

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Ректор

Дата подписания: 22.04.2021 14:59:06

Уникальный программный ключ:

c098bc0c1041cb2a4cf926cf171d6715d99a6aa0ad48a27b55cbe1e2db47c79

Методические указания по выполнению лабораторных работ

1. Начало работы в пакете обработки статистических данных Eviews

Статистический пакет Econometric Views является универсальным программным инструментом для анализа, моделирования и прогнозирования экономических объектов, явлений и процессов в различных предметных областях. Первый релиз программы относится к 1981 году. Пакет имеет дружественный оконный интерфейс стандартного приложения Windows, а также возможности использования командной строки и написания пользовательских программ. В пакете реализованы все основные статистические модели и методы, а, начиная с версии 6, и возможность работы с панельными данными.

Запуск Eviews осуществляется так же, как и обычного приложения Windows:

1 способ. Двойной щелчок левой кнопкой мыши на иконке пакета Eviews.

2 способ. Пуск/ Программы/ EViews 6/ EViews 6.

3 способ. Открыть существующий рабочий файл.

При запуске пакета появится окно, представленное на рисунке 1, включающее следующие элементы:

1) Заголовок программы (**Title Bar**). Если в настоящий момент окно, содержащее пакет, является активным, то заголовок будет темнее остальных. При переключении в другое окно цветовая окраска заголовка изменит цвет.

2) Главное меню (**Main Menu**). При нажатии на соответствующие клавиши появляется раскрывающееся меню (drop-down menu). Доступные в настоящий момент опции являются затемненными. Те пункты, с которыми в настоящий момент работа невозможна, приглушены.

3) Командная строка (**Command Window**). В ней происходит непосредственный набор команд, которые выполняются после нажатия клавиши **Enter (Ввод)**. Для исполнения многих команд отсутствует необходимость их набора – надо выбрать нужный пункт в главном меню.

4) Рабочая область (**Work Area**). Большая часть экрана пакета отведена под рабочую область. В ней размещаются рабочие объекты, переключение между которыми осуществляется с помощью клавиши **F6**.

5) Строка текущего состояния (**Status Line**) пакета (рабочий каталог, текущий файл и др.). Левая секция содержит сообщение для пользователя; далее указан путь, по которому программа будет искать данные. Последние две правые секции отражают имена файла данных и рабочего файла.

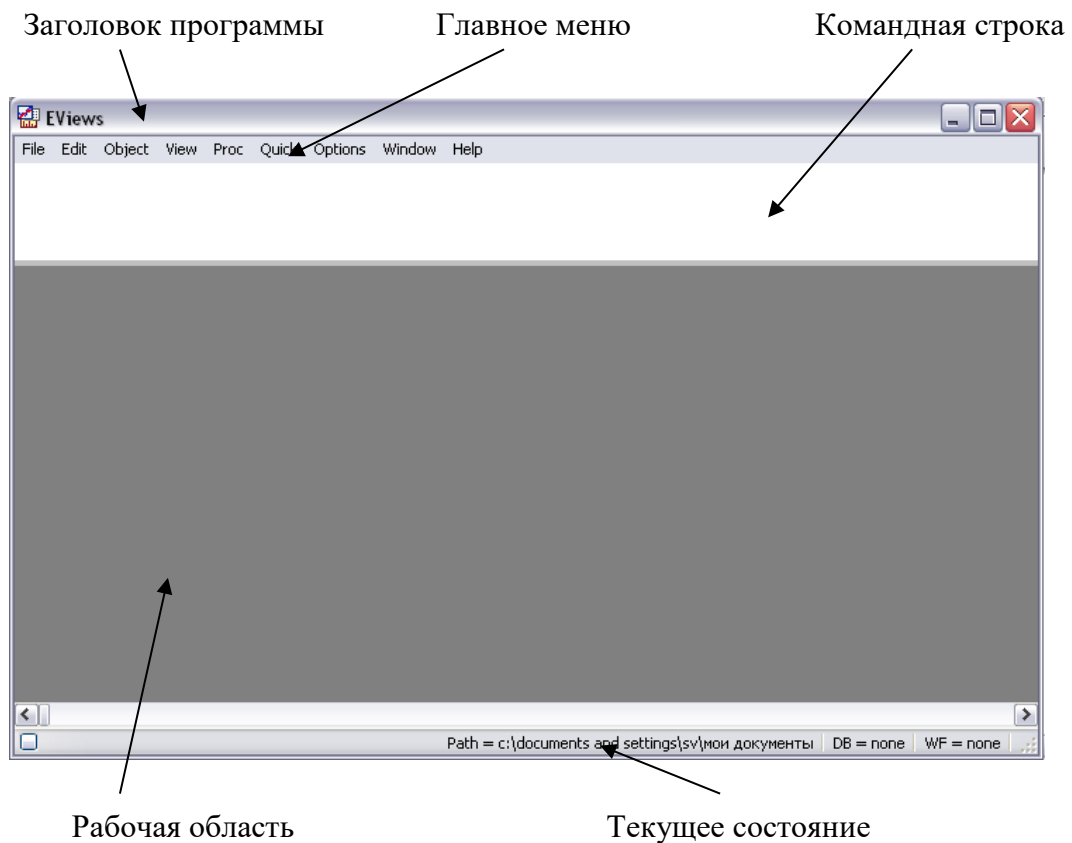


Рис. 1. Стартовое окно пакета EViews 6

Завершение работы с пакетом осуществляется путем выбора в главном меню 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 главного меню **Help**). Страничка пакета в сети интернет расположена по адресу <http://www.eviews.com>.

После запуска пакета необходимо сформировать файл данных (рабочий файл). Для этого в главном меню выбираем **File\New\Workfile...** и в появившемся окне задаем параметры исходных данных. Слева (в поле **Workfile structure type**) указываем тип структуры данных (неструктурированные (unstructured), датированные (dated), панельные (balanced panel)), справа (в поле **Date specification**) – частоту данных, начальную и конечную даты.

