



Молчанов И.Н., Герасимова И.А.

**КОМПЬЮТЕРНЫЙ ПРАКТИКУМ ПО
НАЧАЛЬНОМУ КУРСУ ЭКОНОМЕТРИКИ
(РЕАЛИЗАЦИЯ НА EViews)**

Практикум

Ростов-на-Дону
2001

УДК [330.43](076.5)

М 75

1Л4

Молчанов И.Н., Герасимова И.А. Компьютерный практикум по начальному курсу эконометрики (реализация на Eviews): Практикум /Ростовский государственный экономический университет. - Ростов-н/Д., - 2001. – 58 с. - ISBN 5-7972-0377-4.

Практикум представляет собой попытку создания учебного пособия, ориентированного на специфику преподавания эконометрики в экономическом вузе с использованием специализированного эконометрического пакета Eviews. Практикум ориентирован на начальный курс эконометрики.

Для студентов и аспирантов вузов, обучающихся по экономическим специальностям.

Замечания и предложения просим направлять по адресу:

344007, г.Ростов-на-Дону, ул. Большая Садовая, 69, к. 404, каф. СМиП.

E-mail: **IGORM@APPLECLUB.DONPAC.RU**

Интернет: **HTTP://MOLCHANOV.NAROD.RU/ECONOMETRICS.HTML**

Рецензенты:

Л.И.Ниворожкина, доктор экономических наук, профессор, зав. кафедрой СМиП РГЭУ «РИНХ».

В.С.Князевский, доктор экономических наук, профессор, Заслуженный деятель науки Российской Федерации, РГЭУ «РИНХ».

Утверждено в качестве практикума редакционно-издательским советом РГЭУ

ISBN 5-7972-0377-4

© Ростовский государственный экономический университет «РИНХ», 2001

© Молчанов И.Н., Герасимова И.А., 2001

Предисловие

Эконометрический пакет Eviews обеспечивает особо сложный и тонкий инструментарий обработки данных, позволяет выполнять регрессионный анализ, строить прогнозы в Windows-ориентированной компьютерной среде. С помощью этого программного средства можно очень быстро выявить наличие статистической зависимости в анализируемых данных и затем, используя полученные взаимосвязи, сделать прогноз изучаемых показателей.

Целесообразно выделить следующие сферы применения Eviews:

- анализ научной информации и оценивание;
- финансовый анализ;
- макроэкономическое прогнозирование;
- моделирование;
- прогнозирование состояния рынков.

Особо широкие возможности открывает Eviews при анализе данных, представленных в виде временных рядов.

Подробную информацию об условиях приобретения и распространения пакета можно получить на сайте производителя: <http://www.eviews.com>. Пакет занимает после инсталляции около 12 Мб на жестком диске.

Все используемые в практикуме задания (примеры) доступны в виде файлов в формате Excel и Eviews по адресу: <http://molchanov.narod.ru/econometrics.html>.

При выполнении предлагаемых заданий могут оказаться полезными следующие учебники и пособия:

1. **Айвазян С.А., Мхитарян В.С.** Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998. – 1022 с. ISBN 5-238-00013-8.
2. **Доугерти К.** Введение в эконометрику. - М.: ИНФРА-М, 1997. – XIV, 402 с.: ил. - (Университетский учебник) Библиография: с.384-386. ISBN 5-86225-458-7; 0-19-50346-4.
3. **Елисеева И.И.** Эконометрика: Учебник /И.И.Елисеева и др. – М.: Финансы и статистика, 2001. – ISBN 5-279-01955-0.
4. **Князевский В.С., Житников И.В.** Анализ временных рядов и прогнозирование: Учеб. пособие. – Ростов-на-Дону: РГЭА, 1998. – 161 с.
5. **Князевский В.С., Молчанов И.Н.** Статистические расчеты на компьютере с использованием ППП Microstat. - Ростов-на-Дону: РГЭА, 1996. - 86 с.
6. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика. Начальный курс. – М.: Дело, 2000. – 400 с. ISBN 5-7749-0055-X.
7. Практикум по эконометрике: Учеб. пособие /И.И.Елисеева и др. – М.: Финансы и статистика, 2001. – 192 с. ISBN 5-279-02313-2.
8. **Greene, W.H.** *Econometric analysis*, Prentice Hall, 4th Edition, 2000. – 1004 p.
9. **Verbeek, M.** *A Guide to Modern Econometrics*, Wiley, 2000. – 400 p.

КОМПЬЮТЕРНЫЙ ПРАКТИКУМ ПО НАЧАЛЬНОМУ КУРСУ ЭКОНОМЕТРИКИ (РЕАЛИЗАЦИЯ НА EIEWS)

Практикум

Молчанов Игорь Николаевич
Герасимова Ирина Алексеевна

Ответственная за выпуск
Начальник РИО РГЭУ В.Е. Смейле
Редактирование и корректура авторов
Оригинал-макет И.Н.Молчанов

Лицензия ЛР N 020276 от 18.02.97
Государственного Комитета Российской Федерации по печати

Изд. № 65/5309	Подписано к печати	28.02.2001.
Бумага офсетная.	Печать офсетная.	Формат 60•84/16
Объем 4,0 уч.-изд.л.	Тираж 100 экз.	Заказ № «С» 65

344007, Ростов-на-Дону, ул. Большая Садовая, 69, РГЭУ «РИНХ», Издательство

Отпечатано в копировально-множительном центре.
Ростов-на-Дону, ул. Большая Садовая, 79. ПБОЮЛ Зайчиков О.Б.

Практическое занятие № 1.

«Знакомство с эконометрическим пакетом Eviews»

Eviews (далее **пакет**) установлен в директорий **Program Files/Eviews3**. Запуск осуществляется выбором соответствующего значка в панели Пуск/Программы/**Eviews3/Eviews 3.1** (файл C:\Program Files\EViews3\EViews3.exe) (см. рис. 1) или щелчком (двойным щелчком – в зависимости от установок) по соответствующей пиктограмме на рабочем столе.

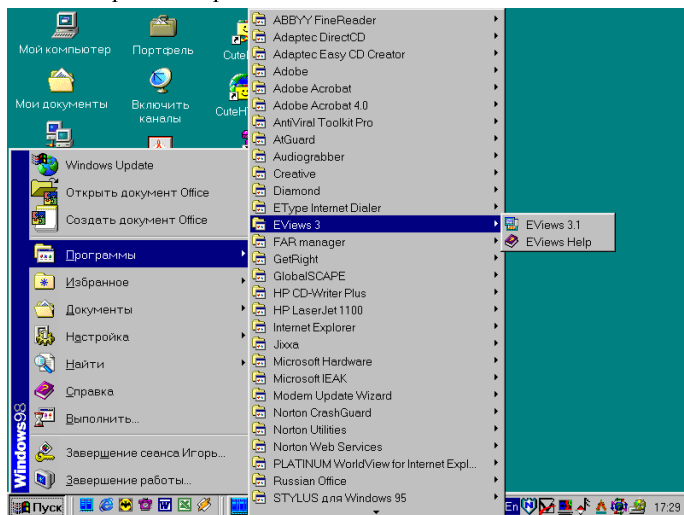


Рис. 1.

Если Вы все сделали правильно, появится стартовое окно пакета (рис.2).

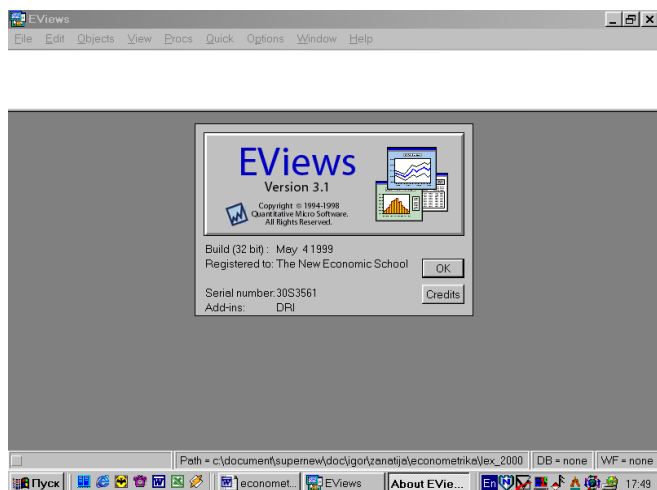


Рис. 2.

Если в настоящий момент окно, содержащее пакет, является **активным**, то первая строка экрана (**Title Bar**) будет темнее остальных. При переключении в другое окно цветовая окраска данной строки изменит цвет на более приглушенный (серый).

Ниже следует строка **основного меню (Main Menu)**. Принцип его построения прост – при нажатии на соответствующие клавиши появляется раскрывающееся меню (**drop-down menu**). Доступные в настоящий момент опции являются затемненными (**darkened menu items**). Те пункты, с которыми в настоящий момент работа невозможна, приглушены (**grayed menu items**).

Далее располагается **командная строка (окно) (command window)**. В нем происходит непосредственный набор команд, которые выполняются после нажатия клавиши **Enter (Ввод)**. Для исполнения многих команд отсутствует необходимость их набора – просто надо выбрать нужный пункт в основном меню.

Большая часть экрана пакета отведена под **рабочую область (work area)**. В ней размещаются рабочие объекты. Переключение между ними осуществляется нажатием клавиши **F6**.

Последняя область экрана показывает **текущее состояние (status line)** пакета (рабочий каталог, текущий файл и др.).

Завершение работы с пакетом осуществляется путем выбора в командной строке опции **File/Exit**. Система предложит сохранить/не сохранить имеющиеся данные. Если имя файла не было задано ранее, автоматически будет предложено имя **UNTITLED**. Его можно изменить на любое другое. Пакет имеет обширную **справочную систему** (пункт основного меню **Help**).

Знакомство с пакетом начнем с файла, содержащего данные о совокупном спросе на деньги (M1) – (aggregate money demand) (**M1**) – **зависимая переменная; независимые:** доход (ВВП) - income (**GDP**); уровень цен (PR) - price level (**PR**); краткосрочная процентная ставка (RS) - short term interest rate (**RS**).

Проведем некоторые преобразования и расчеты.

Первым шагом создадим новый рабочий файл (**workfile**). Его имя должно иметь следующий вид и состоять **только** из латинских букв: **Номер_группы_demo_01.wf1** (расширение wf1 присваивается автоматически). *Например: 451_demo_01.wf1*. Расположить его следует в директории, относящейся к Вашему факультету (**внимательно ознакомьтесь с памяткой в компьютерном классе**). Исходные данные находятся в файле **Excel**. Они должны быть импортированы в пакет. Создание рабочего файла начнем с того, что выберем **File/New/Workfile** в основном меню (см. рис. 3).

После нажатия на кнопке со словом **Workfile** откроется диалоговое окно, с помощью которого можно задать тип вводимых Вами данных (см. рис. 4).

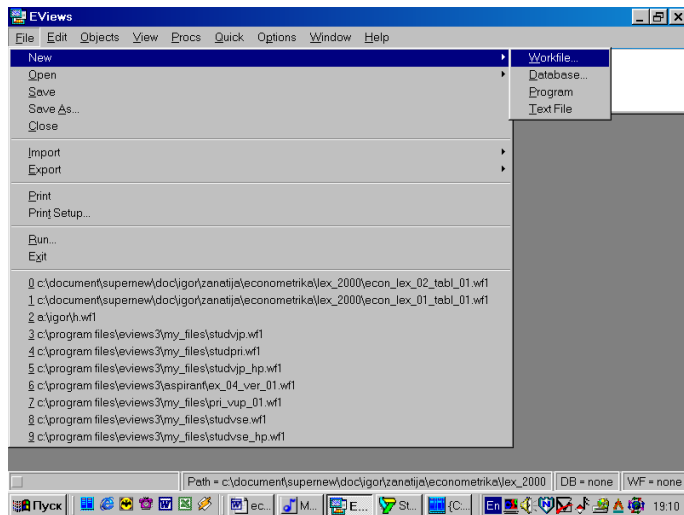


Рис. 3.

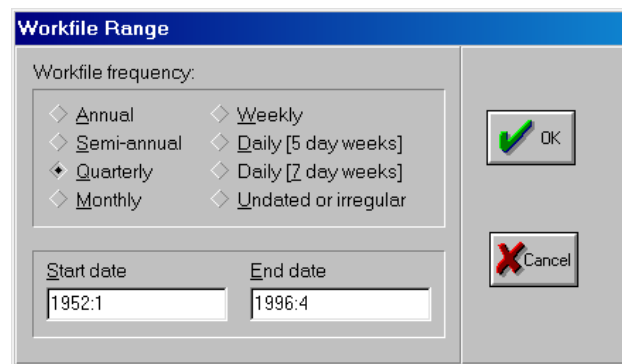


Рис. 4.

Как видно из рис. 4, в пакете допускается восемь типов данных. Это могут быть:

- **Годовые (Annual)** – годы 20 в. идентифицируются по последним двум цифрам (97 эквивалентно 1997), для данных, относящихся к 21 в. необходима полная идентификация (например, 2020);
- **Полугодовые (Semi-annual)** – 1999:1, 2001:2 (формат – год и номер полугодия);
- **Квартальные (Quarterly)** – 1992:1, 65:4, 2005:3 (формат – год и номер квартала);
- **Ежемесячные (Monthly)** – 1956:1, 1990:11 (формат – год и номер месяца);
- **Недельные (Weekly)** и **дневные (5/7 day weeks)** – допускаются форматы Месяц/День/Год (по умолчанию) и (День/Месяц/Год) – настроить эту опцию

можно в меню **Options/Frequency Conversion & Date Display**. Так, введенные числа 8:10:97 будут интерпретированы как Август, 10, 1997. Для установки, принятой в Европе, начальная дата будет выглядеть как Октябрь, 8, 1997;

- **Недатированные или нерегулярные (Undated or irregular)** – допускают работу с данными, строго не привязанными к определенным временным периодам.

Важным является указание **начальной (start) и конечной (end) даты/наблюдения (date/observation)**.

В нашем примере начальным периодом является первый квартал 1952 г. (1952:1), конечным – четвертый квартал 1996 г. (1996:4).

Закончив ввод временных периодов, надо нажать клавишу **ОК**. Пакет создаст рабочий файл без имени, и на дисплее в рабочей области появится окно (см. рис. 5). Все рабочие файлы пакета всегда содержат вектор коэффициентов **C** и серию **RESID**.

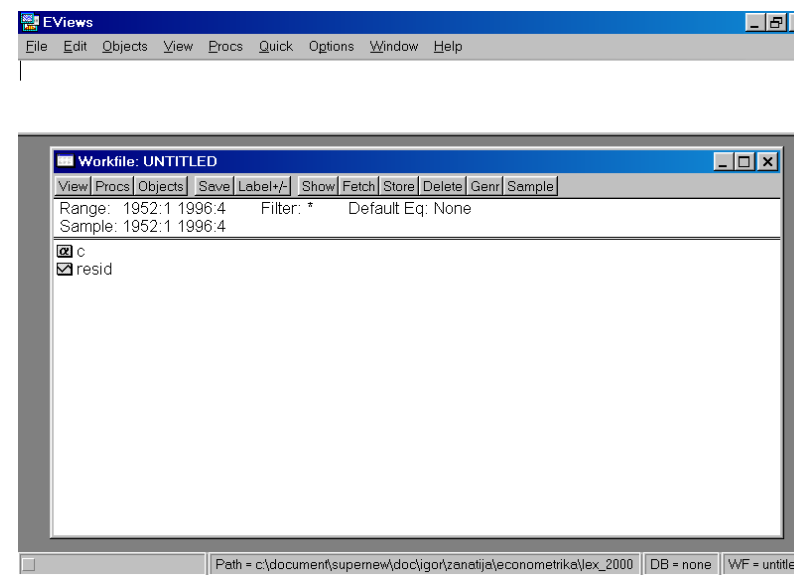


Рис. 5.

Следующим шагом является просмотр исходных данных, содержащихся в исходном файле по адресу **Program Files/Eviews3/Example files/demo.xls** (формат Excel версии 5.0 и младше). **Важное замечание: имеющаяся версия пакета позволяет импортировать файлы Excel не старше версии 5.0. В противном случае будет выдано сообщение об ошибке. Всегда сохраняйте свои файлы как файлы Microsoft Excel 5.0/95.** Для визуализации данных необходимо запустить табличный процессор Excel (действия аналогичны запуску Eviews). Результат представлен на рис. 6. Ознакомившись с данными, файл, подлежащий экспортированию, **необходимо закрыть**.

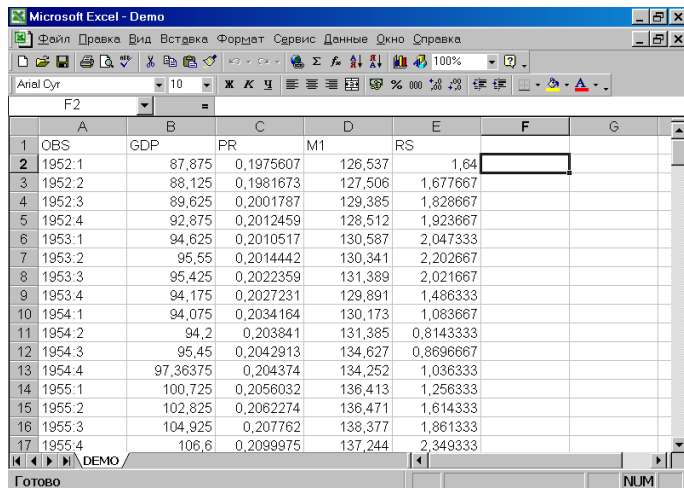


Рис. 6.

Для чтения данных, созданных в других программах, надо выбрать в рабочем файле опцию Procs/Import/Read Text-Lotus-Excel... (см. рис. 7). Появится диалог, представленный на рис. 8.

