведено 4 итерации, n=55.

Статистика Дарбина-Уотсона = 1,61. DW=1,61 >d_U, значит, автокорреляция в остатках отсутствует.

Коррелограмма подтверждает отсутствие автокорреляции малых порядков (рис. 5.33):



Лаг	ACF		PACF		Q-стат.	[р-значение]
1	0,1921		0,1921		2,1420	[0,143]
2	0,2551	*	0,2266	*	5,9919	[0,050]
3	0,1945		0,1237		8,2720	[0,041]
4	0,3594	***	0,2880	**	16,2124	[0,003]
5	0,1983		0,0739		18,6780	[0,002]
6	0,0218		-0,1743		18,7083	[0,005]
7	0,0514		-0,0865		18,8809	[0,009]
8	0,0529		-0,0650		19,0674	[0,015]
9	-0,1400		-0,2508	*	20,4025	[0,016]
10	0,0779		0,1853		20,8257	[0,022]
11	-0,0838		0,0120		21,3257	[0,030]

Рис. 5.33. График автокорреляционных функций - ACF и PACF

(метод Хилдрета-Лу)

Таблица 5.5

Сводная таблица результатов авторегрессионного преобразования

Тип преобразования	Вид уравнения	DW	Стандартные ошибки коэффициентов	Стандартная ошибка модели
Исходная	Y=73,2+0,105X+ε	0,06	ma=1,816 mb=0,008	6,82
Кохрана-Оркатта	Y=-817,68+0,836X+ε	1,77	ma=370,8 mb=0,296	1,57
Хилгрета-Лу	Y=-34595,9+4,948X+ε	1,78	ma=14054,5 mb=1,956	1,57
Прайса-Винстена	Y=75,458+0,123X+ε	1,61	ma=9,135 mb=0,026	1,64

Таким образом, исходя из величин стандартных ошибок коэффициентов регрессии и модели в целом наиболее приемлемо авторегрессионное преобразование с использованием поправки Прайса-Винстена.

11. Построим четыре прогноза:

– по регрессии без последнего наблюдения (k=55, 07.01.2005), x_{55}=378, Y_{55}=124,6.

– с учетом автокорреляции с ρ = 0,993 (метод Кохрана-Оркатта).

– по авторегрессионному преобразованию (метод Хилдрета-Лу).

– по авторегрессионному преобразованию (метод Прайса-Винстена).

Используем последнее наблюдение для контроля точности прогнозов.

12. Первый вариант прогноза. Предварительно подготовим данные: Выборка / Установить диапазон (рис. 5.34):

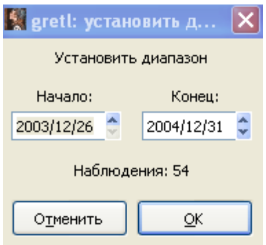


Рис. 5.34. Окно диапазона исходных данных

Построение уравнения регрессии золотовалютных резервов РФ в зависимости от времени: Модель / Метод наименьших квадратов…(рис. 5.35)

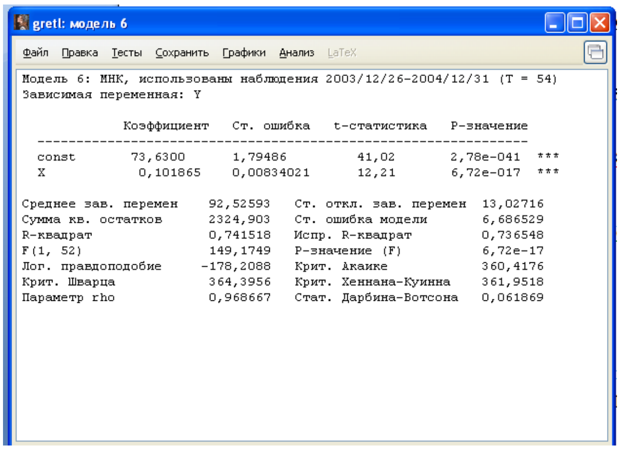


Рис. 5.35. Регрессия по исходным данным

Получим модель: Y=73,63+0,102X+ε. DW=0,06<dL, присутствует положительная автокорреляция. Строим прогноз (рис. 5.36):