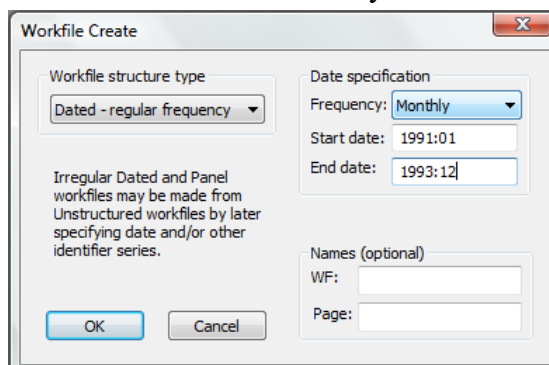
Отметим, что EViews позволяет работать с восьмью типами данных:

- **Годовые** (Annual) – годы 20 века идентифицируются по последним двум цифрам (97 эквивалентно 1997), для данных, относящихся, например, к 21 веку 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.

Рассмотрим процесс создания рабочего файла на примере. Файл *ex1.xls* формата Excel содержит следующую информацию по трем переменным: *gdp* – произведенный ВВП по рыночной стоимости в текущих ценах, трлн. руб., *m0* – денежный агрегат М0, трлн. руб., *r* – ставка рефинансирования ЦБ, % в год. Данные являются ежемесячными с 1991 по 1993 годы, поэтому указываем тип данных и частоту как показано на скриншоте:



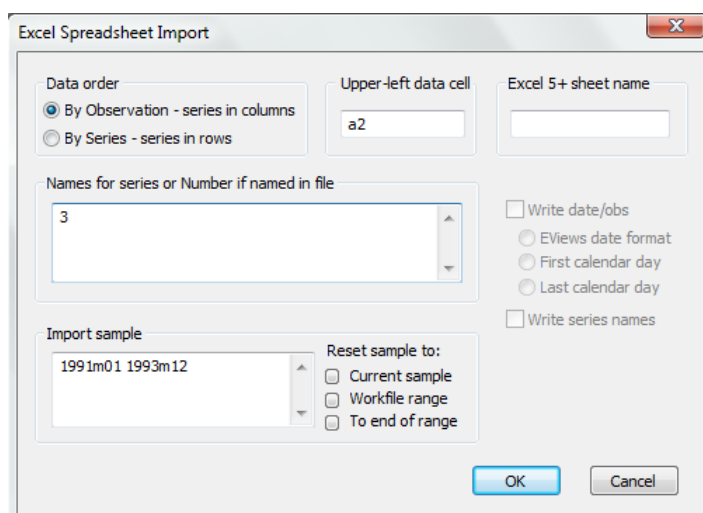
После нажатия кнопки **ОК** в рабочей области появится рабочий файл, содержащий два элемента: вектор коэффициентов *C* и вектор ошибок *RESID*.

Для работы с данными необходимо предварительно ввести их в рабочий файл, где они будут храниться в виде набора переменных (серий).

Если у Вас уже имеется файл с исходными данными, например в формате Excel, то можно осуществить их импорт в EViews. Для этого необходимо в главном меню выбрать **File\Import\Read Text-Lotus-Excel ...** и в появившемся окне указать:

- 1) в поле *Data order* – порядок представления переменных (*in columns* – в столбцах, *in rows* – в строках);
- 2) в поле *Upper-left data cell* – адрес левой верхней ячейки в исходном файле Excel, с которой начинаются данные;
- 3) в поле *Names for series or Number if named in file* – число импортируемых переменных (если их имена уже содержатся в исходном файле). Если же имена переменных в исходном файле не заданы, то их можно ввести латинскими буквами вместо числа импортируемых переменных. 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*).

Вернемся к рассматриваемому примеру. Чтобы осуществить импорт переменных *gdp*, *m0* и *r* из файла *ex1.xls* в окне импорта необходимо установить следующее:



После нажатия кнопки ОК исходные данные переместятся в рабочую область (в рабочем файле появятся их имена).

Также можно открыть файл другого формата непосредственно в Eviews, воспользовавшись опциями меню **File\Open\ForeignData as Workfile...**

Необходимые переменные можно создать и «вручную». Для этого в меню рабочего файла требуется выбрать опцию **Object\New object\Series**, задать имя создаваемой переменной и заполнить её соответствующими числовыми значениями.

На первом этапе работы с данными посмотрим на характеристики отдельных переменных. Например, двойной клик мышью по иконке *gdp* в рабочем файле откроет лист со значениями этой переменной в новом окне. С помощью пунктов меню этого окна **View** и **Proc** можно получить различные характеристики переменной. Для вычисления, например, основных описательных статистик необходимо выбрать **View\Descriptive Statistics & Tests\Stats Table** (в таблице **Mean** – математическое ожидание; **Median** – медиана; **Maximum** – максимальное значение; **Minimum** – минимальное значение; **Std. Dev.** – стандартное отклонение; **Skewness** – коэффициент асимметрии; **Kurtosis** – эксцесс; **Observations** – количество наблюдений). Для просмотра графика переменной выбрать **View/Graph...** и затем в левом подменю **Graph Options** появившегося окна указать тип графика **Line & Symbol**. Гистограмма распределения переменной выводится в этом же окне при выборе **View\Descriptive Statistics & Tests\Histogram and Stats** (рис. 2).