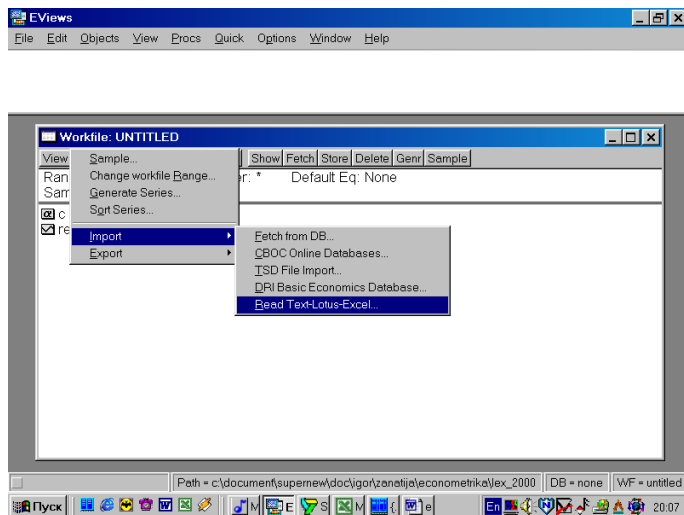


Рис. 7.

Перейдем к папке, содержащей искомый файл (для упрощения поиска в опции **Тип файлов** (Files of type) можно выбрать **Excel.xls** (см. рис.8). Для того, чтобы пакет «помнил» Ваши перемещения по папкам компьютера, можно поставить флажок в опции **Update default directory** (см. рис. 8).

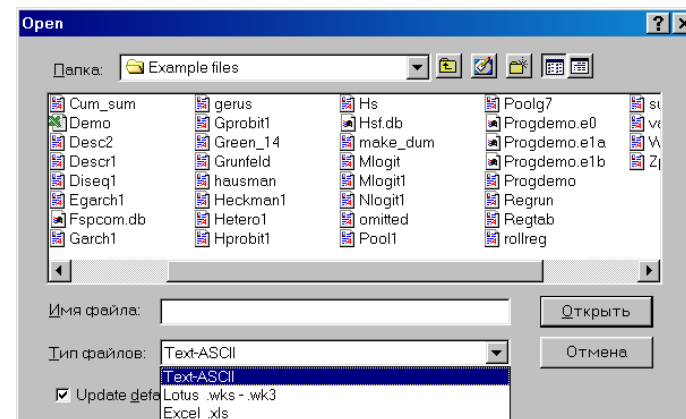


Рис. 8.

Наведем курсор на файл **demo.xls** и нажмем кнопку **Открыть** (см. рис. 8). Появится диалог открытия электронных таблиц формата Excel (см. рис. 9).

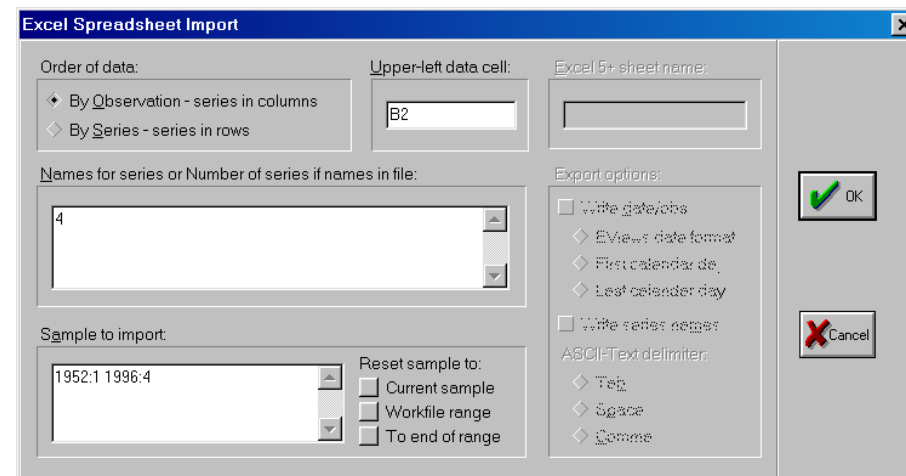


Рис. 9.

По умолчанию в окне, представленном на рис. 9, предполагается, что данные находятся в столбцах (**by observation - series in columns**). Если данные представлены в виде серий в строках, то надо отметить другую опцию (**By series - series in rows**).

Окно **Upper-left data cell** (левая верхняя ячейка данных) автоматически отобразило клетку **B2**. Это означает, что данные будут импортироваться из исходной таблицы с клетки, указанной в этом окне (тем самым первая строка и первый столбец будут пропущены). Это вполне соответствует структуре нашего исходного файла (см. рис. 6). Иногда приходится исправлять адрес такой клетки на актуальный.

В окне **Names for series or Number of series if names in file** (имена для серий или

число серий, если имена содержатся в файле) указываем цифру 4. Это связано с тем, что исходный файл (см. рис.6) содержит 4 переменные, находящиеся в столбцах. Имена для этих переменных будут взяты из первой строки электронной таблицы (клетки B1:E1). В том случае, когда необходимо импортировать часть данных (например, только первые две переменные), надо ввести их количество (цифра 2). Если имена переменных, по каким либо причинам, в исходном файле не заданы, можно вместо цифр ввести их имена (латинскими буквами). Если количество переменных, введенных в рассматриваемом окне, превышает количество реально существующих, то в рабочий файл будет введен столбец с заданным именем без данных (обозначаются такие клетки как NA). Если все другие установки удовлетворяют заданным Вами условиям, то можно нажать кнопку **ОК**. Появится окно, отображенное на рис. 10.

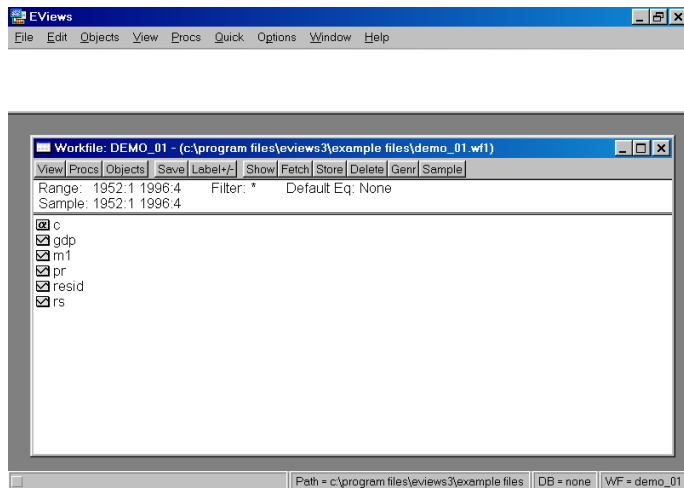


Рис. 10.

После того, как исходные данные перенесены Вами в рабочую область пакета (появились имена переменных), надо провести их верификацию (проверку правильности). Вам необходимо создать новую группу, содержащую все импортированные серии (переменные). Это делается следующим образом: необходимо кликнуть мышкой по имени первой переменной (например, GNP), затем, удерживая клавишу **CTRL** кликнуть по переменным M1, PR и RS. Все серии на экране будут зачернены. Затем необходимо подвести курсор мыши на зачерненную область экрана и кликнуть правой кнопкой. Далее необходимо выбрать опцию **Open**. Пакет откроет диалоговое окно со следующими опциями (см. рис. 11).

Выберем **Open Group** (открыть в одной группе). Пакет создаст группу с именем UNTITLED, в которую войдут все переменные (серии). По умолчанию, данные будут представлены в виде электронной таблицы (возможны другие варианты представления) – см. рис. 12.

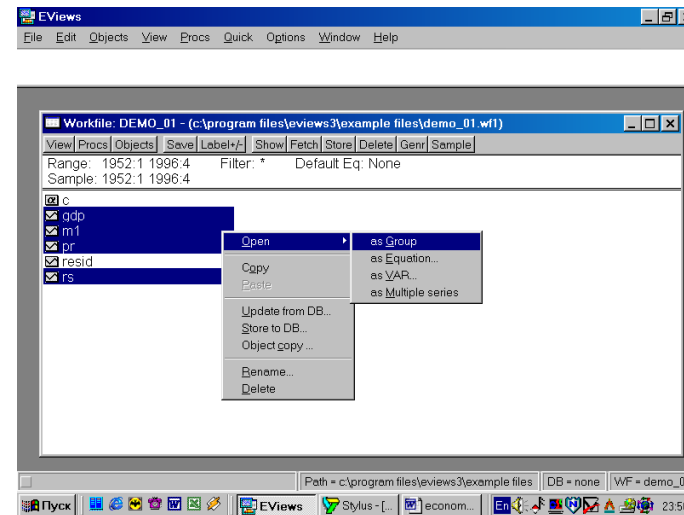


Рис. 11.

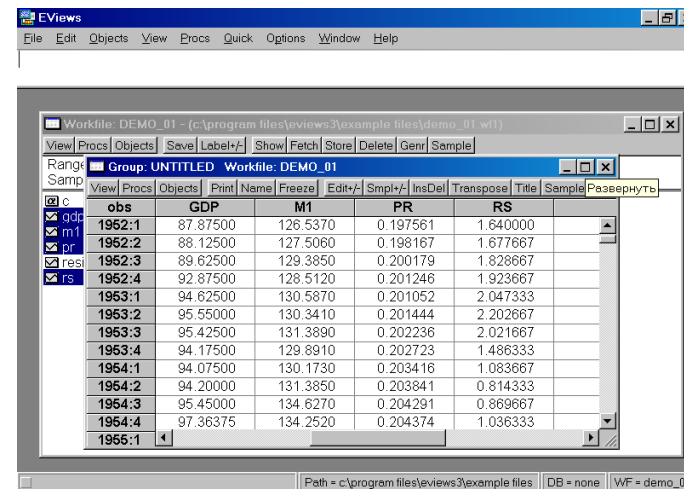


Рис. 12.

Проведите визуальную проверку корректности данных. Сравните, как разместились переменные из исходного файла, обратите внимание на столбец слева от первой переменной (он серого цвета). В нем отображены годы и порядковые номера кварталов. Полученной новой группе данных можно дать имя. Для этого необходимо нажать кнопку **Name** в текущем окне (см. рис. 12). Появится диалоговое окно (рис. 13.). Автоматически будет предложено имя – **GROUP01**. Его можно принять, нажав кнопку **ОК**. В рабочем файле сразу добавится одна переменная с введенным Вами именем. Теперь к ней всегда можно перейти простым нажатием клавиши мыши.

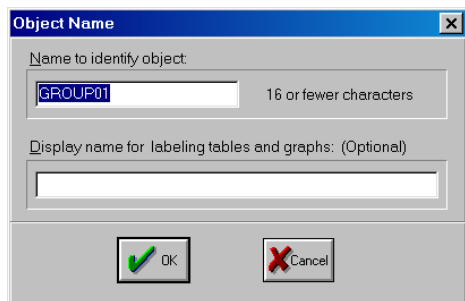


Рис. 13.

Образованную Вами группу можно просматривать не только в виде электронной таблицы. Если, находясь внутри **GROUP01**, выбрать последовательность команд **View/Multiple Graphs/Line** (см. рис. 14), то данные предстанут не в виде таблицы, а как линейные графики по каждой серии (переменной) – см. рис. 15.

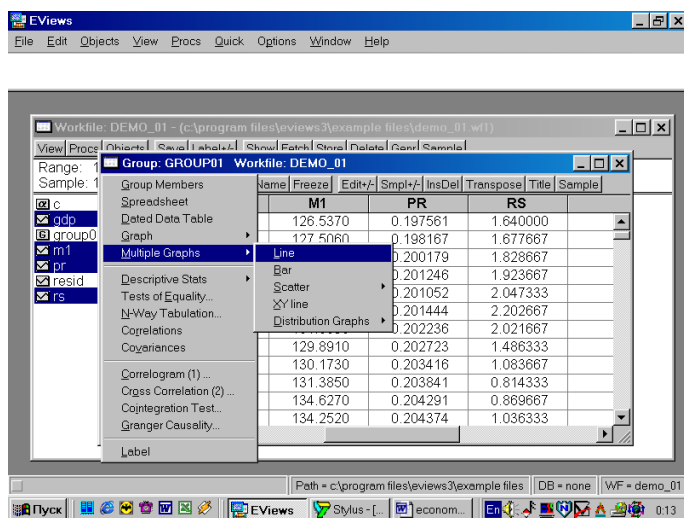


Рис. 14.

Для того, чтобы вернуться к прежней форме представления данных (например, электронной таблице), надо выбрать **View/Spreadsheet**.

Для просмотра числовых характеристик (описательных статистик) отмеченных переменных необходимо выбрать в рабочем файле **View/Descriptive Stats/Individual Samples** (см. рис. 16).

В результате появится окно, представленное на рис. 17. В нем содержатся:

Mean – Среднее арифметическое значение;

Median – Медиана;

Maximum – Максимальное значение;

Minimum – Минимальное значение;

Std. Dev. – Стандартное отклонение (среднее квадратическое отклонение);

Skewness – Коэффициент асимметрии;

Kurtosis – Экцесс;

Probability – Вероятность;

Observations – Количество наблюдений.

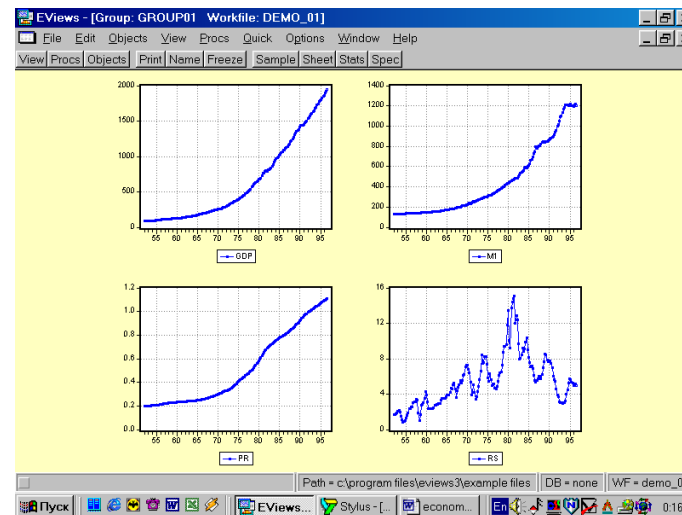


Рис. 15.

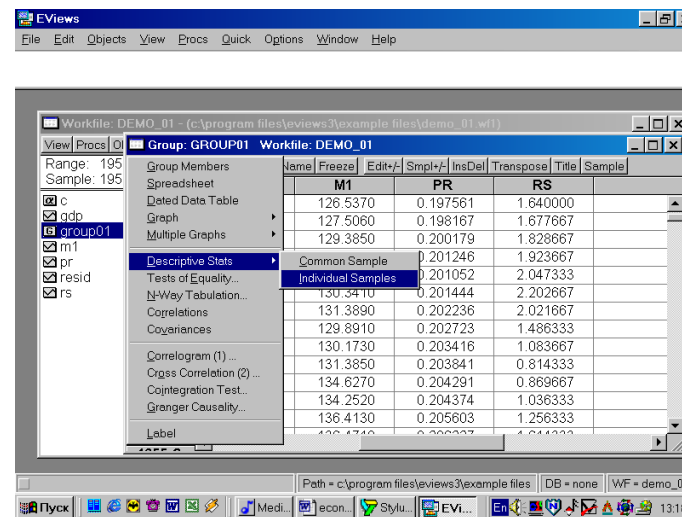


Рис. 16.

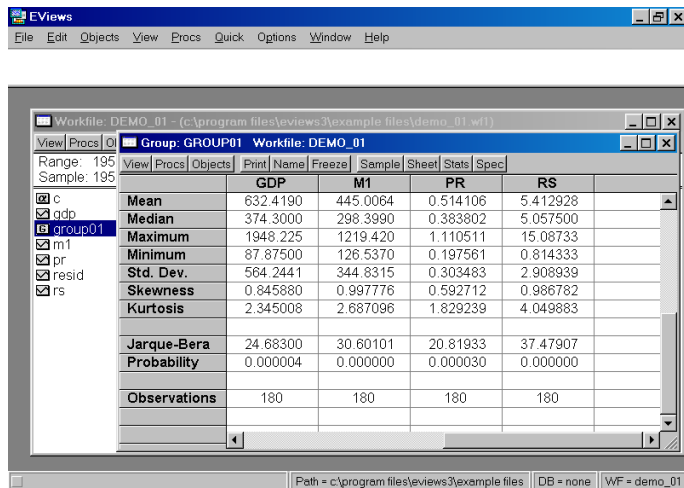


Рис. 17.

Если возникает необходимость проанализировать матрицу коэффициентов корреляции, то необходимо выбрать **View/Correlations**. Результат представлен на рис. 18.

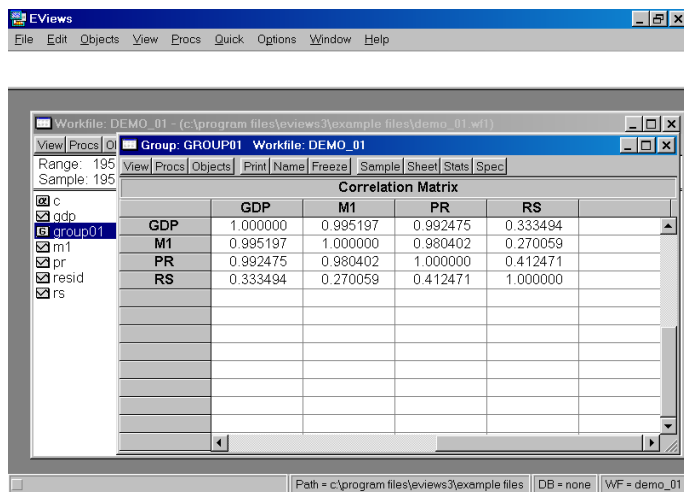


Рис. 18.

Вы также можете исследовать характеристики для отдельных серий (переменных), совместив вывод диаграммы и числовых характеристик. Дважды кликните на имени серии (например, на переменной M1) и выберите в рабочем файле пункт меню **View/Descriptive Stats/Histogram and Stats** (см. рис. 19). Результат наглядно виден на рис. 20.