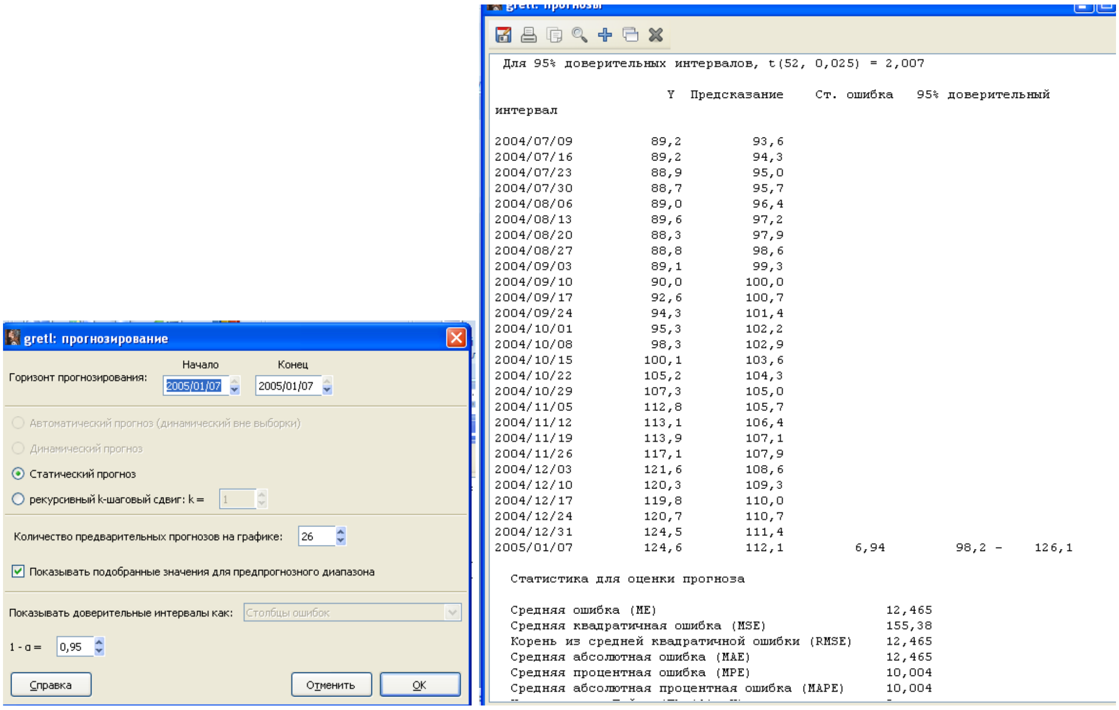


Рис. 5.36. Результаты прогноза по исходным данным

Прогнозное значение Y55=112,1 Доверительный интервал: 98,2; 126,1.

Средняя абсолютная процентная ошибка прогноза (MAPE): 1,11%.

Относительная ошибка прогноза: $OE = \frac{|y_{55} - \hat{y}_{55}|}{y_{55}} = 0{,}1(10\%)$.

Представим расчет относительной ошибки прогноза в меню Gretl: Вид / Сессия / Скаляр.

Ввод скалярных величин: Y55=124,6; YR55=112,1; OE=abs(Y55-YR55) / Y55 (рис. 5.37).

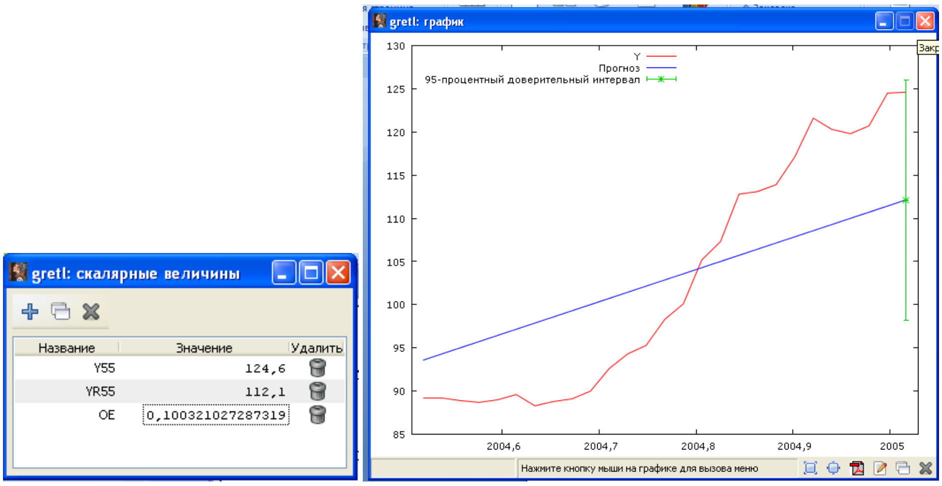


Рис. 5.37. Окно ввода скалярных величин и доверительный интервал (первый вариант прогноза)

13. Второй вариант прогноза (метод Кохрана-Оркатта) :

$\hat{y}*_{55} = \hat{y}_{55} + 0{,}993 \cdot (y_{54} - \hat{y}_{54})$

Для прогноза: Вид/ Сессия/ Скаляр: p=0,993; Y54=124,5; YR54=111,4; YR55_2=YR55+p*(Y54-YR54) (рис. 5.38).