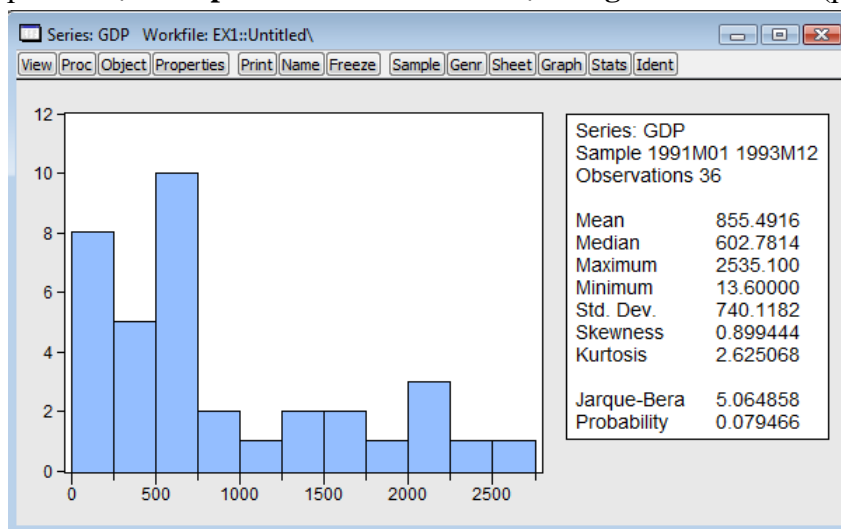
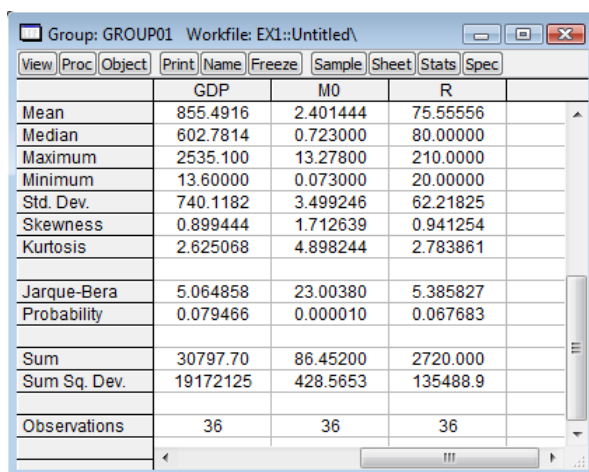


Рис. 2. Гистограмма и основные статистики распределения переменной *gdp*

Для работы с несколькими переменными одновременно необходимо создать группу, содержащую эти переменные: кликнуть мышкой по имени первой переменной (например, *gdp*), затем, **удерживая** клавишу **CTRL** кликнуть по переменным *m0* и *r*. Все выделенные переменные на экране будут затемнены. Затем необходимо подвести курсор мыши на затемненную область экрана, кликнуть правой кнопкой мыши и выбрать опцию **Open** и затем **as Group**. По умолчанию, данные будут представлены в виде электронной таблицы. Для задания имени объекта – группы надо в меню группы нажать **кнопку Name**. По умолчанию имя будет задано как *group01*. Для переменных в группе реализован ряд возможностей. В частности, построение графиков для каждой переменной на одной плоскости **View\Graph...** и далее опция в левой колонке **Line & Symbol**. Альтернативно, выбрав **View\Graph...** и **Multiple graphs** в ниспадающем меню во вкладке **Multiple series** справа в диалоговом окне, получим графики каждой переменной на отдельных плоскостях. Выбор **View\Descriptive Stats\Individual Samples** позволяет получить таблицу описательных статистик, вычисленных для каждой переменной (серии) из группы (рис. 3).

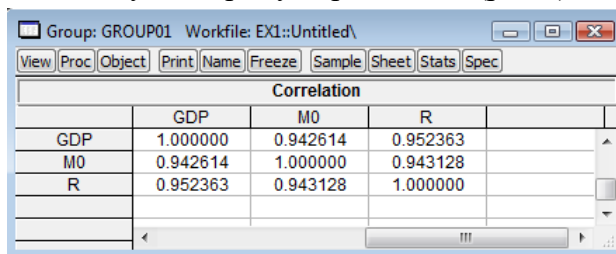


	GDP	M0	R
Mean	855.4916	2.401444	75.55556
Median	602.7814	0.723000	80.00000
Maximum	2535.100	13.27800	210.0000
Minimum	13.60000	0.073000	20.00000
Std. Dev.	740.1182	3.499246	62.21825
Skewness	0.899444	1.712639	0.941254
Kurtosis	2.625068	4.898244	2.783861
Jarque-Bera	5.064858	23.00380	5.385827
Probability	0.079466	0.000010	0.067683
Sum	30797.70	86.45200	2720.000
Sum Sq. Dev.	19172125	428.5653	135488.9
Observations	36	36	36

Рис. 3. Описательные статистики для переменных *gdp*, *m0* и *r*

Если выбрать опцию **Common Samples**, то пакет использует только наблюдения, доступные (не имеющие пропусков) для всех переменных в группе.

Клик по **View/Covariance Analysis...** и выбор **Correlation** в кнопке-флажке позволяет построить корреляционную матрицу переменных (рис.4).



Correlation			
	GDP	M0	R
GDP	1.000000	0.942614	0.952363
M0	0.942614	1.000000	0.943128
R	0.952363	0.943128	1.000000

Рис. 4. Корреляционная матрица переменных *gdp*, *m0* и *r*

Важным этапом предварительного статистического анализа данных является проверка гипотезы о законе распределения переменной. Для проверки на нормальность

распределения переменной *gdp* необходимо открыть эту переменную в отдельном окне и выполнить **View/Descriptive Statistics & Tests/Empirical Distribution Tests...** При этом, если оценки параметров распределения (μ и σ) рассчитываются по имеющимся данным, то соответствующие поля оставляются пустыми и тогда статистика Lilliefors'a соответствует статистике Колмогорова.

The screenshot shows the EViews software interface. On the left is the 'EDF Test' dialog box with 'Normal' distribution selected. On the right is the 'Empirical Distribution Test for GDP' results window.

EDF Test Dialog Box:

- Test Specification: Estimation Options
- Distribution: Normal
- Formula: $f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$
- Parameters: μ and σ input fields.