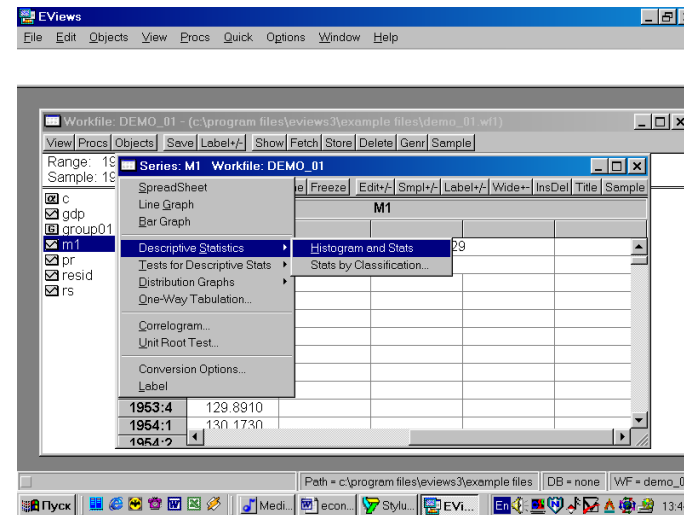


Рис. 19.

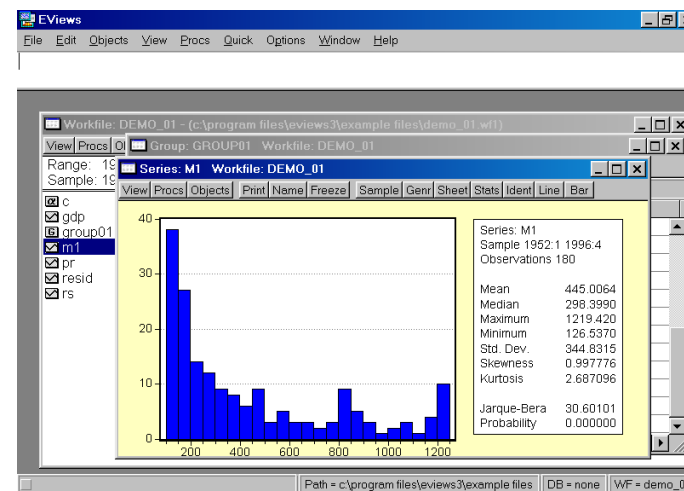


Рис. 20.

С другими возможностями пакета Вы познакомитесь на последующих занятиях. Для индивидуальной работы по предложенной выше схеме предназначены ниже следующие данные. **Подумайте**, все ли данные необходимо заносить в электронную таблицу или импортировать из неё.

Пример 1. Стоимость однокомнатных квартир в Москве [6].

Данные из газеты «Из рук в руки» за период с декабря 1996 г. по сентябрь 1997г.

Была выбрана Юго-Западная часть города, в которой высок спрос на жилые площади (всего 69 наблюдений). Файл **example_01.xls**.

Переменные:

- N** - Номер по порядку.
distc Удаленность. от центра, км.
distm Удаленность от метро, мин.
totsq Общая площадь квартиры, кв.м.
kitsq Площадь кухни, кв.м.
livsq Площадь комнаты, кв.м.
floor Этаж. 0-первый/последний, 1-нет.
cat Категория дома. 1-кирпичный, 0-нет.
price Цена квартиры, тыс. USD.

Найдите среднее арифметическое, выборочное стандартное отклонение и другие статистики параметров. Найдите коэффициенты корреляции параметров с ценой квартиры. Соответствуют ли полученные значения экономической интуиции?

N	region	distc	distm	totsq	kitsq	livsq	floor	cat	price
1	Фрунзенская	4	10	34,00	7,50	19,00	1	1	54
2	Ленинский пр.	5,7	7	36,00	10,00	20,00	0	0	35
3	Ленинский пр.	5,7	12	45,00	13,00	20,00	1	1	59
4	Академическая	7,6	10	35,30	10,00	20,00	1	0	35
5	Университет	8,7	6	33,00	5,50	22,00	1	0	33
6	Нов Черемуш.	10,3	3	33,00	8,50	18,00	1	1	57
7	Юго-Западная	13,3	10	37,00	10,00	19,00	1	0	43
8	Коньково	14,8	2	38,00	8,50	19,10	1	0	39
9	Фрунзенская	4	15	54,00	9,20	27,20	1	1	70
10	Университет	8,7	15	35,00	6,00	20,00	0	1	43
11	Пр Вернадск.	11,4	10	31,40	5,20	21,30	1	0	33
12	Ленинский пр.	5,7	7	32,00	6,00	21,00	1	0	37
13	Нов Черемуш	10,3	7	38,00	8,00	19,00	0	0	33
14	Университет	8,7	10	31,60	8,80	14,00	0	0	31
15	Юго-Запад	13,3	5	32,00	8,00	17,00	1	0	37
16	Юго-Запад	13,3	10	37,00	10,00	19,00	1	0	43
17	Ленинский пр.	5,7	5	32,00	8,00	17,00	1	1	38
18	Академическая	7,6	10	37,00	8,00	19,00	1	1	51
19	Академическая	7,6	15	32,20	6,50	17,00	0	1	30
20	Коньково	14,8	3	33,00	8,00	19,00	1	0	30
21	Коньково	14,8	5	37,50	9,60	19,80	1	0	36
22	Коньково	14,8	10	33,00	7,00	19,00	1	0	33
23	Университет	8,7	15	32,00	6,00	21,50	1	0	35
24	Пр Вернадск.	11,4	5	29,70	6,00	16,10	0	0	28
25	Пр Вернадск.	11,4	15	36,00	8,60	18,00	0	0	40
26	Юго-Запад	13,3	15	36,00	10,00	19,00	0	0	33
27	Ленинский пр.	5,7	2	31,60	6,00	21,60	1	1	35
28	Ленинский пр.	5,7	5	52,00	12,00	34,00	1	1	75

N	region	distc	distm	totsq	kitsq	livsq	floor	cat	price
29	Коньково	14,8	3	36,00	10,00	19,00	1	0	40
30	Коньково	14,8	5	33,00	8,00	18,00	1	0	30
31	Университет	8,7	5	32,00	5,50	20,10	1	0	31
32	Академическая	7,6	15	35,00	9,80	20,00	1	0	37
33	Нов Черемуш	10,3	15	38,00	10,00	19,50	1	0	40
34	Коньково	14,8	1	39,00	8,50	19,00	1	0	40
35	Фрунзенская	4	5	34,00	8,00	19,00	1	1	58
36	Фрунзенская	4	10	38,00	6,50	18,00	0	1	48
37	пр Вернадск.	11,4	3	35,00	10,00	20,00	1	0	40
38	Юго-запад	13,3	7	36,00	9,00	19,50	1	0	42
39	Нов Черемуш.	10,3	7	34,00	8,00	18,00	1	1	51
40	Коньково	14,8	5	38,00	8,50	19,00	1	0	43
41	Коньково	14,8	7	33,00	6,00	19,00	1	0	30
42	Коньково	14,8	10	32,00	8,00	17,00	1	0	40
43	Коньково	14,8	10	38,00	8,50	19,10	1	0	43
44	Академическая	7,6	5	43,00	8,50	25,00	0	1	53
45	Академическая	7,6	10	30,00	6,00	18,30	1	1	28
46	Коньково	14,8	7	34,80	7,80	17,80	0	0	29
47	Коньково	14,8	15	35,00	10,00	19,60	1	0	37
48	Коньково	14,8	3	32,80	6,50	18,50	1	0	30
49	НовЧеремуш.	10,3	10	39,00	9,00	19,00	1	0	45
50	Университет	8,7	15	49,00	9,00	20,50	0	1	52
51	Фрунзенская	4	3	32,00	6,20	19,00	1	1	53
52	Пр Вернадск	11,4	10	33,00	6,50	19,00	1	0	32
53	Пр Вернадск	11,4	15	32,30	6,00	21,90	0	0	28
54	Юго-Запад	13,3	10	30,00	7,00	19,80	1	0	34
55	Юго-Запад	13,3	10	34,00	9,00	19,00	1	0	42
56	Юго-Запад	13,3	7	33,00	7,00	19,00	0	0	33
57	Академическая	7,6	10	30,00	6,00	18,30	1	1	28
58	Академическая	7,6	15	32,00	6,00	18,00	1	0	30
59	Коньково	14,8	5	33,10	7,50	18,00	1	0	32
60	Коньково	14,8	2	38,00	7,50	19,00	1	0	41
61	Коньково	14,8	7	38,00	8,60	19,00	1	0	43
62	Коньково	14,8	5	37,30	6,50	19,00	1	0	31
63	Ленинский пр.	5,7	8	31,40	5,60	21,00	1	0	33
64	Ленинский пр.	5,7	7	52,00	10,00	34,00	1	1	60
65	Нов Черемуш	10,3	15	30,00	6,00	17,00	1	1	37
66	Нов Черемуш	10,3	5	36,00	11,00	20,00	1	0	41
67	Пр Вернадск.	11,4	5	28,00	6,70	14,40	1	0	35
68	Пр Вернадск	11,4	10	31,40	5,20	21,30	1	0	33
69	Юго-Запад	13,3	5	32,00	8,00	17,00	1	0	37

В дальнейшем мы продолжим работу с этим файлом.

ВНИМАНИЕ!!! Все файлы, созданными Вами во время практического занятия, необходимо сохранять на личных дисках (желательно в двух экземплярах на разных носителях).

Практическое занятие № 2.

«Применение Eviews при построении и анализе линейной однофакторной модели регрессии»

Пример 2. Имеются следующие данные по 10 фермерским хозяйствам области:

№ п/п	1	2	3	4	5	6	7	8	9	10
Урожайность зерновых ц/га	15	12	17	21	25	20	24	14	23	13
Внесено удобрений на 1 га посевов, кг	4,0	2,5	5,0	5,8	7,5	5,7	7,0	3,0	6,0	3,5

Необходимо:

1. Создать файл с исходными данными в среде Excel (файл **example_02.xls**).
2. Осуществить импорт исходных данных в Eviews.
3. Создать workfile (рабочий файл).
4. Найти значения описательных статистик по каждой переменной и объяснить их.
5. Построить поле корреляции моделируемого (результативного) и факторного признаков. Объяснить полученные результаты.
6. Найти значение линейного коэффициента корреляции и пояснить его смысл.
7. Определить параметры уравнения парной регрессии и интерпретировать их. Объяснить смысл полученного уравнения регрессии.
8. Оценить статистическую значимость коэффициента регрессии b и уравнения в целом. Сделать выводы.
9. Объяснить полученное значение R^2 .
10. Построить эмпирическую и теоретическую линию регрессии и объяснить их.
11. Построить и проанализировать график остатков.
12. С вероятностью 0,95 построить доверительный интервал для ожидаемого значения урожайности \hat{y} по точечному значению x_i .
13. Оформить отчет по занятию.

Порядок выполнения задания

1. В Excel исходные данные должны быть организованы таким образом, чтобы в каждой колонке были представлены данные по соответствующей переменной (рис. 21). Имена переменных набираются латинскими буквами. Файл необходимо сохранить в формате **Excel 5.0/95** (рис. 22). Введем обозначения: урожайность зерновых – переменная **Productivity** (зависимая, Y); внесено удобрений на 1 га посевов – **Fertilizers** (независимая, X).

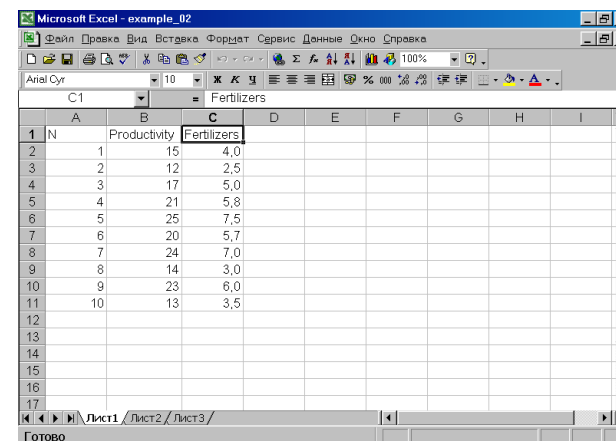


Рис. 21.

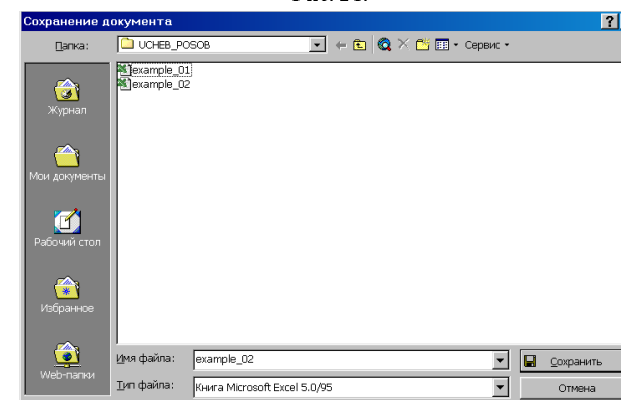


Рис. 22.

2. Создаем рабочий файл для импортирования исходных данных из Excel в Eviews, работая с диалоговым окном **File/New/Workfile** (рис. 23), далее выбираем: **Procs/Import/Read Text-Lotus-Excel** (рис. 24).

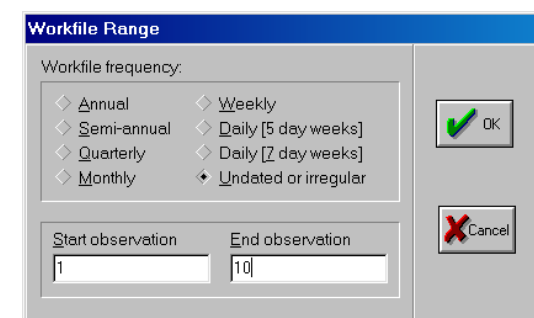


Рис. 23.

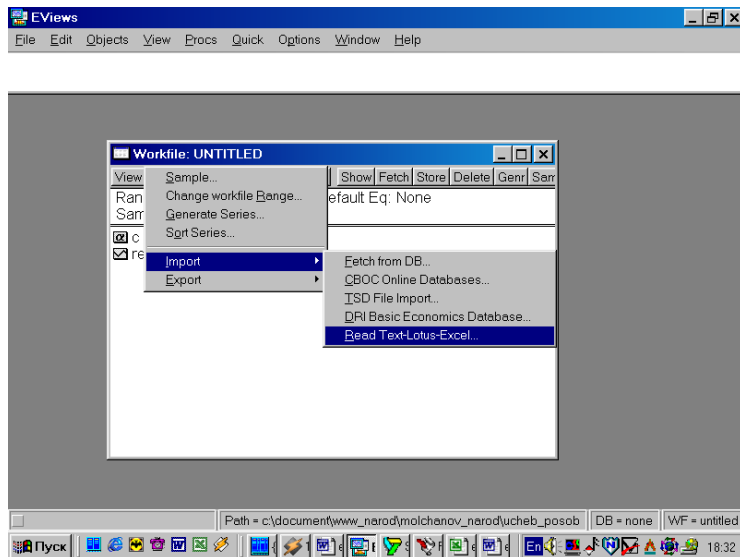


Рис. 24.

3. Далее в открывшемся окне находим и выбираем файл Excel с исходными данными (файл *не должен* в этот момент *использоваться* любыми программами), осуществляя автоматический импорт исходных данных в workfile (рис. 25). В следующем открывшемся диалоговом окне нужно указать адрес ячейки, в которой записаны данные первого по счету наблюдения и число переменных в рассматриваемом примере (рис. 26).. Если все выполнено правильно, то в открывшемся окне workfile должны появиться имена переменных, а также константа (c) и остатки (resid) (рис. 27).

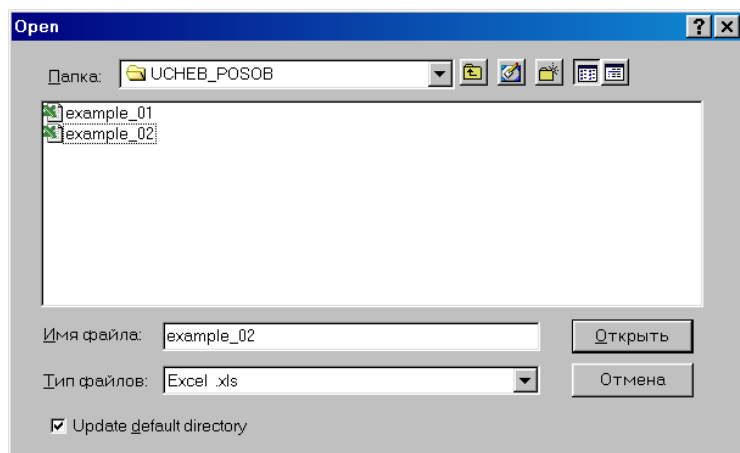


Рис. 25.

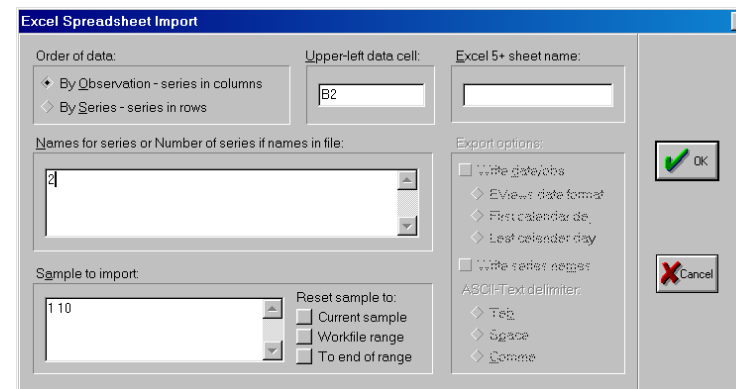


Рис. 26.