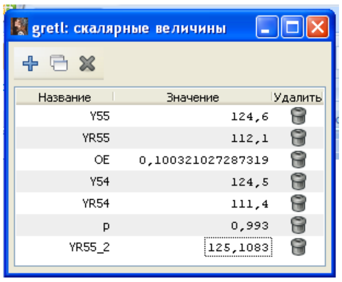


Рис. 5.38. Окно ввода скалярных величин (второй вариант прогноза)

Прогнозное значение Y55=125,1.

Относительная ошибка прогноза: OE=abs(Y55-YR55_2)/Y55.

OE_2 =0,004 или 0, 4%.

По сравнению с первым вариантом относительная ошибка прогноза уменьшилась в 25 раз.

14. Третий вариант прогноза (метод Хилдрета-Лу). Проводим авторегрессионное преобразование: Модель/ Временные ряды/ Кохрана-Оркатта (рис. 5.39):

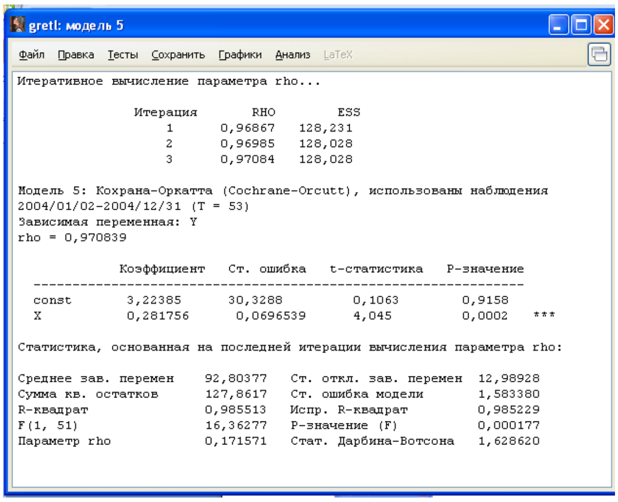


Рис. 5.39. Регрессия с учетом поправки (метод Хилдрета-Лу)

Получим модель: Yt=3,22+0,202*t+ε

Далее в окне модели: Анализ / Прогнозы (рис. 5.40):