Empirical Distribution Test for GDP Results:

Hypothesis: Normal
Date: 12/29/09 Time: 12:10
Sample: 1991M01 1993M12
Included observations: 36

Method	Value	Adj. Value	Probability
Lilliefors (D)	0.222467	NA	0.0001
Cramer-von Mises (W2)	0.311203	0.315525	0.0002
Watson (U2)	0.275409	0.279234	0.0003
Anderson-Darling (A2)	1.697277	1.735584	0.0002

Method: Maximum Likelihood - d.f. corrected (Exact Solution)

Parameter	Value	Std. Error	z-Statistic	Prob.
MU	855.4916	123.3530	6.935311	0.0000
SIGMA	740.1182	88.46105	8.366600	0.0000

Log likelihood: -288.4269
No. of Coefficients: 2

Mean dependent var.: 855.4916
S.D. dependent var.: 740.1182

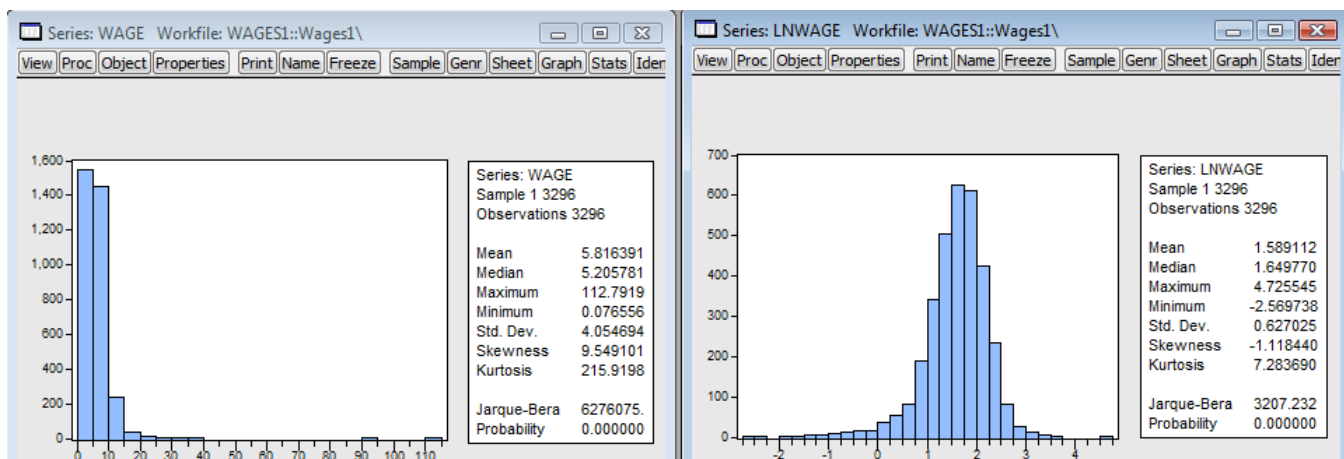
Исходные данные желательно измерять в величинах одного порядка. Для этого, как правило, переменные стандартизируют с помощью вычитания среднего и деления на стандартное отклонение, например, набрав в командной строке команду **series gdpn=(gdp-@mean(gdp))/@stdevp(gdp)** и нажав **ENTER**, получим новую стандартизованную переменную *gdpn* с нулевым математическим ожиданием и единичной дисперсией.

Часто зависимая переменная имеет распределение близкое к логнормальному, в этом случае логарифм такой переменной будет иметь распределение близкое к нормальному (log – натуральный логарифм): **series lgdp=log(gdp)**.

2. Построение регрессии

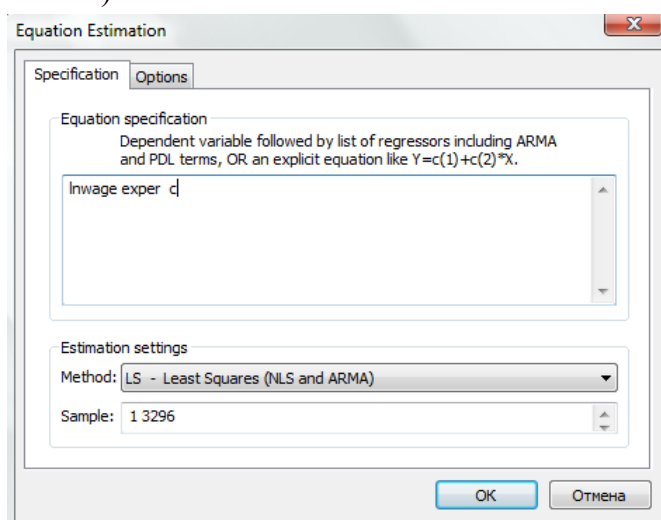
Откроем файл *wages1*: в меню **File/Foreign Data as Workfile ...** выбрав необходимый файл и далее, следуя указаниям подменю. В окне переменных вы увидите четыре переменные: *exper* – стаж работы в годах, *male* – пол: 1 – для мужчин, 0 – для женщин, *school* – число лет образования, *wage* – доход в 1980 году, \$/час. Файл содержит наблюдения по 3296 американским индивидам (данные National Longitudinal Survey).

Предварительный анализ данных рекомендуется выполнить самостоятельно. Одним из результатов такого анализа является факт логнормального распределения переменной *wage*. Поэтому будем использовать новую переменную *lnwage*, полученную после логарифмирования переменной *wage*.



Построим парную регрессию, выбрав в качестве зависимой переменной *lnwage*, а в качестве регрессора *exper*. Регрессию можно построить несколькими способами:

1. В командной строке задать команду `ls lnwage c exper`, где *ls* – оператор применения метода наименьших квадратов, *lnwage* – зависимая переменная, *c* – константа, *exper* – независимая переменная.
2. Выбрать пункты меню **Quick\Estimate\Equation...** и в окне **Equation specification** указать `ls lnwage c exper` и нажать ОК.
3. Выделив последовательно переменные **lnwage** и **exper** (выделение необходимо начинать с зависимой переменной), нажать правую кнопку мыши для вывода контекстного меню, выбрать **Open\as Equation ...** (при этом константа в спецификацию будет добавлена автоматически) и нажать ОК.