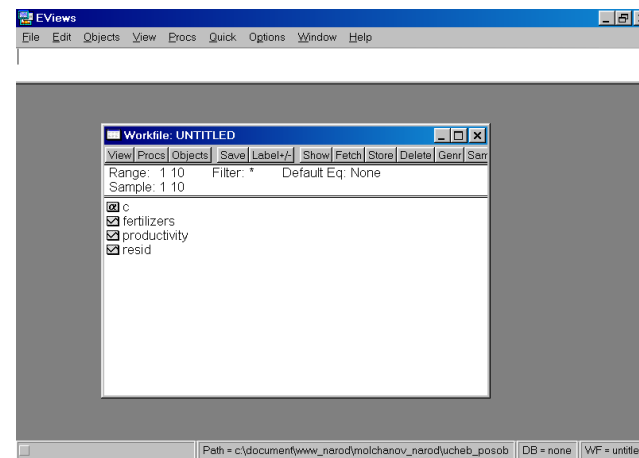


Рис. 27.

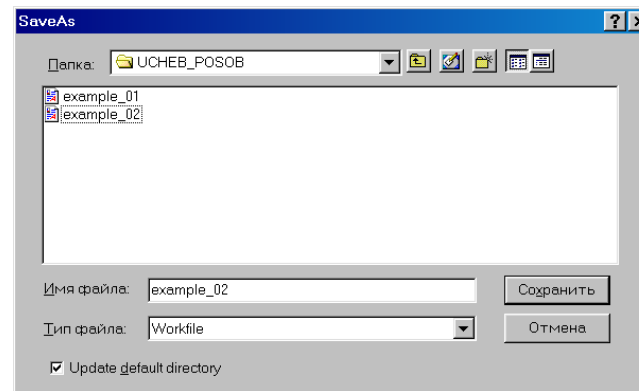


Рис. 28.

Сохраним рабочий файл (рис. 28).

- Значения описательных статистик находим следующим образом: в окне **workfile** выделяем переменные, щелкаем мышкой по выделенной части и далее выбираем: **Open/As Group/** (рис. 29). Открывается окно с исходными данными. Новую группу можно сохранить, выбрав опцию **Name** (рис. 30). Для просмотра описательных статистик **View/Descriptive Stats/Common Sample** (рис 31). Результат представлен на рис. 32.

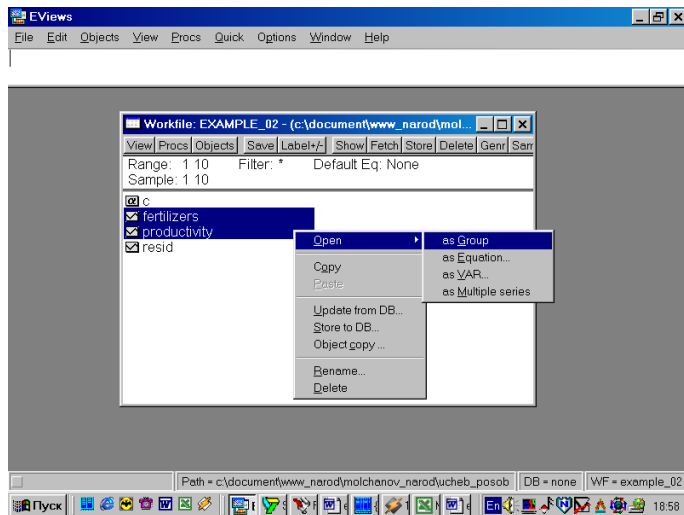


Рис. 29.

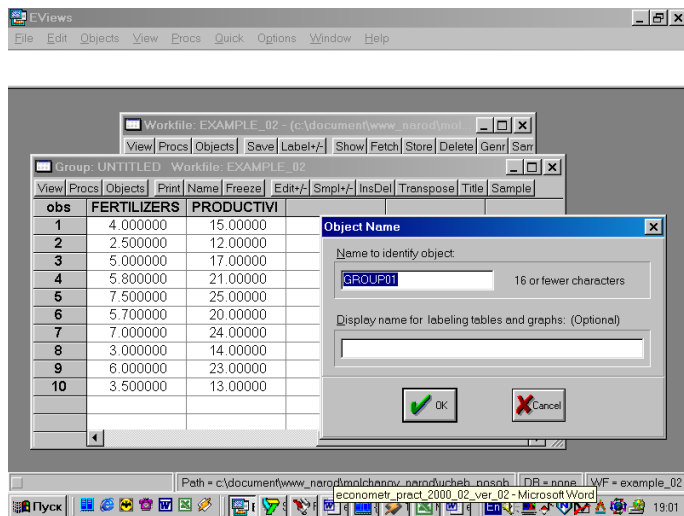


Рис. 30.

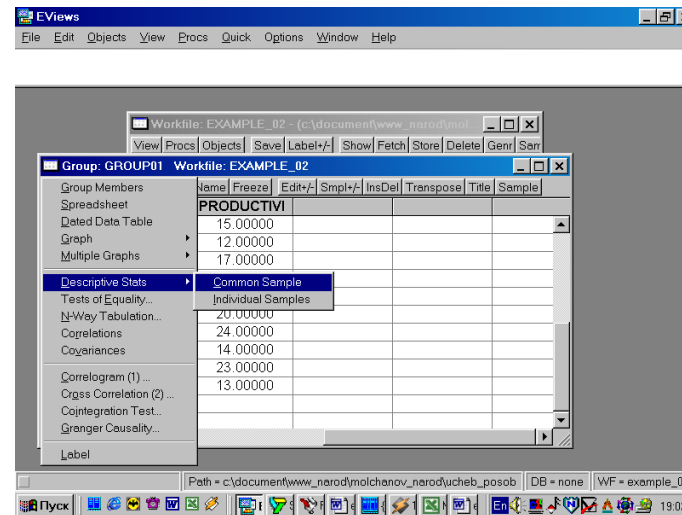


Рис. 31.

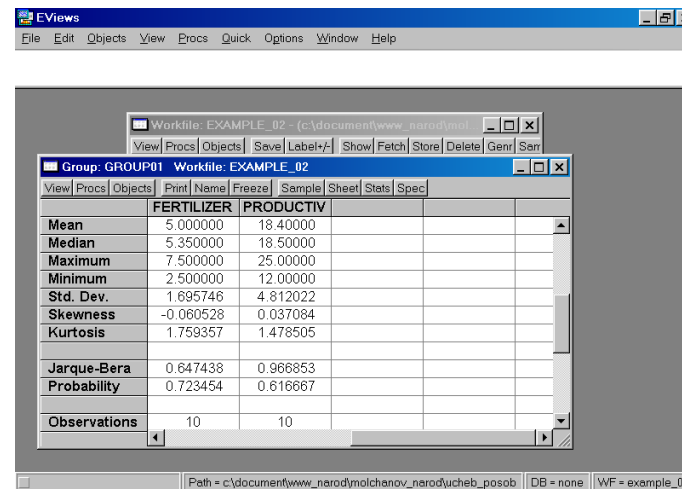


Рис. 32.

- В окне **workfile** (рис. 32) для построения поля корреляции необходимо выбрать следующие пункты меню: **VIEW/GRAPH/SCATTER/SIMPLE SCATTER/** (рис. 33). Полученный в результате график представляет собой поле корреляции резульативного и факторного признаков (рис. 34).
- В окне **Workfile** (используя созданную группу из двух переменных) выбрать: **/VIEW/CORRELATION/** (рис. 35). Полученная таблица - корреляционная матрица, в которой отражено значение коэффициента парной корреляции резульативного и факторного признаков (рис. 36).

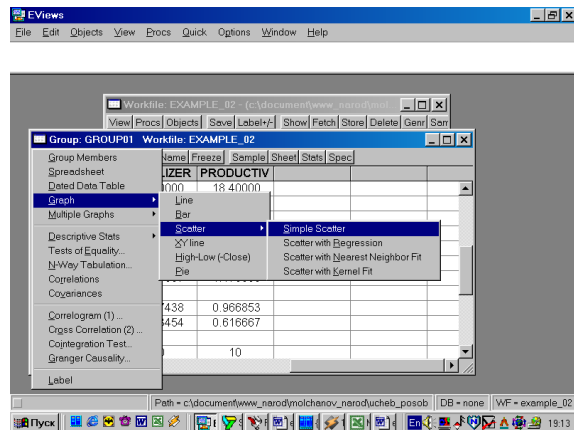


Рис. 33.

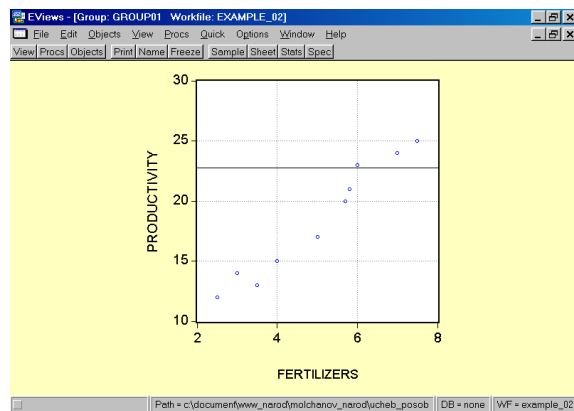


Рис. 34.

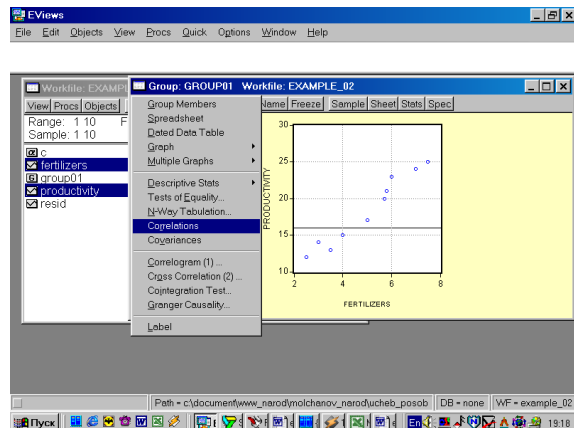


Рис. 35.

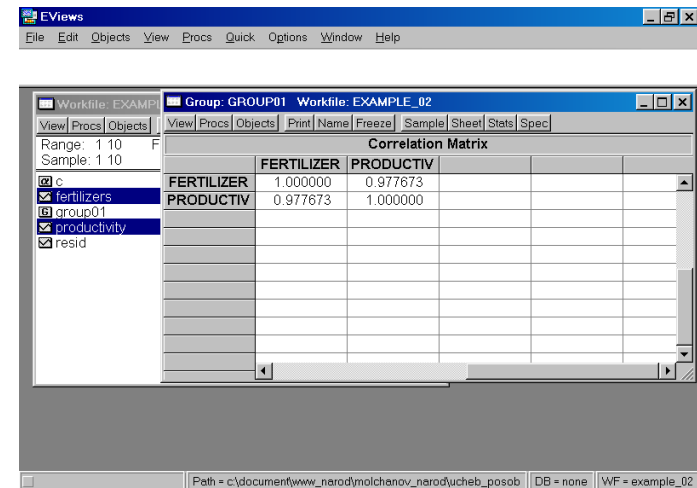


Рис. 36.

- В диалоговом окне описать в общем виде искомое уравнение: LS PRODUCTIVITY C FERTILIZERS <Enter> (метод наименьших квадратов (LS) эндогенная переменная, константа, экзогенная переменная), или выбрать в строке главного меню EIEWS: **QUICK/ESTIMATE EQUATION/PRODUCTIVITY C FERTILIZERS** (рис. 37). В открывшемся окне (рис. 38) должны быть переменные: зависимая переменная, применяемый метод, число наблюдений, параметры уравнения регрессии, стандартные ошибки, значения t – статистик и соответствующие им вероятности, значение R^2 и ряд других показателей.

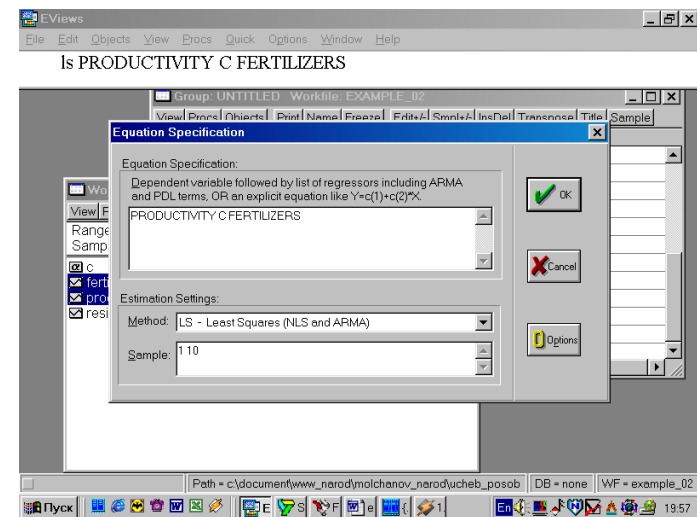


Рис. 37.

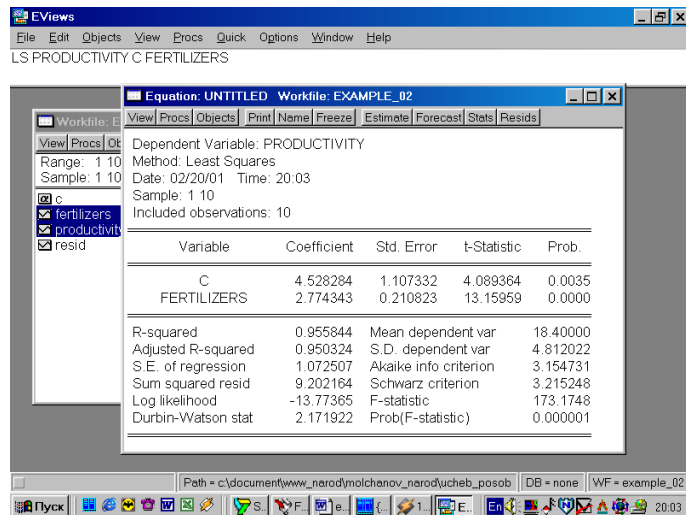


Рис. 38.

8. и 9. Результаты выполнения п.7 позволяют оценить статистическую значимость параметров уравнения регрессии и объяснить полученное значение R^2 .

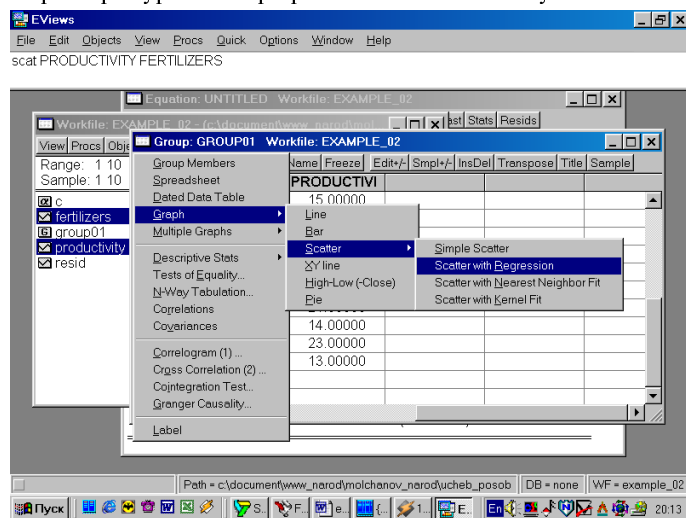


Рис. 39.

10. Для построения эмпирической линии регрессии в окне workfile выделить группу переменных и выбрать: **VIEW/GRAPH/SCATTER/SCATTER WITH REGRESSION/** (рис. 39). В промежуточном окне (рис. 40) необходимо нажать <Ok>. Полученный график (рис. 41) – эмпирическая линия регрессии. Чтобы построить теоретическую (подогнанную) линию регрессии, необходи-

мо найти теоретические (вычисленные с помощью уравнения регрессии) значения резульативного признака. Для этого открыть окно с параметрами уравнения регрессии, далее выбрать **Forecast** (рис. 42). Появится окно (рис. 43), в котором к исходным добавилась новая переменная **PRODUCTIVIf** (прогнозное, (теоретическое, выровненное) значение переменной **PRODUCTIVITY**). Затем, выделив все переменные (включая теоретическое значение резульативного признака), в командной строке записать **SCAT FERTILIZERS PRODUCTIVITY PRODUCTIVIf**. Полученный график (рис. 44) – теоретическая (подогнанная) линия регрессии.

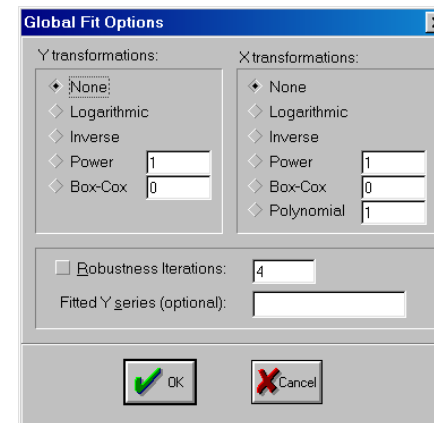


Рис. 40.

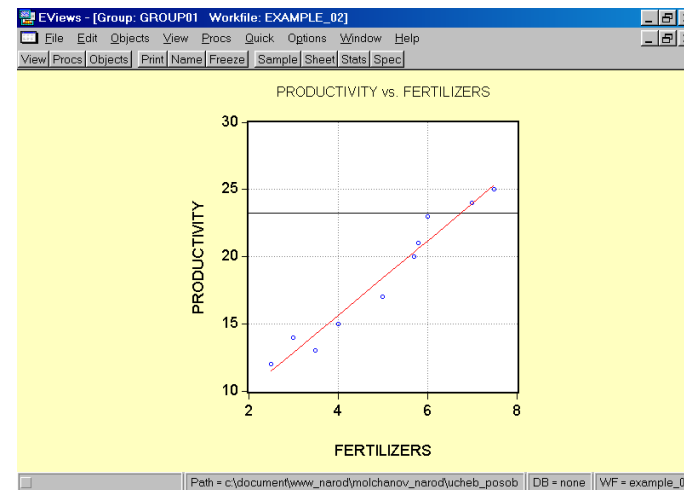


Рис. 41.