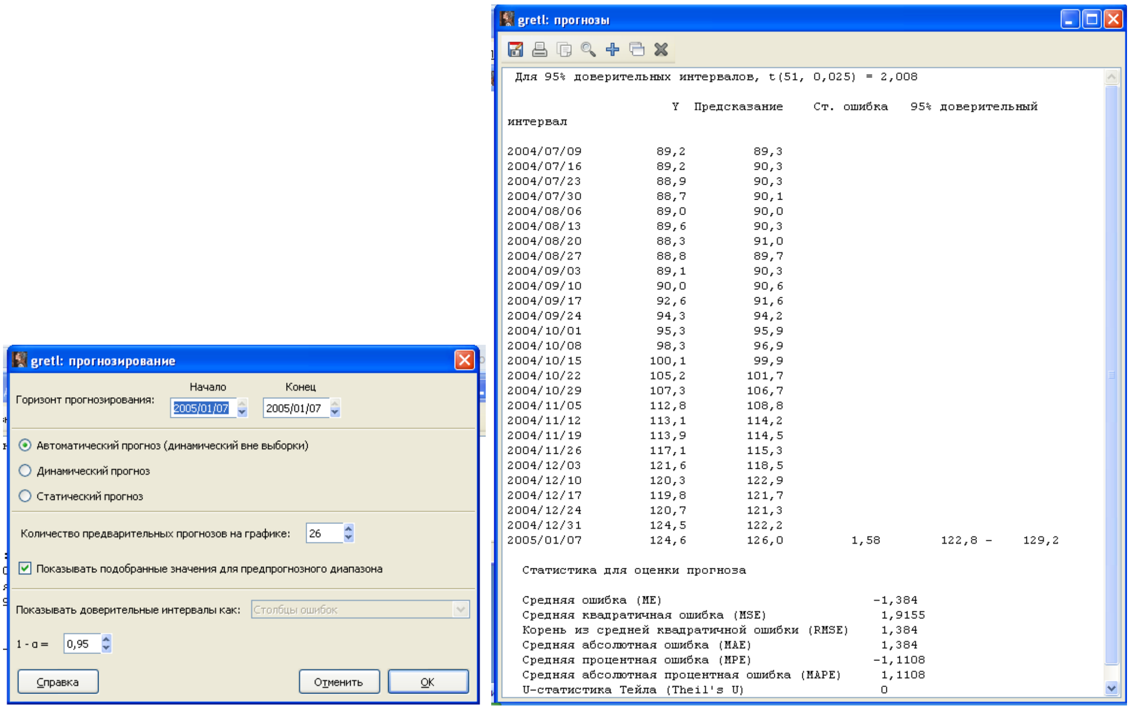


Рис. 5.40. Результаты прогноза (метод Хилдрета-Лу)

Прогноз на 07.01.2005 – 126,0. Доверительный интервал: [122,8; 129,2].

Средняя абсолютная процентная ошибка прогноза (MAPE): 1,11%.

Для расчета относительной ошибки прогноза OE_3: Вид / Сессия / Скаляр: Ввод скалярных величин: YR55_3=126,0; OE_3=abs(Y55-YR55_3) / Y55 (рис. 4.41).

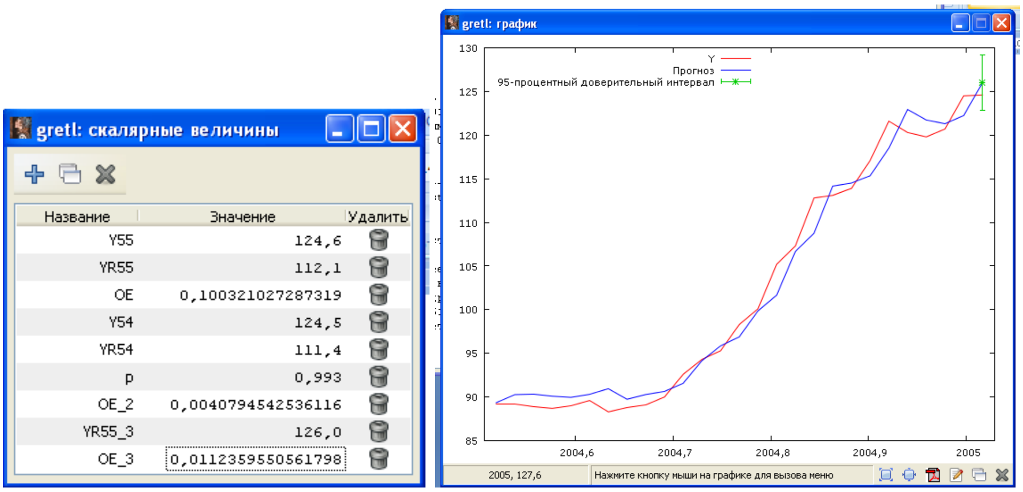
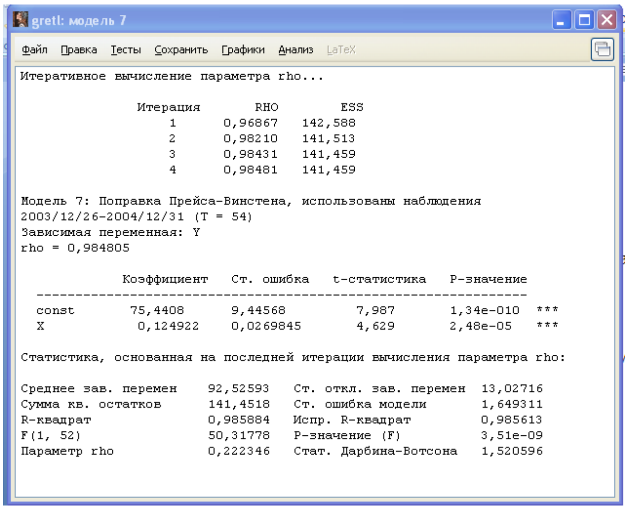


Рис. 4.41. Окно ввода скалярных величин и доверительный интервал (метод Хилдрета-Лу)

Относительная ошибка прогноза: OE_3=0,0112 или (1,12%).