Появится окно с уравнением регрессии, рядом необходимых статистик и дополнительной информацией. Если в окне уравнения нажать кнопку **Name**, то можно задать имя, под которым уравнение будет сохранено в рабочем файле (по умолчанию eq1).

	Coefficient	Std. Error	t-Statistic	Prob.
EXPER	0.024067	0.004751	5.066272	0.0000
C	1.395566	0.039722	35.13305	0.0000
R-squared	0.007732	Mean dependent var		1.589112
Adjusted R-squared	0.007431	S.D. dependent var		0.627025
S.E. of regression	0.624691	Akaike info criterion		1.897487
Sum squared resid	1285.446	Schwarz criterion		1.901189
Log likelihood	-3125.059	Hannan-Quinn criter.		1.898812
F-statistic	25.66711	Durbin-Watson stat		1.825442
Prob(F-statistic)	0.000000			

В представленном окне eq01:

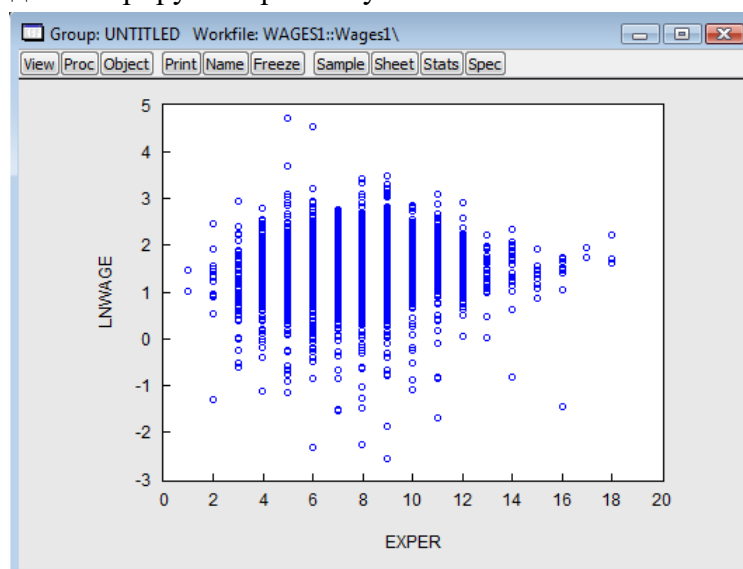
Coefficient – значения коэффициента регрессии (в строке EXPER) и свободного члена (в строке C); *Std. Error* – стандартные ошибки параметров уравнения регрессии; *t-Statistic* – наблюдаемые значения критерия Стьюдента для соответствующих параметров уравнения регрессии; *Prob.* – вероятность получения соответствующего значения t-статистики (предельный уровень значимости), зная которую можно принять или отклонить гипотезу о статистической не значимости соответствующего параметра регрессионной модели (против двусторонней альтернативной гипотезы): например, если тест выполняется на 5%-ом уровне значимости, то значение Prob., большее 0,05, свидетельствует о необходимости принять основную гипотезу о статистической не значимости соответствующего параметра; *R-squared* – коэффициент детерминации; *Adjusted R-squared* – скорректированный коэффициент детерминации; *S.E. of regression* – стандартная ошибка оценки уравнения регрессии; *Sum squared resid* – сумма квадратов остатков; *Log likelihood* – значение логарифма функции правдоподобия; *F-statistic* – наблюдаемое значение критерия Фишера; *Prob (F-statistic)* – вероятность получения соответствующего значения F-статистики; *Mean dependent var* – среднее значение зависимой переменной; *S.D. dependent var* – стандартное отклонение зависимой переменной; *Akaike info criterion*, *Schwarz criterion*, *Hannan-Quinn criter.* – значение соответствующего информационного критерия, основанного на логарифме функции правдоподобия; *Durbin-Watson stat.* – наблюдаемое значение статистики Дарбина-Уотсона, используемое для проверки гипотезы об отсутствии автокорреляции в остатках.

Опция **View\Representations** позволяет получить вид построенного уравнения.

Выбрав в окне уравнения вкладку Estimate, получим возможность переопределить уравнение, включив в него квадрат переменной *exper*.

	Coefficient	Std. Error	t-Statistic	Prob.
EXPER	0.162804	0.021420	7.600415	0.0000
EXPER*EXPER	-0.008439	0.001271	-6.640035	0.0000
C	0.869896	0.088458	9.833982	0.0000
R-squared	0.020842	Mean dependent var	1.589112	
Adjusted R-squared	0.020247	S.D. dependent var	0.627025	
S.E. of regression	0.620645	Akaike info criterion	1.884794	
Sum squared resid	1268.463	Schwarz criterion	1.890346	
Log likelihood	-3103.140	Hannan-Quinn criter.	1.886781	
F-statistic	35.04647	Durbin-Watson stat	1.826978	
Prob(F-statistic)	0.000000			

Чтобы построить график зависимости логарифма дохода от стажа работы выделим последовательно в окне рабочего файла переменные *exper* и *lnwage*, далее выберем с помощью правой кнопки мыши опцию **Open as Group View\Graph...XY Line**. Построенный график демонстрирует выраженную нелинейность.



Обе построенные регрессии являются значимыми по критерию Фишера, а коэффициенты регрессий значимы по критерию Стьюдента. Вместе с этим коэффициент детерминации мал.

В окне уравнения имеется ряд возможностей.

По **View\Actual, Fitted, Residual** получаем либо таблицу значений, либо графики фактического и подогнанного значений зависимой переменной и остатков регрессии.