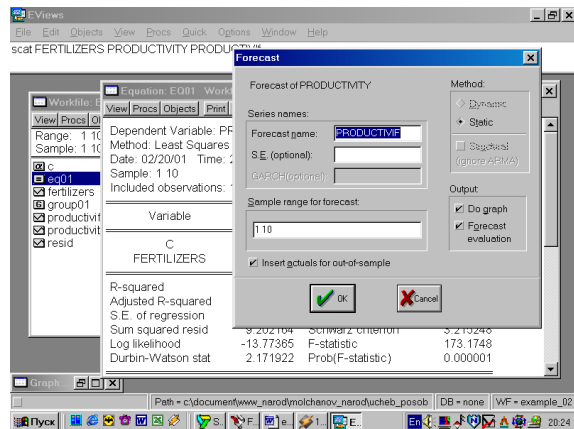


Рис. 42.

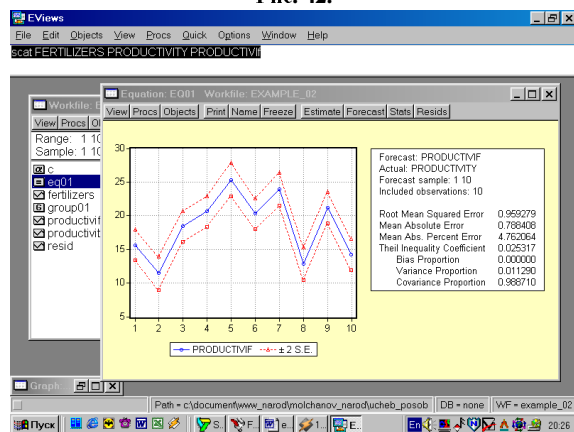


Рис. 43.

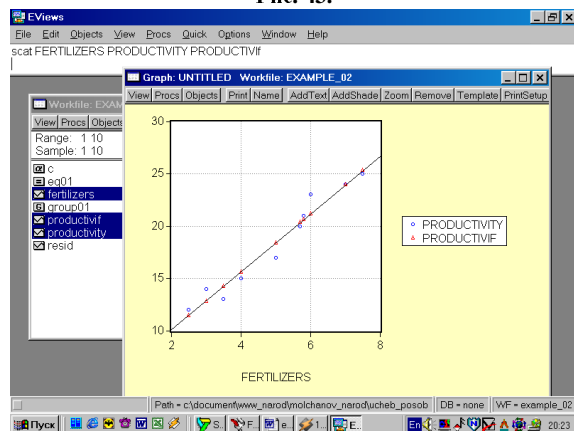


Рис. 44.

11. Данная операция возможна только в том случае, если ей предшествует построение регрессионного уравнения. В окне Workfile можно дважды щелкнуть на переменной **Resid** (рис. 45). Далее, выбрать: **VIEW/LINE GRAPH/** или, открыв окно с параметрами уравнения регрессии, выбрать: **VIEW/ACTUAL, FITTED.../ACTUAL, FITTED...TABLE/** (рис. 46). Результат представлен на рис. 47. Другой вариант вывода (фактические, предсказанные значения переменных, остатки, график остатков) – рис. 48.

Series: RESID	Value
1	-0.625657
2	0.535858
3	-1.400000
4	0.380526
5	-0.335858
6	-0.342040
7	0.051314
8	1.148686
9	1.825657
10	-1.238485

Рис. 45.

Label	Coefficient	Std. Error	t-Statistic	Prob.
C	4.528284	1.107332	4.089364	0.0035
FERTILIZERS	2.774343	0.210823	13.15959	0.0000

R-squared	0.955844	Mean dependent var	18.40000
Adjusted R-squared	0.950324	S.D. dependent var	4.812022
S.E. of regression	1.072507	Akaike info criterion	3.154731
Sum squared resid	9.202164	Schwarz criterion	3.215248
Log likelihood	-13.77365	F-statistic	173.1748
Durbin-Watson stat	2.171922	Prob(F-statistic)	0.000001

Рис. 46.

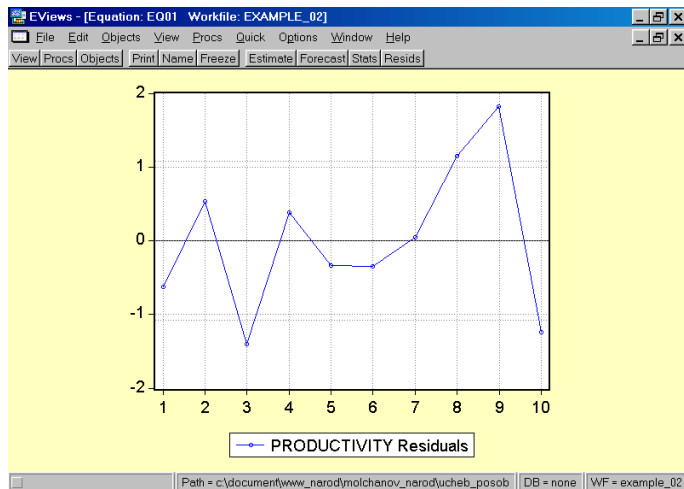


Рис. 47.

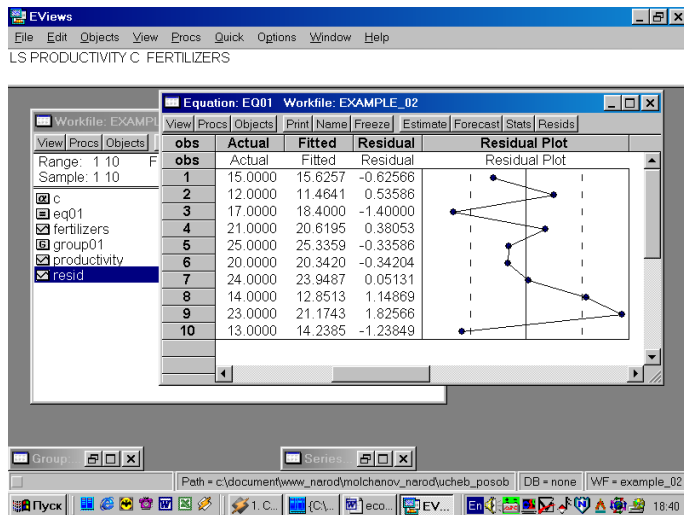


Рис. 48.

12. Для нахождения границ доверительного интервала в командной строке необходимо указать (рис. 49):

$$\text{GENR XK} = 5 * 1.05$$

$$\text{GENR YFK} = 4.53 + 2.77 * \text{XK}$$

$$\text{GENR h} = ((1 + 0.25^2) / 1.6957^2)^{0.5}$$

$$\text{GENR CI} = 2.31 * (1.07 / 10^{0.5}) * \text{h}$$

В результате искомые границы определяются следующим образом:

$\text{YFK} \pm \text{CI}$, т.е. от $\text{YFK} + \text{CI}$ до $\text{YFK} - \text{CI}$ (см. рис. 50).

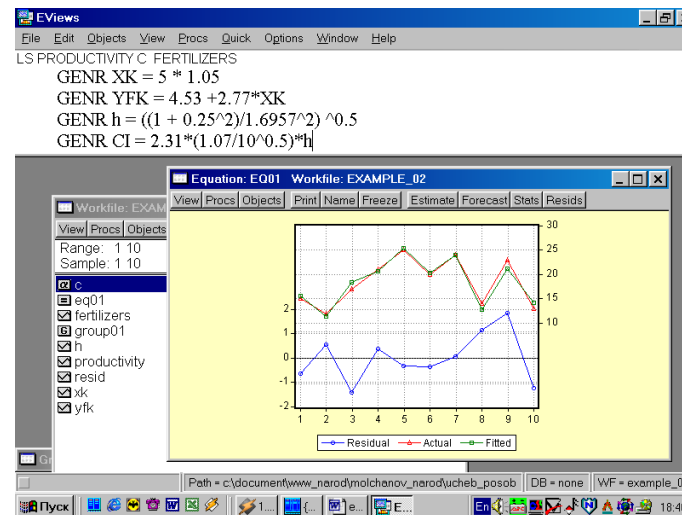


Рис. 49.

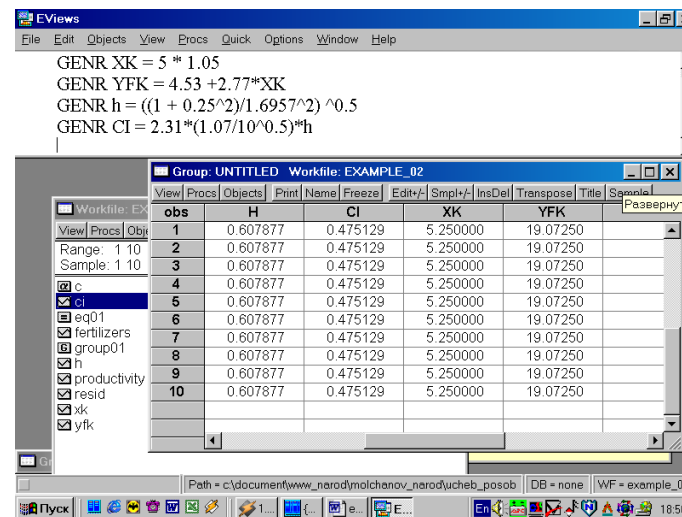


Рис. 50.

13. Оформить отчет по занятию.

Отчет должен содержать: подробные пояснения расчетов, ссылки на используемые формулы, результаты работы Eviews в виде экранных копий, другую, необходимую на Ваш взгляд, информацию.

Практическое занятие № 3.

«Применение Eviews при построении и анализе линейной однофакторной модели регрессии»

Выполняется самостоятельно.

Пример 3. Компания American Express Company в течение долгого времени полагала, что владельцы ее кредитных карточек имеют тенденцию путешествовать более интенсивно, как по делам бизнеса, так и для развлечений. Как часть объемного исследования, проведенного Нью-Йоркской компанией рыночных исследований по заказу American Express Company, было осуществлено определение взаимосвязи между путешествиями и расходами владельцев кредитных карточек. Исследовательская фирма случайным образом выбрала 25 владельцев карточек из компьютерного файла American Express Company и записала суммы их общих расходов за определенный период времени. Для выбранных владельцев карточек фирма так же подготовила и разослала по почте вопросы о числе миль, которые провел в путешествиях владелец карточки за изучаемый период. Данные, полученные из опроса, составляют исходную информацию анализа (X – число миль, проведенных в пути; Y – расходы путешественников (усл. ден ед.)¹).

№ п\п	Miles (X)	Costs (Y)
1	1211	1802
2	1345	2405
3	1422	2005
4	1687	2511
5	1849	2332
6	2026	2305
7	2133	3016
8	2253	3385
9	2400	3090
10	2468	3694
11	2699	3371
12	2806	3998
13	3082	3555
14	3209	4692
15	3466	4244
16	3643	5298
17	3852	4801
18	4033	5147
19	4267	5738
20	4498	6420
21	4533	6059
22	4804	6426
23	5090	6321
24	5233	7026
25	5439	6964

¹ Ниворожкина Л.И. Текст лекций по начальному курсу эконометрики для аспирантов.

1. Создать файл с исходными данными в среде Excel (файл **example_03.xls**).
2. Осуществить импорт исходных данных в Eviews.
3. Создать рабочий файл (workfile).
4. Найти значения описательных статистик по каждой переменной и объяснить их (рис. 51).
5. Построить поле корреляции моделируемого (результативного) и факторного признаков (рис. 52). Объяснить полученные результаты.
6. Найти значение линейного коэффициента корреляции и пояснить его смысл (рис. 53).
7. Определить параметры уравнения парной регрессии и интерпретировать их. Объяснить смысл полученного уравнения регрессии (рис. 54).
8. Оценить статистическую значимость коэффициента регрессии b и уравнения в целом. Сделать выводы.
9. Объяснить полученное значение R^2 .
10. Построить эмпирическую и теоретическую линию регрессии и объяснить их (рис. 55).
11. Построить и проанализировать график остатков (рис. 56).
12. С вероятностью 0,95 построить доверительный интервал для оценки ожидаемого значения средних расходов владельцев карточек, дальность путешествий которых составила 4000 миль (рис. 57).
13. Оформить отчет по занятию.

Результаты расчетов:

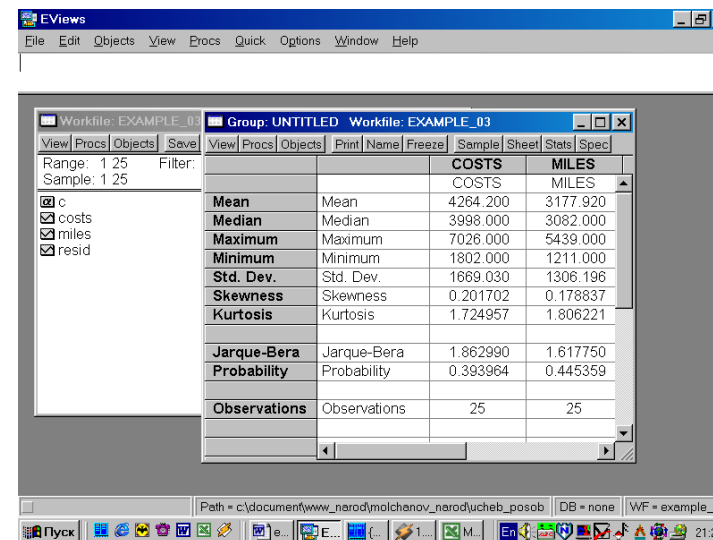


Рис. 51.

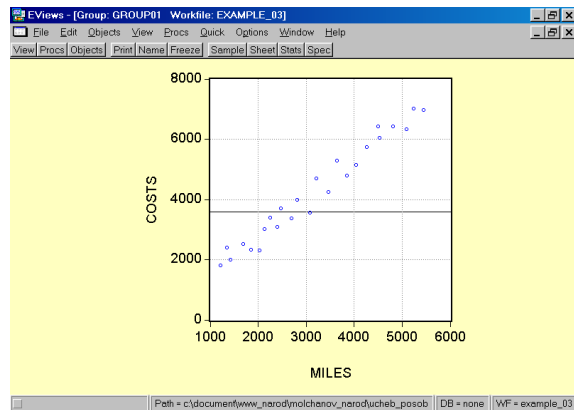


Рис. 52.

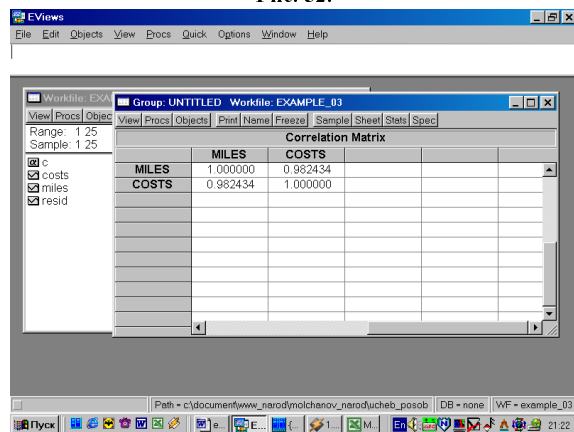


Рис. 53.

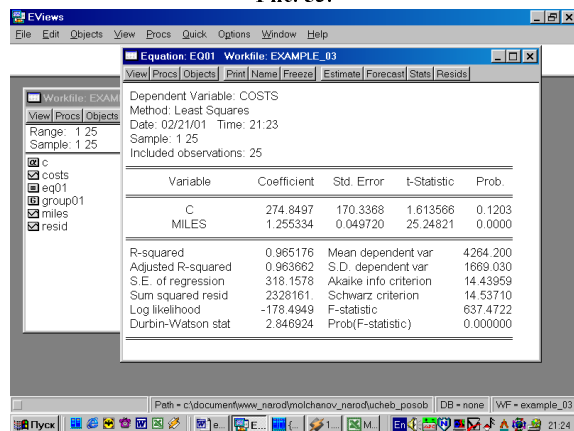


Рис. 54.

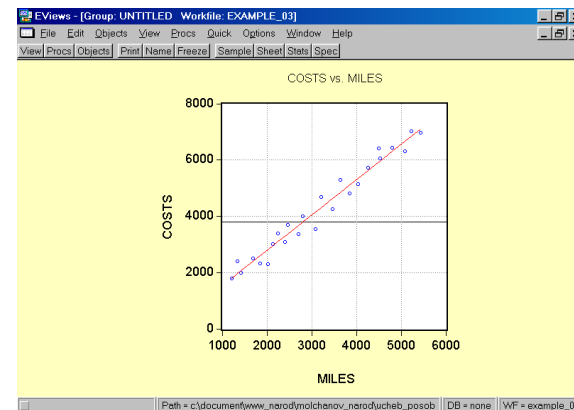


Рис. 55.

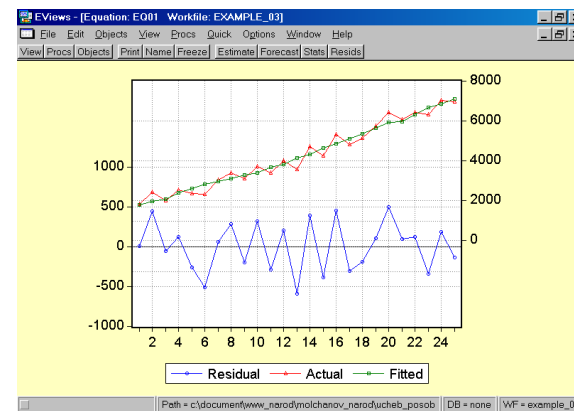


Рис. 56.

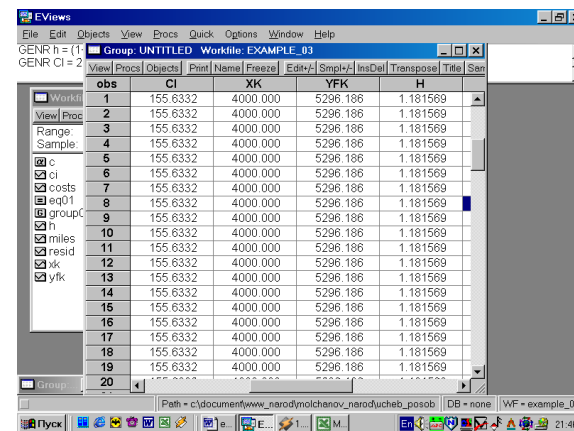


Рис. 57.