15. Четвертый вариант прогноза. Проводим авторегрессионное преобразование: Модель / Временные ряды / Прайса-Винстена (рис.5.42):



```
gretl: модель 7
Файл  Правка  Тесты  Сохранить  Графики  Анализ  LaTeX

Итеративное вычисление параметра rho...

               Итерация       RHO        ESS
                    1       0,96867   142,588
                    2       0,98210   141,513
                    3       0,98431   141,459
                    4       0,98481   141,459

Модель 7: Поправка Прейса-Винстена, использованы наблюдения
2003/12/26-2004/12/31 (T = 54)
Зависимая переменная: Y
rho = 0,984805

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const        75,4408      9,44568         7,987       1,34e-010  ***
  X             0,124922    0,0269845       4,629       2,48e-05   ***

Статистика, основанная на последней итерации вычисления параметра rho:

Среднее зав. перемен    92,52593   Ст. откл. зав. перемен  13,02716
Сумма кв. остатков      141,4518   Ст. ошибка модели       1,649311
R-квадрат               0,985884   Испр. R-квадрат         0,985613
F(1, 52)                50,31778   Р-значение (F)          3,51e-09
Параметр rho            0,222346   Стат. Дарбина-Вотсона   1,520596
```

Рис. 5.42. Регрессия с учетом поправки (метод Прайса-Винстена)

Получим модель: Y=75,44+0,125X+ε

Результаты анализа точностных свойств представлены в окне модели: Анализ / Прогнозы (рис. 5.43):

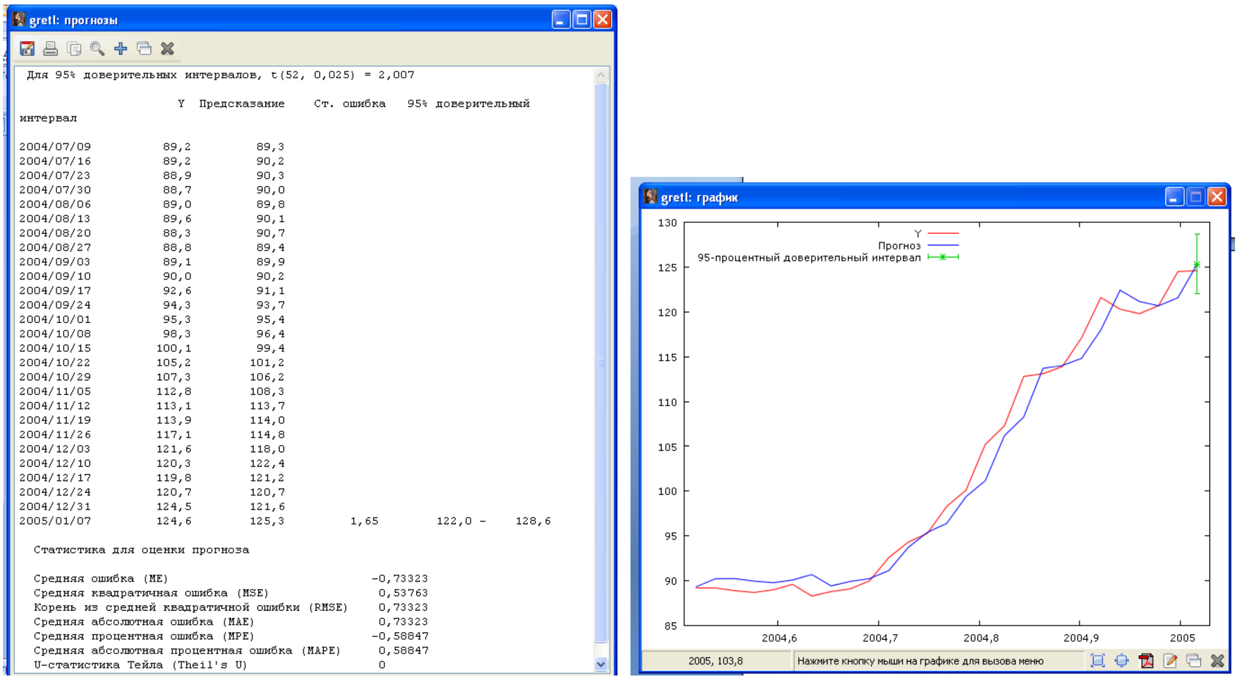


Рис. 5.43. Результаты прогноза (метод Прайса-Винстена)

Прогнозное значение: Y55=125,3. Доверительный интервал: [122,0;128,6].