Для анализа остатков уравнения получим серию значений остатков регрессии, выбрав в окне уравнения **Proc\Make Residual Series ...** и задав имя для переменной (по умолчанию *resid01*). Просмотр графика остатков в зависимости от изменения *exper* (рис. 5) позволяет визуальнo убедиться в непостоянстве дисперсии остатков (методы коррекции модели в этом случае обсуждаются ниже).

Можно получить рассчитанные по модели значения объясняемой переменной, выбрав в окне уравнения **Forecast** и задав имя переменной (по умолчанию в нашем примере *lnwagef*).

Для того, чтобы получить уравнения регрессии в зависимости от значений категоризированной переменной пола *male*, необходимо в окне рабочего файла выбрать

Sample и в опции **IF condition** задать условие для отбора наблюдений $male=0$. Получим уравнение регрессии для женщин (табл. 1).

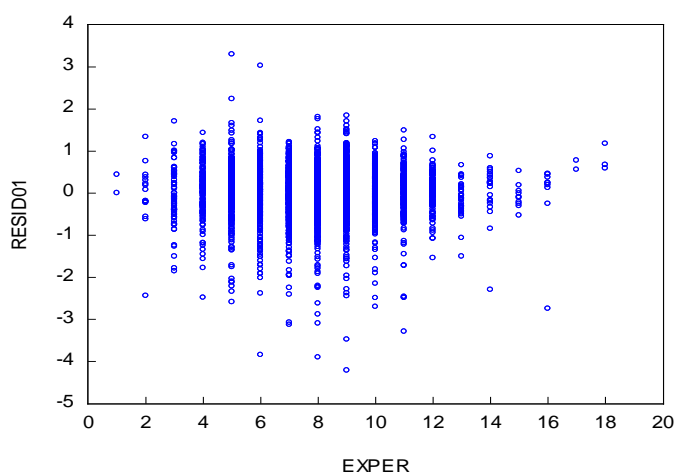


Рис. 5. График остатков регрессионной модели

Таблица 1

Результаты оценивания регрессионной модели зависимости логарифма дохода от стажа работы и его квадрата (пол – женский)

Dependent Variable: LNWAGE
 Method: Least Squares
 Sample: 1 3296 IF MALE=0
 Included observations: 1569

	Coefficient	Std. Error	t-Statistic	Prob.
EXPER	0.219526	0.035052	6.262859	0.0000
EXPER*EXPER	-0.011950	0.002238	-5.339764	0.0000
C	0.548093	0.134057	4.088519	0.0000
R-squared	0.033275	Mean dependent var		1.474751
Adjusted R-squared	0.032041	S.D. dependent var		0.630775
S.E. of regression	0.620587	Akaike info criterion		1.885609
Sum squared resid	603.1115	Schwarz criterion		1.895855
Log likelihood	-1476.261	Hannan-Quinn criter.		1.889418
F-statistic	26.95149	Durbin-Watson stat		1.937693
Prob(F-statistic)	0.000000			

Аналогично, изменив условие на $male=1$, получим уравнение регрессии для мужчин (табл. 2). При этом значение коэффициента детерминации не высоко, несмотря на значимость уравнения в целом.

Для женщин стаж работы, при котором максимален логарифм дохода, будет составлять $0,219-2 \cdot 0,012 \text{exper}=0$ и $\text{exper}^*=9,1$ лет, а для мужчин $0,116-2 \cdot 0,006 \text{exper}=0$ и $\text{exper}^*=9,7$ лет.

Таблица 2

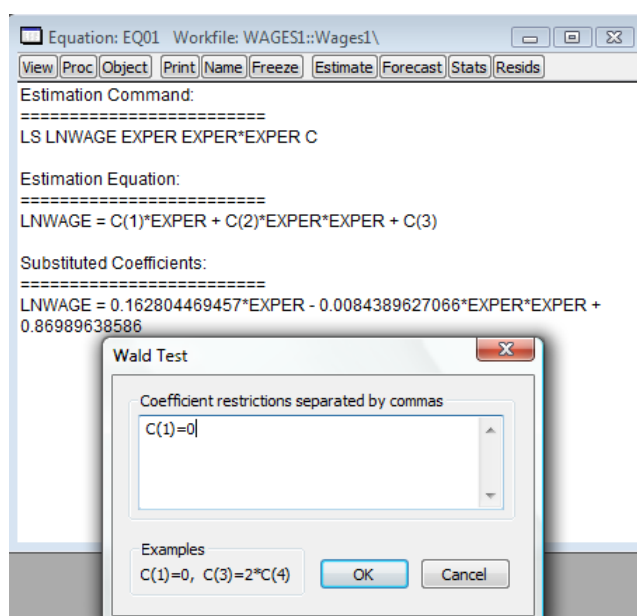
Результаты оценивания регрессионной модели зависимости логарифма дохода от стажа работы и его квадрата (пол – мужской)

Dependent Variable: LNWAGE
 Method: Least Squares
 Sample: 1 3296 IF MALE=1
 Included observations: 1727

	Coefficient	Std. Error	t-Statistic	Prob.
--	-------------	------------	-------------	-------

EXPER	0.115865	0.027955	4.144659	0.0000
EXPER*EXPER	-0.006462	0.001580	-4.091131	0.0000
C	1.212294	0.121068	10.01334	0.0000
R-squared	0.009893	Mean dependent var	1.693011	
Adjusted R-squared	0.008744	S.D. dependent var	0.605320	
S.E. of regression	0.602668	Akaike info criterion	1.826835	
Sum squared resid	626.1720	Schwarz criterion	1.836310	
Log likelihood	-1574.472	Hannan-Quinn criter.	1.830340	
F-statistic	8.612843	Durbin-Watson stat	1.828111	
Prob(F-statistic)	0.000190			

Вернемся к построенному ранее по всей выборке уравнению eq1, предварительно убрав в **Sample** условие $male=1$. Для проверки гипотез о коэффициентах регрессии служит опция **View\Coefficients tests\Wald – Coefficient Restrictions**. Например, для проверки значимости коэффициента при *exper* окно ограничения будет иметь вид:



В результате получим значимость коэффициента на 1% уровне.