Практическое занятие № 4.

«Применение EViews при построении и анализе многофакторной модели регрессии. Выявление мультиколлинеарности и гетероскедастичности в модели. Проверка спецификации модели»

Пример 4. Имеются данные о вариации дохода кредитных организаций США за период 25 лет в зависимости от изменений годовой ставки по сберегательным депозитам и числа кредитных учреждений².

Введем следующие обозначения:

Y – прибыль кредитных организаций, %;

X_{1i} – чистый доход на 1\$ депозита;

X_{2i} – число кредитных учреждений.

Год	X_{1i} (Income)	X_{2i} (Credit institutions)	Y (Profit)
1	3,92	7298	0,75
2	3,61	6855	0,71
3	3,32	6636	0,66
4	3,07	6506	0,61
5	3,06	6450	0,7
6	3,11	6402	0,72
7	3,21	6368	0,77
8	3,26	6340	0,74
9	3,42	6349	0,9
10	3,42	6352	0,82
11	3,45	6361	0,75
12	3,58	6369	0,77
13	3,66	6546	0,78
14	3,78	6672	0,84
15	3,82	6890	0,79
16	3,97	7115	0,7
17	4,07	7327	0,68
18	4,25	7546	0,72
19	4,41	7931	0,55
20	4,49	8097	0,63
21	4,7	8468	0,56
22	4,58	8717	0,41
23	4,69	8991	0,51
24	4,71	9179	0,47
25	4,78	9318	0,32

1. Создать файл с исходными данными в среде Excel (файл **example_04.xls**).
2. Осуществить импорт исходных данных в Eviews.
3. Создать **workfile**.
4. Найти значения описательных статистик по каждой переменной и объяс-

² Ниворожина Л.И. Текст лекций по начальному курсу эконометрики для аспирантов.

нить их (рис. 58).

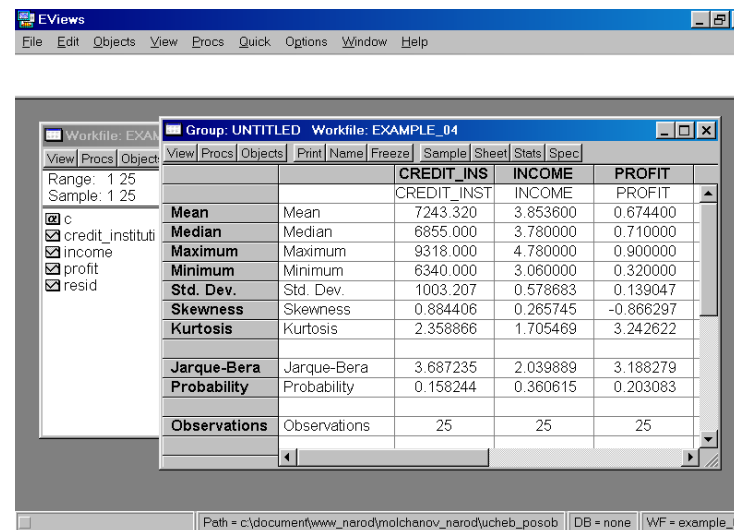


Рис. 58.

5. Построить корреляционную матрицу для всех переменных, включенных в модель (рис. 59).

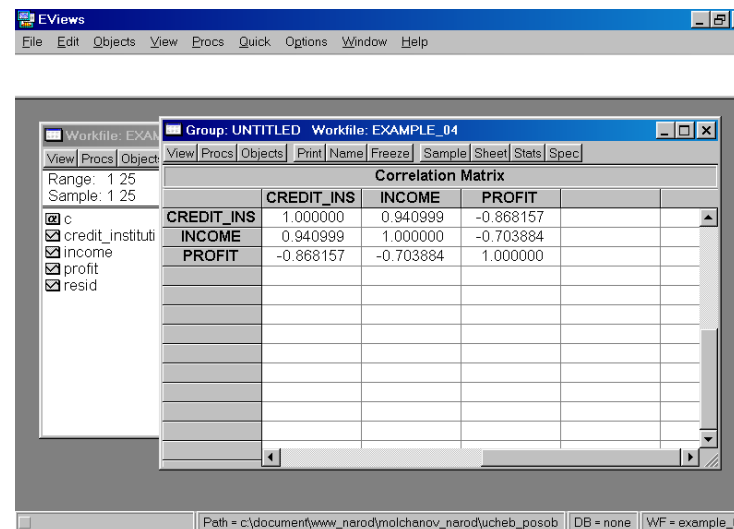


Рис. 59.

6. Построить регрессионное уравнение МНК, в котором зависимая переменная – прибыль кредитных организаций, а независимые – чистый доход на 1\$ депозита и число кредитных учреждений (рис. 60, 61).

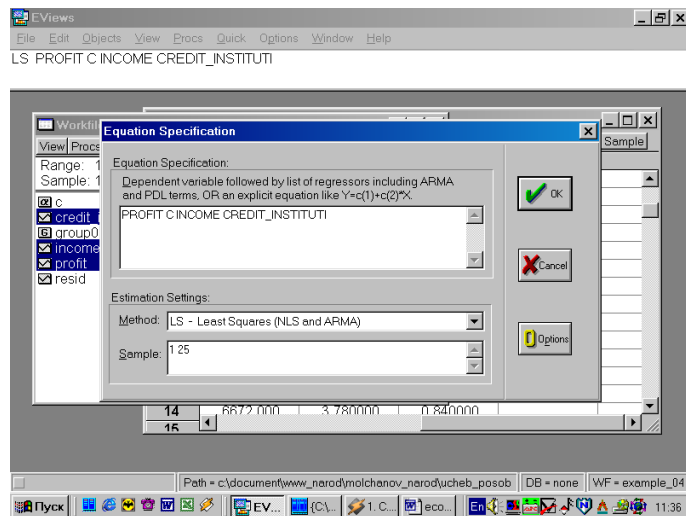


Рис. 60.

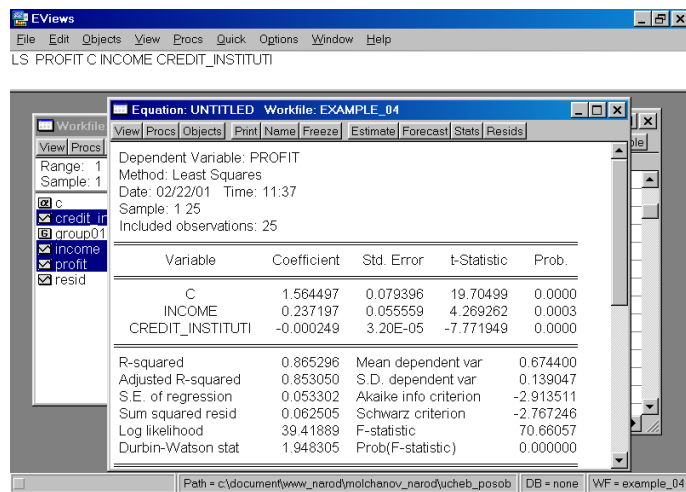


Рис. 61.

Уравнение примет следующий вид:

$$Y_i = \alpha_0 + \beta_1 INCOME + \beta_2 CREDIT_INSTITUTI + u_i.$$

Подставим полученные оценки из итоговой формы вывода:

$$\hat{Y}_i = 1,5645 + 0,2372INCOME - 0,00025CREDIT_INSTITUTI.$$

- Оценить статистическую значимость параметров полученного уравнения и всей модели в целом.
- Проверить наличие мультиколлинеарности в модели. Сделать вывод.

Мультиколлинеарность – это коррелированность двух или нескольких объясняющих переменных в уравнении регрессии. В результате высококоррелированные объясняющие переменные действуют в одном направлении и имеют недостаточно независимое колебание, чтобы дать возможность модели изолировать влияние каждой переменной. Проблема мультиколлинеарности возникает только в случае множественной регрессии. Мультиколлинеарность особенно часто имеет место при анализе макроэкономических данных (например, доходы, производство). Получаемые оценки оказываются нестабильными как в отношении статистической значимости, так и по величине и знаку (например, коэффициенты корреляции). Следовательно, они ненадежны. Значения коэффициентов R^2 могут быть высокими, но стандартные ошибки тоже высоки, и отсюда t -критерии малы, отражая недостаток значимости.

Для проверки появления мультиколлинеарности применяются два метода, доступные во всех статистических пакетах³:

- Вычисление матрицы коэффициентов корреляции для всех объясняющих переменных. Если коэффициенты корреляции между отдельными объясняющими переменными очень велики, то, следовательно, они коллинеарны. Однако, при этом не существует единого правила, в соответствии с которым есть некоторое пороговое значение коэффициента корреляции, после которого высокая корреляция может вызвать отрицательный эффект и повлиять на качество регрессии.
- Для измерения эффекта мультиколлинеарности используется показатель VIF – «фактор инфляции вариации»:

$$VIF(X_h) = \frac{1}{1 - R_h^2}, \text{ где } R_h^2 - \text{значение коэффициента множественной корреляции, полученное для регрессора } X_h \text{ как зависимой переменной и остальных переменных } X_i.$$

При этом степень мультиколлинеарности, представляемая в регрессии переменной X_h , когда переменные X_1, X_2, \dots, X_k включены в регрессию, есть функция множественной корреляции между X_h и другими переменными X_1, X_2, \dots, X_k .

- Если $VIF > 10$, то объясняющие переменные, коррелирующие между собой, считаются мультиколлинеарными.

Существует еще ряд способов, позволяющих обнаружить эффект мультиколлинеарности:

- Стандартная ошибка регрессионных коэффициентов близка к нулю.
- Мощность коэффициента регрессии отличается от ожидаемого значения.
- Знаки коэффициентов регрессии противоположны ожидаемым.
- Добавление или удаление наблюдений из модели сильно изменяют значения оценок.

³ Ниворожжина Л.И. Текст лекций по начальному курсу эконометрики для аспирантов.

- Значение F-критерия существенно, а t-критерия – нет.
- Для устранения мультиколлинеарности может быть принято несколько мер:
- Увеличивают объем выборки по принципу, что больше данных означает меньшие дисперсии оценок МНК. Проблема реализации этого варианта решения состоит в трудности нахождения дополнительных данных.
- Исключают те переменные, которые высококоррелированы с остальными. Проблема здесь заключается в том, что возможно переменные были включены на теоретической основе, и будет неправомерным их исключение только лишь для того, чтобы сделать статистические результаты «лучше».
- Объединяют данные кросс-секций и временных рядов. При этом методе берут коэффициент из, скажем, кросс-секционной регрессии и заменяют его на коэффициент из эквивалентных данных временного ряда.

Проделанные манипуляции позволяют предположить, что мультиколлинеарность может присутствовать (оценки любой регрессии будут страдать от нее в определенной степени, если только все независимые переменные не окажутся абсолютно некоррелированными), однако в данном примере это не влияет на результаты оценки регрессии. Следовательно, выделять «лишние» переменные не стоит, так как это отражается на содержательном смысле модели.

9. Проверить спецификацию модели. Объяснить полученные результаты.

Подробно теоретические вопросы, связанные с проблемами спецификации эконометрических моделей, были рассмотрены в лекционном курсе.

В нашем случае мы ограничимся тем, что попробуем исключить поочередно независимые переменные. Первой исключаем переменную **CREDIT_INSTITUTI** (рис. 62). Коэффициент при переменной **INCOME** изменил знак на противоположный.

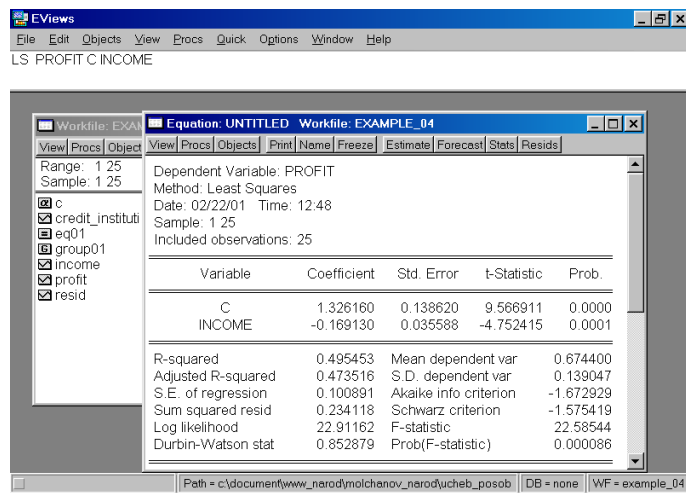


Рис. 62.

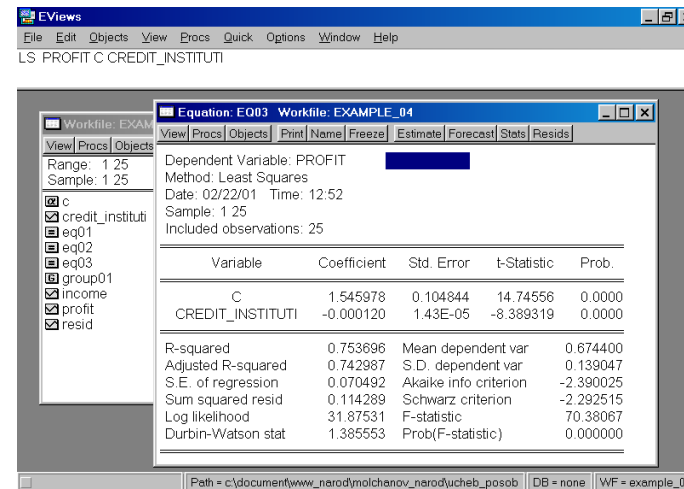


Рис. 63.

В случае исключения из первоначальной модели переменной **INCOME**, знак регрессионного коэффициента при переменной **CREDIT_INSTITUTI** остался без изменения (рис. 63). Представляется разумным разделять эффект двух независимых переменных на зависимую переменную в модели с совместным их влиянием в регрессионном уравнении. Данный пример иллюстрирует важность использования множественной регрессии вместо парной в случае, когда изучаемое явление существенно детерминирует несколько независимых переменных.

10. Проверить наличие гетероскедастичности в модели. Объяснить полученные результаты.

Если остатки имеют постоянную дисперсию, они называются **гомоскедастичными**, но если они непостоянны, то **гетероскедастичными**. Гетероскедастичность приводит к тому, что коэффициенты регрессии больше не представляют собой лучшие оценки или не являются оценками с минимальной дисперсией, следовательно, они больше не являются наиболее эффективными коэффициентами.

Воздействие гетероскедастичности на оценку интервала прогнозирования и проверку гипотезы заключается в том, что хотя коэффициенты не смещены, дисперсии и, следовательно, стандартные ошибки этих коэффициентов будут смещены. Если смещение отрицательно, то оценочные стандартные ошибки будут меньше, чем они должны быть, а критерий проверки будет больше, чем в реальности. Таким образом, мы можем сделать вывод, что коэффициент значим, когда он таковым не является. И наоборот, если смещение положительно, то оценочные ошибки будут больше, чем они должны быть, а критерии проверки – меньше. Значит, мы можем принять нулевую гипотезу, в то время как она должна быть отвергнута.

Проверкой на гетероскедастичность служит тест Голдфелда-Кванта. Он требует, чтобы остатки были разделены на две группы из n наблюдений, одна группа с низ-

кими, а другая – с высокими значениями. Обычно срединная одна шестая часть наблюдений удаляется после ранжирования в возрастающем порядке, чтобы улучшить разграничение между двумя группами. Отсюда число остатков в каждой группе составляет $(n - c)/2$, где c представляет одну шестую часть наблюдений.

Критерий Голдфелда-Кванта – это отношение суммы квадратов отклонений (СКО) высоких остатков к СКО низких остатков:

$$(n - c)/2.$$

Этот критерий имеет t – распределение с $(n - c)/(2 - k)$ степенями свободы.

Чтобы решить проблему гетероскедастичности, нужно исследовать взаимосвязь между значениями ошибки и переменными и трансформировать регрессионную модель так, чтобы она отражала эту взаимосвязь. Это может быть достигнуто посредством регрессии значений ошибок по различным формам функций переменной, которая приводит к гетероскедастичности, например,

$$e_i = \alpha + \beta \cdot X_i^H,$$

где X_i - независимая переменная (или какая-либо функция независимой переменной), которая предположительно является причиной гетероскедастичности, а H отражает степень взаимосвязи между ошибками и данной переменной, например, X^2 или $X^{1/n}$ и т. д.

Следовательно, дисперсия коэффициентов запишется:

$$E(\sigma_i^2) = \sigma^2 X_i^H.$$

Отсюда если $H = 1$, мы трансформируем регрессионную модель к виду:

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\alpha}{\sqrt{X_i}} + \beta_i \frac{e_i}{\sqrt{X_i}}.$$

Если $H = 2$, т.е. дисперсия увеличивается в пропорции к квадрату рассматриваемой переменной X , трансформация приобретает вид:

$$\frac{Y_i}{X_i} = \frac{\alpha}{X_i} + \beta_i \frac{e_i}{X_i}.$$

Используя Eviews, можно провести проверку и устранение гетероскедастичности следующим образом:

- Запустить стандартную регрессию.
- Вычислить остатки.
- Запустить регрессию с использованием квадрата остатков как зависимой переменной и оценить зависимую переменную \hat{u} как независимую переменную (тест White).

- Оценить nR^2 , где n – объем выборки, R^2 – коэффициент детерминации.
- Использовать статистику χ^2 с одной степенью свободы (в EViews – используется F – статистика) для проверки существенности отличия nR^2 от нуля.
- Основным способом устранения гетероскедастичности является применение взвешенного метода наименьших квадратов.

Выбираем тест White (см. рис. 64).