Средняя абсолютная процентная ошибка прогноза (MAPE): 0,588%.

Для расчета относительной ошибки прогноза ОЕ: Вид/ Сессия/ Скаляр: Ввод скалярных величин: YR55_4=125,3; OE_4=abs(Y55-YR55_4)/Y55 (рис. 5.44). OE_4=0,006 или 0,6%.

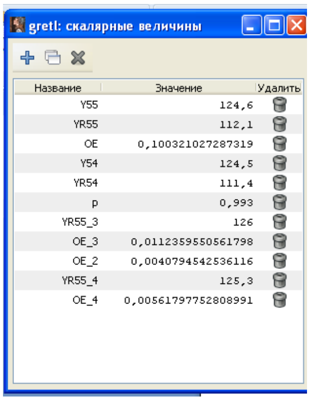


Рис. 5.44. Окно ввода скалярных величин (метод Прайса-Винстена)

Полученные результаты прогнозирования обобщены в таблице 5.6.

Таблица 5.6

Сводная таблица результатов прогноза

Тип прогноза	Вид уравнения	W	Прогнозная оценка, млрд. долл.	Относительная ошибка прогноза, %
МНК по исходным данным	Y=73,63+0,102X+ε	0,06	112,1	10
МНК с поправкой Кохрана-Оркатта	$\hat{y}*_{55} = \hat{y}_{55} + 0{,}993 \cdot (y_{54} - \hat{y}_{54})$	1,77	125,1	0,4
АР методом Кохрана-Оркатта	Y=3,22+0,202X+ε	1,63	126,0	1,12
АР методом Прайса-Винстена	Y=75,44+0,125X+ε	1,52	125,3	0,6

Таким образом, исходя из относительной ошибки прогноза наиболее приемлемы прогнозные оценки, полученные методом Кохрана-Оркатта и методом Прайса-Винстена.

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. (http://www.znanium.com/bookread.php?book=196548)

2. Айвазян С.А. Эконометрика - 2: продвинутый курс с приложениями в финансах: Учеб. / С.А.Айвазян, Д. Фантаццини; Московская школа экономики МГУ им. М.В. Ломоносова (МШЭ) - М.: Магистр: НИЦ ИНФРА-М, 2014. - 944 с.: 70x100 1/32. (п) ISBN 978-5-9776-0333-1, 100 экз. (http://znanium.com/catalog.php?bookinfo=472607)

3. Берндт, Э. Р. Практика эконометрики: классика и современность [Электронный ресурс]: Учебник для студентов вузов/ Э. Р. Берндт; пер. с англ. под ред. проф. С. А. Айвазяна. - М.: ЮНИТИ-ДАНА, 2012. - 863 с. (Серия «Зарубежный учебник») - ISBN 0-201-17628-9 (англ.), ISBN 5-238-00859-7 (русск.).(http://www.znanium.com/bookread.php?book=389506)

4. Дайитбегов Д.М. Компьютерные технологии анализа данных в эконометрике: Монография / Д.М. Дайитбегов. - 3-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с.: 70x100 1/16. - (Научная книга). (переплет) ISBN 978-5-9558-0275-6, 500 экз. (http://www.znanium.com/bookread.php?book=365692)

5. Елисеева И. И. Эконометрика: Учебник. - М.: Юрайт, серия "Магистр", 2012. - 464 с.

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

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

8. Плохотников К.Э. Основы эконометрики в пакете STATISTICA.: Учебное пособие / К.Э. Плохотников. - М.: Вузовский учебник, 2010. - 298 с.: 60x90 1/16 + CDROM. (переплет) ISBN 978-5-9558-0114-8, 2000 экз. (http://www.znanium.com/bookread.php?book=177719)

9. Wooldridge J. M. Introductory Econometrics. A modern approach, 5th edition. Michigan State University: South-Western Cengage Learning, 2013. 909p.

10. Gujarati D. N. Basic Econometrics. 5th ed. MacGraw-Hill, 2009.

11. Hill R. C., Griffiths W. E., Lim G. C. Principles of Econometrics. 4th ed. Wiley, 2011.

12. Green W.H. Econometric analysis. 7th ed. Prentice Hall, 2011.- 1231 p.