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	57.76631	(1, 3293)	0.0000
Chi-square	57.76631	1	0.0000

Для гипотезы о том, что оба коэффициента регрессии равны нулю $C(1)=C(2)=0$ (т.е. коэффициенты при *exper* и *exper*exper* равны нулю одновременно или уравнение регрессии не значимо) получим результат:

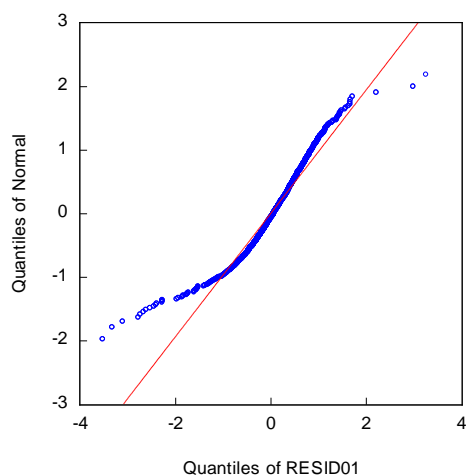
Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	35.04647	(2, 3293)	0.0000

Chi-square 70.09294 2 0.0000

Таким образом, гипотеза о не значимости уравнения регрессии отвергается.

Одним из способов тестирования нормальности распределения остатков регрессии является построение квантиль-квантильных графиков распределения.



Для полученной переменной *resid01* построим график **View\Graph...** и далее тип графика **Quantile-Quantile**. Визуально получаем, что квантили для остатков регрессии достаточно далеко расположены от квантилей нормального распределения на концах интервала.

Задание для самостоятельной работы.

В файле *flat.wfl* содержатся данные о стоимости квартир в Москве. Переменные: *totsp* – общая площадь, кв. м., *price* – цена квартиры, тыс. дол., *rooms* – количество жилых комнат, *livsp* – жилая площадь, кв. м., *kitsp* – площадь кухни, кв. м., *dist* – расстояние до центра, км., *metrdist* – расстояние до ближайшей станции метро, мин., *walk* – если до метро пешком – 1, иначе – 0, *brick* – если дом кирпичный – 1, иначе – 0, *tel* – если есть телефон – 1, иначе – 0, *bal* – если есть балкон или лоджия – 1, иначе – 0, *floor* – если этаж первый или последний – 0, иначе – 1.

Постройте модель парной регрессии стоимости квартиры в зависимости от общей площади. Проверьте качество построенной модели. Проинтерпретируйте результаты. Сделайте выводы.

3. Спецификация уравнения множественной регрессии

Традиционной задачей эмпирических исследований в микроэкономике является оценивание кривых Энгеля. Эрнст Энгель установил, что при увеличении дохода семьи доля расходов на питание уменьшается (закон Энгеля). В современных микроэкономических терминах это означает, что эластичность расходов на питание по доходу меньше единицы.¹ При этом имеют в виду, что еда является необходимым

¹Напомним, что эластичность *y* по *x* - это $\frac{x}{y} \frac{dy}{dx}$ или $d[\ln y]/d[\ln x]$. Она показывает, на сколько процентов изменяется *y*, если *x* увеличивается на 1%.

товаром, а не предметом роскоши. Зависимость расходов на приобретение некоторого вида товара от доходов называется *кривой Энгеля*. В настоящее время принято, как правило, вместо дохода рассматривать полные расходы.

Файл `expend.wfl` содержит данные, полученные из архива данных журнала *Journal of Applied Econometrics*. Для наших целей мы взяли данные о годовых расходах на питание, отдых и другие товары за период с октября 1986 г. по сентябрь 1987 г. (427 наблюдений). Описание переменных: *f3* – расходы на питание семьи в голландских гульденах (Dfl); *v3* – расходы на отдых, Dfl; *tot3* – полные расходы, Dfl; *prov* – провинция, в которой проживает семья; *reg* – регион проживания; *scl* – социальный класс (1 – нижний класс, ..., 5 – верхний класс); *nahm* – число членов семьи старше 11 лет; *durb* – степень урбанизации (1 – маленькая деревня, ..., 13 – большой город); *nch06* – число детей младше 6 лет в семье; *nch711* – число детей от 7 до 11 лет; *nch1217* – число детей от 12 до 17 лет; *nch18* – число детей старше 18 лет.

Вычислим описательные статистики для всех переменных (в окне группы **View\Descriptive Stats\Individual Samples**).

Статистика	F3	TOT3	PROV	REG	SCL	NAHM	DURB	NCH06	NCH1217	NCH18	NCH711
Среднее	6890,88	35762,81	7,28	3,37	3,16	2,18	8,26	0,27	0,23	1,12	0,25
Медиана	6838,43	31847,12	7,00	4,00	3,00	2,00	8,00	0,00	0,00	1,00	0,00
Максимум	17270,75	124564,20	13,00	6,00	5,00	5,00	13,00	3,00	2,00	4,00	3,00
Минимум	883,98	8728,86	1,00	1,00	1,00	1,00	3,00	0,00	0,00	0,00	0,00
Стандартное отклонение	2745,74	18053,31	3,02	1,42	1,09	0,88	2,95	0,63	0,54	1,04	0,57
Асимметрия	0,29	1,47	-0,25	-0,03	-0,44	1,01	0,11	2,33	2,29	0,24	2,29
Экцесс	3,23	6,42	2,23	1,67	2,21	3,88	1,70	7,79	7,07	1,96	7,56

Для определения тесноты линейной связи между некоторыми из переменных вычислим корреляционную матрицу (**Open\as Group\View\Covariance Analysis** и опция **Correlation** в появившемся меню) для переменных *f3*, *tot3* и *nahm*:

Переменные	F3	Tot3	Nahm
F3	1,00	0,50	0,51
Tot3	0,50	1,00	0,33
Nahm	0,51	0,33	1,00

Как и ожидалось, линейная связь имеется между расходами на питание и полными расходами семьи, а также между расходами на питание и числом членов семьи, что выражается в высоких значениях коэффициентов корреляции (0,50 и 0,51 соответственно).