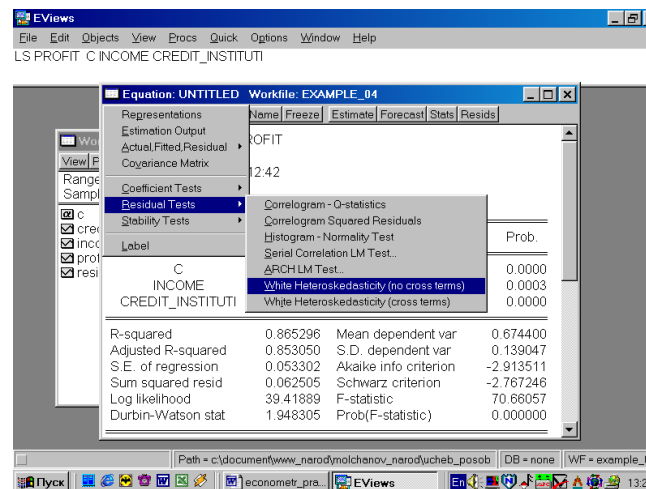


Рис. 64.

Итог формы вывода представлен на рис. 65.

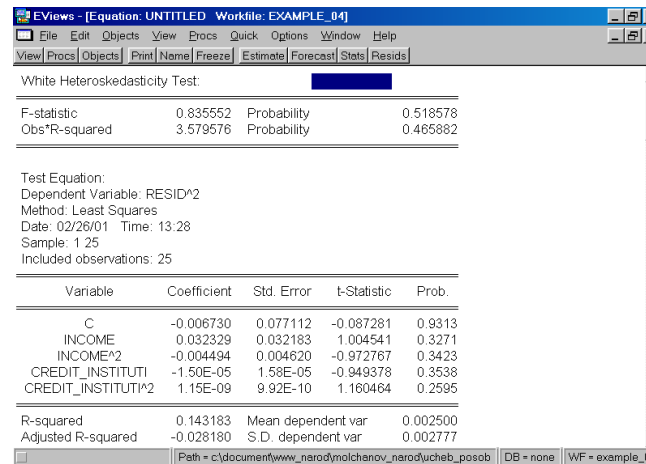


Рис. 65.

Как следует из приведенной распечатки, вероятность ошибки первого рода равна 51,86%. Следовательно, нулевую гипотезу (об отсутствии гетероскедастичности)

нельзя отклонить.

Для случая, когда гетероскедастичность присутствует, проблему гетероскедастичности можно решать следующим образом:

Выбираем в пунктах меню текущего окна опцию **Proc/Specify/Estimate...** (рис. 66). Появляется окно оценки регрессии, где необходимо нажать клавишу **Options** и в появившемся окне отметить **Heteroskedasticity** (рис. 67).

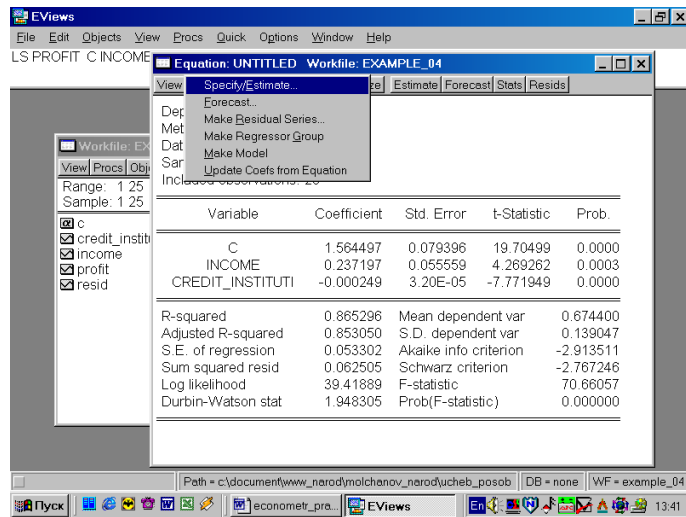


Рис. 66.

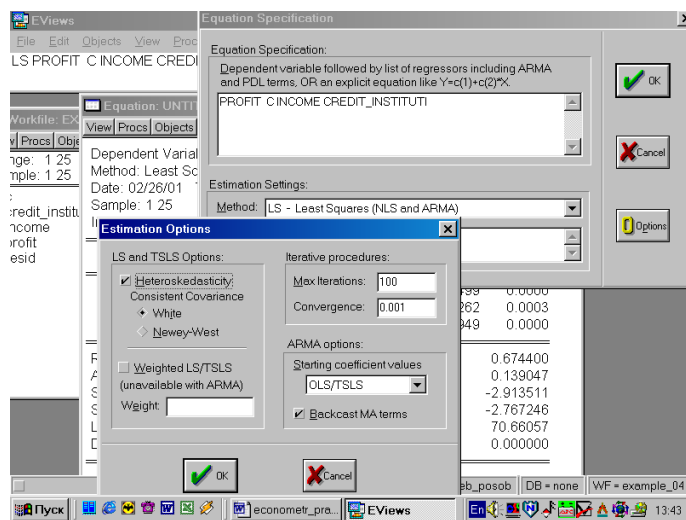


Рис. 67.

Появилось новое, переоцененное уравнение (рис. 68). Полученное уравнение можно вновь проверить по тесту White.

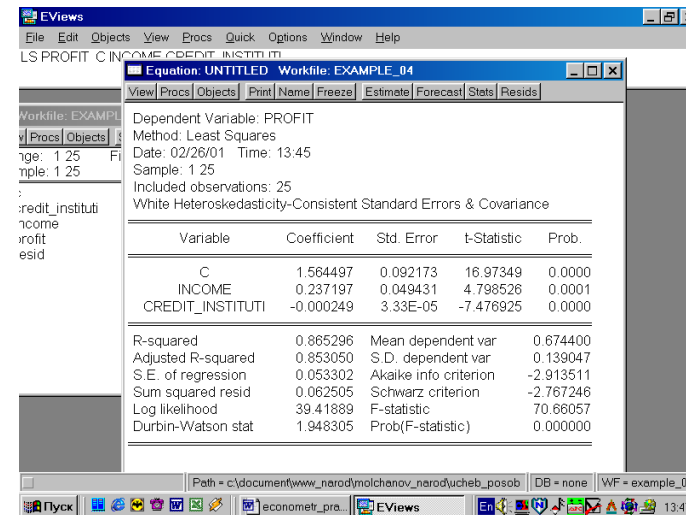


Рис. 68.

11. Оформить отчет.

Практическое занятие № 5.

«Фиктивные переменные»

Иногда необходимо включение в регрессионную модель одной или более качественных переменных (например, разделение по полу: мужской и женский; по уровню образования: общее и профессиональное и т.д.). Альтернативно может понадобиться сделать качественное различие между наблюдениями одних и тех же данных. Так, если проверяется взаимосвязь между размером компании и месячными доходами по акциям, может быть желательным включение качественной переменной, представляющей месяц январь, по причине хорошо известного «январского эффекта» во временных рядах доходов по ценным бумагам. Данный «январский эффект» - это феномен, заключающийся в том, что средние доходы по акциям, особенно небольших компаний, в среднем выше в январе, чем в другие месяцы. Таким образом, если мы рассматриваем январские наблюдения как качественно отличные от других наблюдений, фиктивная переменная (D) позволит произвести подобное качественное различие.

Фиктивные переменные бывают двух типов - сдвига и наклона. Фиктивная переменная сдвига - это переменная, которая меняет точку пересечения линии регрессии с осью ординат в случае применения качественной переменной (рис. 69). Фиктивная переменная наклона - это та переменная, которая изменяет наклон линии регрессии в случае использования качественной переменной (рис. 70). Оба типа фиктивных переменных будут иметь значение $+1$ или -1 , когда наблюдения данных совпадают с уместной количественной переменной, но будут иметь нулевое значение при совпадении с наблюдениями, где эта качественная переменная отсутствует.

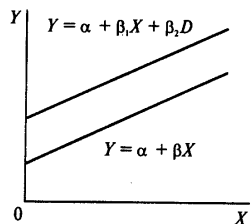


Рис. 69.

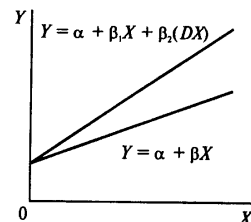


Рис. 70.

Пример 5. По данным примера 1 (файл `example_01.xls`.) дать интерпретацию бинарным, «фиктивным» переменным, принимающим значения 0 или 1: `floor` – принимает значение 0, если квартира расположена на первом или последнем этаже, `cat` – принимает значение 1, если квартира находится в кирпичном доме.

Построим регрессионное уравнение вида `LS PRICE C CAT FLOOR` (рис 71). Тем самым мы предполагаем (хотя в действительности это может быть и не так), что на цену квартиры оказывают влияние только две, указанные выше, составляющие. В результате получится уравнение следующего вида (рис 72):

$$PRICE = \alpha + \beta_1 CAT + \beta_2 FLOOR + \varepsilon.$$

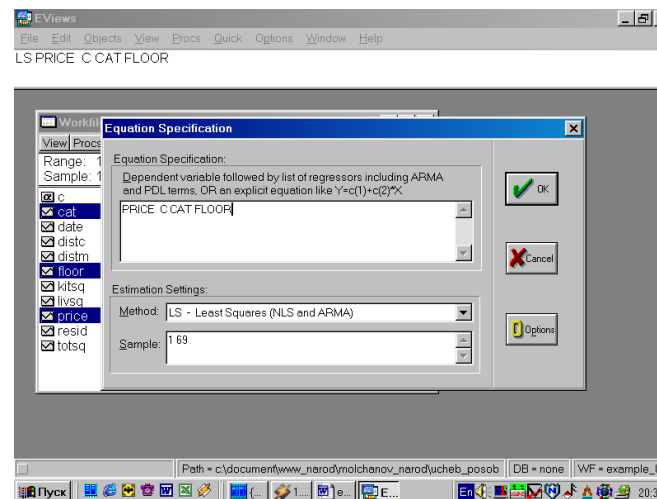


Рис. 71.

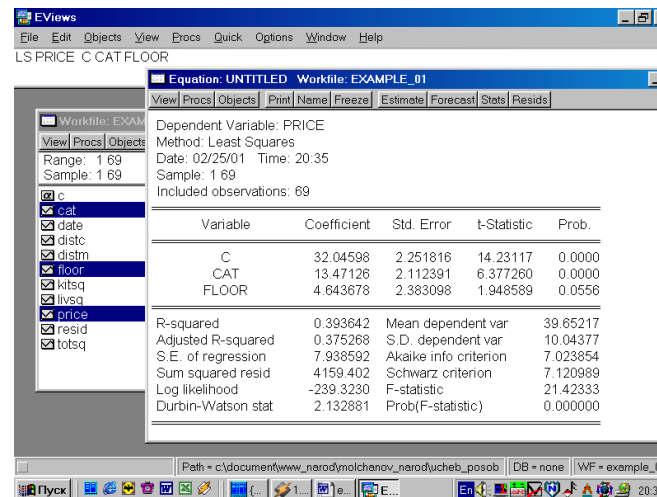


Рис. 72.

Используя результаты оценивания уравнения, содержащиеся в форме вывода (рис. 72), можно записать такое уравнение:

$$PRICE = 32,04598 + 13,47126 \cdot CAT + 4,643678 \cdot FLOOR.$$

Как же можно интерпретировать полученные результаты? Полученный коэффициент при `CAT` означает, что квартиры в кирпичных домах стоят в среднем на \$13471 дороже аналогичных квартир в панельных домах. Коэффициент при `FLOOR` может быть интерпретирован так: квартиры на не первом/последнем этажах стоят в среднем на \$4644 дороже аналогичных, расположенных на первом/последнем этажах.

Практическое занятие № 6.

«Однофакторные стохастические модели динамических процессов»

Процедуры анализа временных рядов и построения простейших прогнозов включают в себя очень широкий спектр проблем, возникающих в процессе эконометрического исследования. В рамках однофакторных стохастических моделей предполагается, что рассматриваемый процесс состоит из компонентов анализируемых временных рядов (если будущие значения рассматриваемой переменной сколько-нибудь предсказуемы, то они являются функцией от прошлых значений этой переменной). Проведем анализ составляющих однофакторного стохастического процесса в контексте авторегрессионного анализа, процессов со скользящей средней и степени интегрирования. Эти три подпроцесса в зарубежной научной литературе объединены под названием авторегрессионные интегрированные модели со скользящей средней (*ARIMA*)¹.

Перед применением *ARIMA* необходимо определить разности уровней с целью получения стационарного ряда, то есть нужно знать порядок этих разностей. Таким образом, процесс *ARIMA* обладает тремя параметрами: P - порядок авторегрессии, d - требуемый порядок предварительно определяемых разностей и q - порядок скользящей средней в модели.

Понятие интегрирования означает, какого порядка разности должны быть рассчитаны для того, чтобы получить стационарный временной ряд. Нахождение разностей - это всего лишь нахождение изменений значения переменной в последующие периоды, т.е. нахождение величины $\Delta Y_t = Y_t - Y_{t-1}$. Ряд значений ΔY_t - это ряд разностей.

Если во временном ряду должны быть рассчитаны первые разности, чтобы получить стационарный ряд, то первоначальный ряд называется интегрированным рядом первого порядка, или $I(1)$. Если же требуется рассчитать вторые разности для получения стационарного ряда, то это интегрированный ряд второго порядка, или $I(2)$. Если же в ряду вообще не требуется вычислять разности, то он называется интегрированным рядом нулевого порядка, или $I(0)$.

Если ряд $I(0)$, т.е. стационарен, то его дисперсия будет конечна. Изменения рассматриваемой переменной будут иметь только промежуточное влияние на временной ряд. Коэффициенты автокорреляции будут постепенно убывать таким образом, что их сумма станет конечной. И наоборот, если ряд $I(1)$, то изменения будут

¹ Уотшем Т.Дж., Паррамоу К. Количественные методы в финансах: Учебное пособие для вузов /Пер. с англ. под ред. М.Р.Ефимовой. – М.: Финансы, ЮНИТИ, 1999. – 527 с.

иметь постоянный эффект, дисперсия с течением времени будет возрастать до бесконечности.

Проведя расчет разностей, можно переходить к моделированию полученного стационарного ряда (ряда, обладающего постоянной средней и дисперсией, ковариация в котором зависит только от временного интервала между двумя отдельными наблюдениями) с помощью *ARMA*.

Начнем анализ *ARIMA* с рассмотрения авторегрессионного процесса. Авторегрессионным называется процесс, при котором значение ряда находится в линейной зависимости от предыдущих значений. Например, если текущее наблюдаемое значение является функцией всего лишь одного значения, непосредственно предшествующего наблюдению, т.е. процесс зависит всего лишь от одного значения рассматриваемой переменной, то процесс называется авторегрессионным процессом первого порядка и обозначается $AR(1)$. Это можно обобщить следующим образом: если анализируемый динамический процесс зависит от значений, отстоящих от 1 до n временных лагов назад, то это авторегрессионный процесс порядка n , т.е. $AR(n)$. Например, процесс $AR(3)$ можно отобразить следующим образом:

$$\hat{Y}_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \varepsilon_t. \quad (6.1)$$

Из формулы (6.1) следует, что текущее значение Y - это функция от трех наиболее недавних предыдущих значений; ε_t - остаток или ошибка (погрешность).

Модель скользящей средней (*MA*)² - это модель, где моделируемая величина задается линейной функцией от прошлых ошибок, то есть разностей между прошлыми смоделированными значениями и прошлыми фактическими наблюдениями:

$$\hat{Y}_t = B_0 + B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + B_3 \varepsilon_{t-3} + \varepsilon_t, \quad (6.2)$$

$$\text{где } \varepsilon_t = Y_t - \hat{Y}_t.$$

Используемый термин «скользящая средняя» не следует путать со схожим термином, относящимся к технике сглаживания данных.

Таким образом, модель *ARMA(pq)* имеет p временных лагов в авторегрессионном процессе и q интервалов в модели скользящей средней. Например, *ARMA(3,2)* будет иметь следующий вид:

$$\hat{Y}_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + B_3 \varepsilon_{t-3} + u_t, \quad (6.3)$$

где u_t - остаточный член ошибки в данном уравнении.

² Используемый термин «скользящая средняя» не следует путать со схожим термином, относящимся к технике сглаживания данных.

Для определения степени автокорреляции временных рядов необходимо определить силу связи между текущими и прошлыми значениями рассматриваемой переменной. Одним из способов измерения этой связи являются коэффициенты автокорреляции (*ACC*), совокупность которых образует функции автокорреляции (*ACF*). Коэффициент автокорреляции измеряет связь между текущими и прошлыми наблюдениями временного ряда и рассчитывается следующим образом:

$$\rho_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y}) \cdot (Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (6.4)$$

где k - количество лагов.

Таким образом, коэффициент автокорреляции первого порядка будет рассчитан с лагом в один период, коэффициент автокорреляции второго порядка будет учитывать степень связи между значениями, отстоящими на два временных периода, и т.д. Рассчитываются коэффициенты автокорреляции всех порядков и затем проводится статистическая проверка для определения, при каких лагах коэффициенты статистически значимы. Только лаги, являющиеся статистически значимыми, оставляются в модели. Иногда проверка значимости коэффициентов автокорреляции проводится при помощи критерия стандартной ошибки и Q -критерия Бокса-Пирса³. Два критерия предлагаются потому, что существуют два подхода к проверке наличия автокорреляции. При первом подходе, то есть при использовании критерия стандартной ошибки, проверяются коэффициенты автокорреляции каждого порядка отдельно, чтобы выявить, какие из них значимы. Второй подход использует Q -критерий Бокса-Пирса для того, чтобы проверить на значимость все множество коэффициентов как группу. В настоящее время, с развитием компьютерных программ и средств, проверку удобнее проводить по методу P -value (методом определения уровня вероятности, то есть получением P -значения, которое в случае верности нулевой гипотезы представляет собой вероятность получения величины стандартизованного критерия проверки, большего по абсолютному значению, чем рассчитанный критерий проверки).

Частный коэффициент автокорреляции (*PAC*), лежащий в основе частной функции автокорреляции (*PAF*), измеряет связь между текущим значением пере-

³ Box, G.E.P. and Pierce, D.A. «Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models». – Journal of the American Statistical Association, 1970, № 65. – P. 1509-1526.

менной X_t и последующими значениями этой переменной $X_{t-1}, X_{t-2}, \dots, X_{t-k}$, когда влияние всех промежуточных временных лагов устранено. Таким образом, частный коэффициент автокорреляции первого порядка будет равен коэффициенту автокорреляции первого порядка, так как нет промежуточных лагов. Но частные коэффициенты второго и следующих порядков будут уже отличаться друг от друга.

Частный коэффициент автокорреляции используется для определения степени автокорреляции внутри временного ряда. Например, ряд, обозначенный $AR(m)$, показывает, что последний статистически значимый частный коэффициент автокорреляции рассчитан с лагом m . Таким образом, в ряде $AR(2)$ текущее значение переменной обладает значимой корреляцией только со значениями, отстоящими на 1 и 2 временных лага назад. В ряде $AR(4)$ значимыми будут частные коэффициенты автокорреляции с лагами от одного до четырех периодов, но коэффициенты с более высокими лагами не будут значимо отличаться от нуля.

В динамическом процессе $AR(m)$ частные коэффициенты автокорреляции значимо отличаются от нуля для временных лагов от 1 до m и затем резко падают до нуля для интервалов $m+1$ и больше.

Зная поведение коэффициента автокорреляции и частного коэффициента автокорреляции, можно попытаться определить, содержит ли ряд элемент скользящей средней. Если ряд скорее AR чем MA , то автокорреляция не будет показывать порядок MA -процесса. Хотя, если значение частных коэффициентов автокорреляции падает по экспоненте, а не опускается резко до нуля, то можно предположить, что ряд содержит процесс скользящей средней, а не AR .

Для проверки автокорреляции в рядах, где присутствуют элементы и авторегрессии и скользящей средней, используется критерий Лjung-Бокса (LB) (Ljung-Box)⁴. Критерий LB рассчитывается следующим образом:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{1}{n-k} \right) \cdot r_k^2 \approx \chi_{m-p-q}^2, \quad (6.5)$$

где m - максимальное число временных лагов, рассматриваемых в модели;

p - порядок авторегрессии;

q - порядок процесса скользящей средней.

Как уже отмечалось выше, интеграция означает, в какой степени ряд должен быть преобразован с помощью разностей различного порядка, чтобы стать стационарным. Это очень важно, так как многие методы анализа временных рядов подразумевают, что анализируемый ряд в действительности является стационарным. Провер-

⁴ Ljung, G.M. and Box, G.E.P. «On a Measure of Lack Of Fit in Time Series Models». – Biometrika, 1978, № 66. – P. 67-72.

ка стационарности производится при помощи теста единичного корня (unit root). Если говорить более строго, то проверка стационарности производится на основе анализа корней характеристического уравнения (единичный корень соответствует границе области стационарности). Если данные показывают единичный корень, то ряд является $I(1)$.

Подход к проверке стационарности и степени интеграции называют критерием Дики-Фуллера (DF)⁵. С помощью этого критерия проверяется, имеет ли коэффициент α в уравнении (1) значение, равное единице или меньше единицы:

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t, \quad (6.6)$$

Если α равно единице, то данные имеют единичный корень и степень интегрирования равна 1, т.е. ряд является $I(1)$. Если же α меньше единицы и больше нуля, то ряд стационарен, т.е. $I(0)$. Обычно α бывает не больше 1, поскольку это подразумевает взрывные ряды. Такие ряды маловероятны, поскольку давление экономической среды не позволяет переменной принимать бесконечно большие значения. Существуют некоторые теоретические проблемы с уравнением (6.6), поскольку возможность нестационарности нарушает допущения регрессии МНК, которая подразумевает постоянную дисперсию остатков.

Тогда требуется уравнение, выражающее изменения Y_t следующим образом:

$$\Delta Y_t = \beta Y_{t-1} + e_t, \quad (6.7)$$

где $\beta = (\alpha - 1)$.

Если β равно нулю, то говорят, что ряд Y обладает единичным корнем и является $I(1)$, и ряд ΔY будет стационарным. Если же β меньше нуля, то есть α меньше единицы, то сам ряд Y является стационарным, $I(0)$.

Проблемы при проверке стационарности, когда существует автокорреляция остатков, решаются применением расширенного критерия Дики-Фуллера (DF). При использовании этого метода прошлые значения независимой переменной включаются в уравнение регрессии с лагом, достаточным для того, чтобы избавиться от автокорреляции остатков. Это уравнение может иметь вид:

$$\Delta Y_t = \alpha_0 + \beta_1 Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_n \Delta Y_{t-n} + e_t. \quad (6.8)$$

Все вышеизложенное проиллюстрируем примером конкретного расчета.

Пример 6. Имеются данные о численности студентов в вузах Ростовской области (1958-1998 гг.) (файл **example_06.xls**). В созданном файле изначально присутствуют четыре переменные: STUDVSE – общая численность студентов (чел.);

⁵ Dickey, D.A. and Fuller, W.A. «Distribution of the Estimators for Autoregressive Time Series with a Unit Root». – Journal of the American Statistical Association, 1979, № 74. – P. 427-431.

STUDDNEV – количество студентов на дневной форме обучения (чел.); STUDVECH – количество студентов на вечерней форме обучения (чел.); STUDZAO – количество студентов на заочной форме обучения (чел.). Фрагмент окна с данными представим на рис. 73.

obs	STUDDNEV	STUDVECH	STUDVSE	STUDZAO
1958	30223.00	4250.000	50310.00	15837.00
1959	27781.00	4828.000	50850.00	18241.00
1960	27239.00	5797.000	54177.00	21141.00
1961	28174.00	7212.000	60216.00	24830.00
1962	29321.00	8416.000	67611.00	29874.00
1963	30680.00	9855.000	74392.00	33857.00
1964	33129.00	11167.00	81006.00	36710.00
1965	33345.00	12404.00	85004.00	39255.00
1966	35789.00	12735.00	88549.00	40025.00
1967	39049.00	11711.00	89909.00	39149.00
1968	41450.00	12251.00	91411.00	37710.00
1969	44565.00	12131.00	93914.00	37218.00
1970	47259.00	11891.00	94199.00	35049.00
1971	49455.00	11518.00	94870.00	33897.00
1972	52410.00	11425.00	97590.00	33755.00
1973	55013.00	11005.00	99075.00	33057.00
1974	56321.00	11167.00	101188.00	33700.00
1975	57380.00	11188.00	101842.00	33274.00
1976	58251.00	11160.00	102437.00	33026.00
1977	59029.00	11164.00	103818.00	33625.00
1978	59408.00	11218.00	106380.00	36785.00

Рис. 73. Исходные данные

Учитывая иллюстрационный характер проводимых вычислений, дальнейшие расчеты будем производить с использованием данных о численности студентов по дневной форме обучения (STUDDNEV). Представим исходные цифры в виде линейного графика (рис. 74).

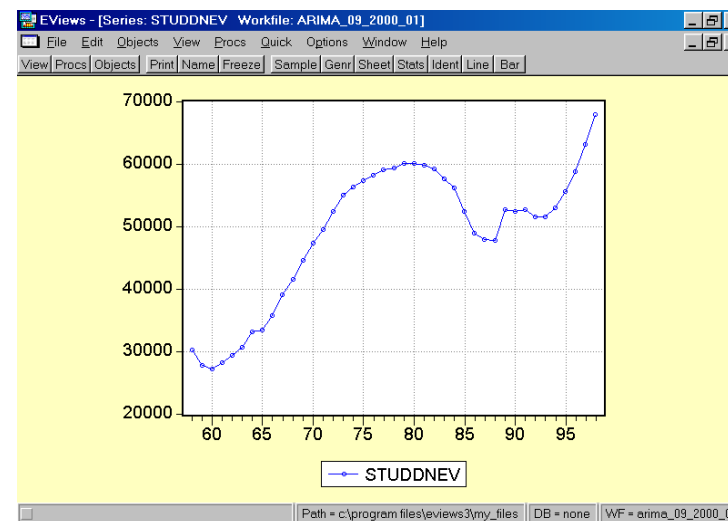


Рис. 74. Динамика численности студентов дневной формы обучения

Для приведения ряда к стационарному найдем первые разности, введя переменную DSTUDDNEV. Визуализируем полученные данные (рис. 75)

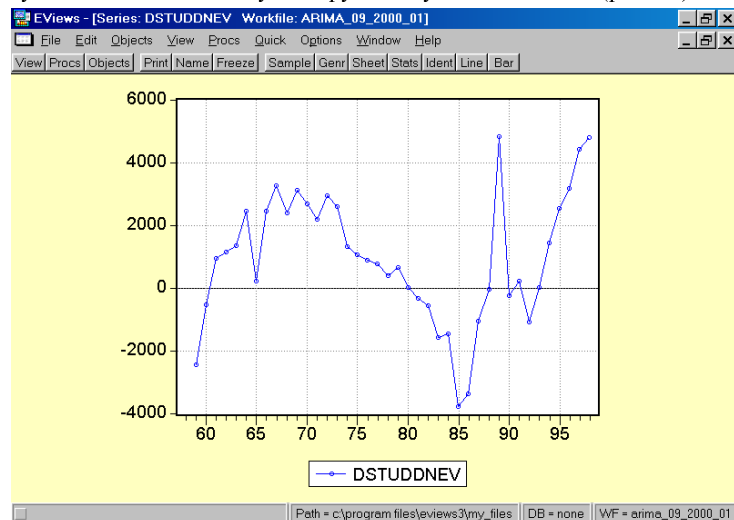


Рис. 75. Первые разности переменной STUDDNEV

Стационарность данной переменной проверим с помощью критерия DF (рис.

76).

ADF Test Statistic	-1.756750	1% Critical Value*	-3.6117
		5% Critical Value	-2.9399
		10% Critical Value	-2.6080

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(DSTUDDNEV)
 Method: Least Squares
 Date: 09/17/00 Time: 14:13
 Sample(adjusted): 1961 1998
 Included observations: 38 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DSTUDDNEV(-1)	-0.246303	0.140204	-1.756750	0.0877
D(DSTUDDNEV(-1))	-0.103576	0.170774	-0.606508	0.5481
C	388.4561	269.8087	1.439746	0.1588
R-squared	0.131069	Mean dependent var		140.5526
Adjusted R-squared	0.081416	S.D. dependent var		1551.744
S.E. of regression	1487.234	Akaike info criterion		17.52288
Sum squared resid	77415306	Schwarz criterion		17.65216
Log likelihood	-329.9347	F-statistic		2.639700
Durbin-Watson stat	1.945933	Prob F-statistic		0.085554

Рис. 76. Проверка переменной DSTUDDNEV с помощью критерия DF

Как видно из рис. 75 и рис. 76, переменная DSTUDDNEV не является стационарной.

В файле данных создадим новую серию D2STUDDNEV, как первые разности переменной DSTUDDNEV (см. рис. 77).

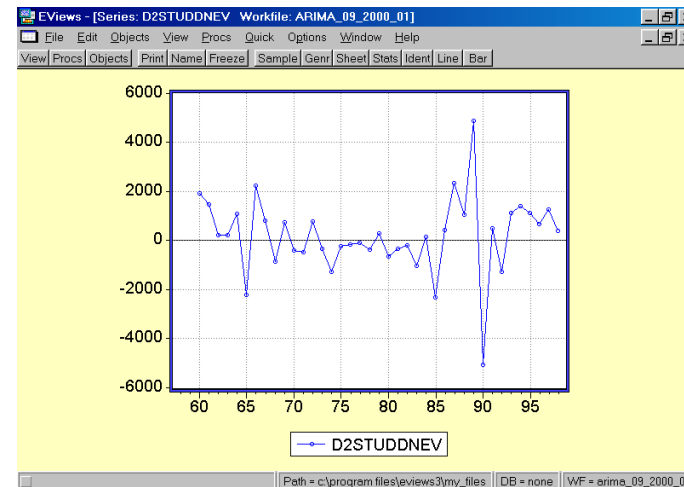


Рис. 77. Первые разности переменной DSTUDDNEV

Даже на глаз видно, что новый ряд является стационарным. Подтвердим наше предположение с помощью расчета критерия DF (рис. 78).

ADF Test Statistic	-4.359137	1% Critical Value*	-3.6171
		5% Critical Value	-2.9422
		10% Critical Value	-2.6092

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(D2STUDDNEV)
 Method: Least Squares
 Date: 09/17/00 Time: 14:19
 Sample(adjusted): 1962 1998
 Included observations: 37 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2STUDDNEV(-1)	-1.143815	0.262395	-4.359137	0.0001
D(D2STUDDNEV(-1))	-0.096798	0.165913	-0.583428	0.5635
C	122.0629	255.3559	0.478011	0.6357
R-squared	0.640568	Mean dependent var		-29.89189
Adjusted R-squared	0.619425	S.D. dependent var		2490.451
S.E. of regression	1536.378	Akaike info criterion		17.58985
Sum squared resid	80255511	Schwarz criterion		17.72046
Log likelihood	-322.4122	F-statistic		30.29691
Durbin-Watson stat	1.989844	Prob(F-statistic)		0.000000

Рис. 78. Проверка переменной D2STUDDNEV с помощью критерия DF

Теперь есть все основания отвергнуть нулевую гипотезу о наличии единичного корня (ADF t-статистика больше критических значений критерия на всех трех уровнях значимости). Другими словами, ряд D2STUDDNEV является стационарным.

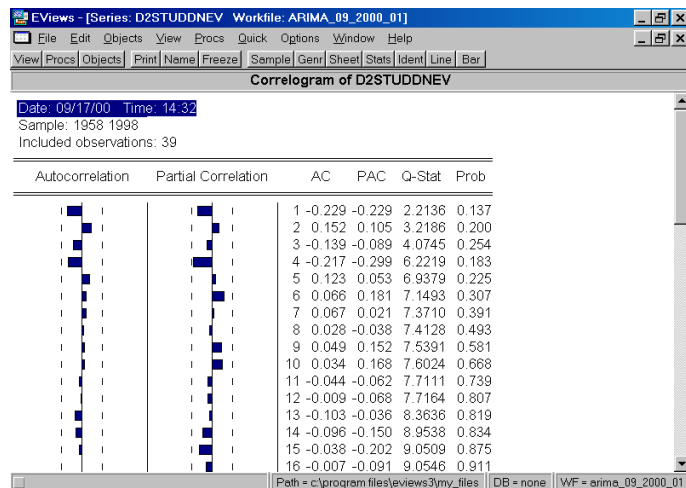


Рис. 79. Коррелограмма переменной D2STUDDNEV

Перед определением порядка авторегрессионного процесса полезным будет рассмотрение коррелограммы (рис. 79). Можно предположить наличие авторегрессии максимального порядка 3.

Проведем оценивание авторегрессионной модели скользящей средней (ARMA). Для этого в командном окне пакета необходимо набрать: *LS D2STUDDNEV C AR(2) AR(3) MA(2)*. То есть, мы подразумеваем модель вида *ARMA(3,2)* с исключенными периодами *ar(1)* и *ma(1)*. Итоговый расчет представлен на рис. 80.

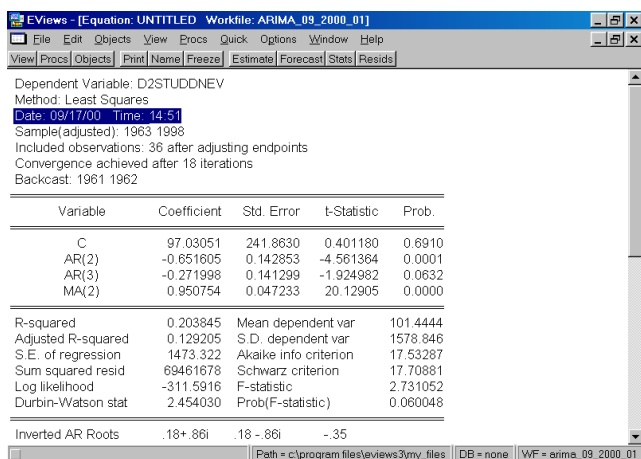


Рис. 80. Результаты расчетов по *ARMA(3,2)*.

Получены довольно значимые наблюдаемые статистики t-критериев, которые могут быть приняты на уровне значимости 6,5%.

График остатков такой модели представлен на рис. 81.

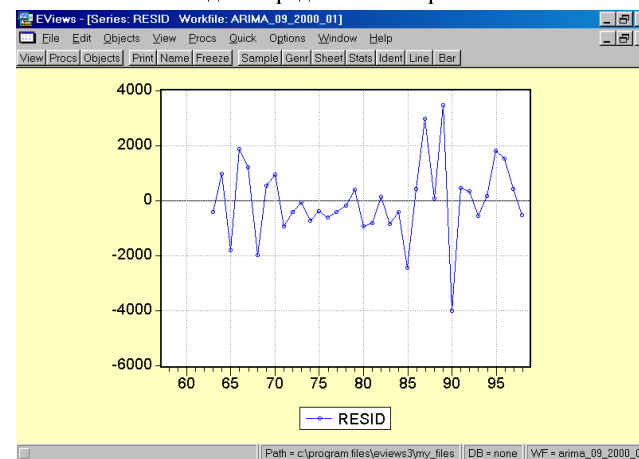


Рис. 81. Остатки уравнения *ARMA(3,2)*.

Анализ рис. 81 позволяет сделать предположение о его стационарности.

СОДЕРЖАНИЕ

1. Предисловие	3
2. Практическое занятие № 1. «Знакомство с эконометрическим пакетом Eviews»	4
3. Практическое занятие № 2. «Применение Eviews при построении и анализе линейной однофакторной модели регрессии»	18
4. Практическое занятие № 3. «Применение Eviews при построении и анализе линейной однофакторной модели регрессии»	32
5. Практическое занятие № 4. «Применение Eviews при построении и анализе многофакторной модели регрессии. Выявление мультиколлинеарности и гетероскедастичности в модели. Проверка спецификации модели»	36
6. Практическое занятие № 5. «Фиктивные переменные»	46
7. Практическое занятие № 6. «Однофакторные стохастические модели динамических процессов»	48