Построим линейную регрессионную модель для объяснения логарифма расходов на питание *lf3* в зависимости от логарифма полных расходов *ltot3*:

$$lf3 = \beta_0 + \beta_1 \cdot ltot3 + u.$$

Результаты оценивания коэффициентов модели методом наименьших квадратов представлены в таблице 3.

Таблица 3

Результаты оценивания регрессионной модели зависимости логарифма расходов на питание от логарифма полных расходов

Dependent Variable: LF3 Method: Least Squares				
Sample: 1 427 Included observations: 427				
	Coefficient	Std. Error	t-Statistic	Prob.
LTOT3	0.549213	0.039579	13.87632	0.0000
C	3.046584	0.410839	7.415526	0.0000

R-squared	0.311799	Mean dependent var	8.741329
Adjusted R-squared	0.310180	S.D. dependent var	0.475891
S.E. of regression	0.395253	Akaike info criterion	0.986094
Sum squared resid	66.39575	Schwarz criterion	1.005095
Log likelihood	-208.5310	Hannan-Quinn criter.	0.993599
F-statistic	192.5523	Durbin-Watson stat	1.988620
Prob(F-statistic)	0.000000		

Оцененная эластичность расходов на питание по доходу (точнее по полным расходам, которые являются прокси-переменной, т.е. переменной заменителем для доходов) равна 0,54. Доверительный 95% интервал для оцененной эластичности имеет вид:

$$\beta = \hat{\beta} \pm t_{\varepsilon/2, n-2} \sqrt{V(\hat{\beta})} = 0,54 \pm 1,960 \cdot 0,04 \text{ или } 0,462 \leq \beta \leq 0,618.$$

Однако в оцененной регрессии не принимается во внимание размер семьи. Это приводит к смещенности оценки, поскольку полные доходы зависят от количества взрослых членов семьи. Поэтому мы, возможно, переоцениваем эластичность по доходу.

Построим регрессию $lf3$ на переменные $ltot3$, $nahm$, $nch06$, $nch711$, scl , $durb$ и константу. Получим согласно результатам расчетов, представленным в таблице 4, регрессию:

$$lf3 = 4,21 + 0,38ltot3 + 0,04scl + 0,11nch711 + 0,14nch06 + 0,18nahm - 0,01dur.$$

Таблица 4

Результаты оценивания регрессионной модели зависимости логарифма расходов на питание от логарифма полных расходов, социального класса, числа членов семьи младше 6 лет, от 7 до 11 лет, старше 11 лет и степени урбанизации

Dependent Variable: LF3	Method: Least Squares			
Sample: 1 427	Included observations: 427			
	Coefficient	Std. Error	t-Statistic	Prob.
LTOT3	0.389599	0.041336	9.425186	0.0000
SCL	0.037727	0.016537	2.281301	0.0230
NCH711	0.115297	0.029163	3.953605	0.0001
NCH06	0.136342	0.026470	5.150799	0.0000
NAHM	0.187000	0.021102	8.861905	0.0000
DURB	-0.012629	0.005763	-2.191474	0.0290
C	4.210594	0.442655	9.512145	0.0000
R-squared	0.497318	Mean dependent var	8.741329	
Adjusted R-squared	0.490137	S.D. dependent var	0.475891	
S.E. of regression	0.339809	Akaike info criterion	0.695390	
Sum squared resid	48.49740	Schwarz criterion	0.761895	
Log likelihood	-141.4658	Hannan-Quinn criter.	0.721658	
F-statistic	69.25302	Durbin-Watson stat	1.971984	
Prob(F-statistic)	0.000000			

Как мы и предполагали, эластичность уменьшилась и стала равна 0,38. Проверим, отличается ли влияние количества детей в возрасте до 6 лет от влияния их количества в возрасте от 7 до 11 лет, т.е. равенство коэффициентов при переменных $nch711$ и $nch06$.

Тест Вальда дает результаты, не позволяющие говорить о значимых различиях во влиянии количества детей разных возрастов на величину потребления продуктов.

Wald Test:

Null Hypothesis: $C(3)=C(4)$

F-statistic	0.271716	Probability	0.602457
Chi-square	0.271716	Probability	0.602183

Построим график зависимости переменной $lf3$ от переменной $ltot3$ (рис. 6).

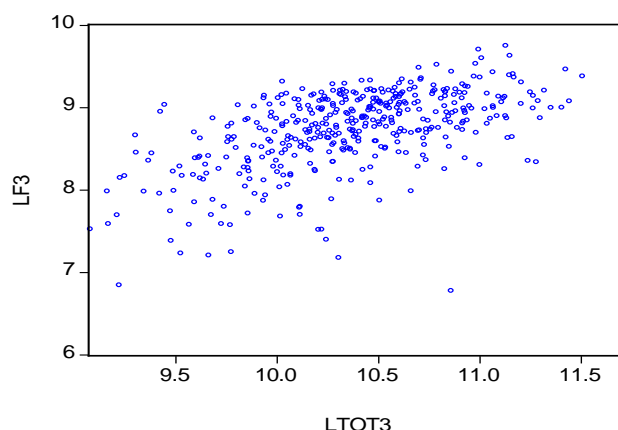


Рис. 6. График зависимости логарифма расходов на питание от логарифма полных расходов

На графике видна параболическая зависимость от переменной $ltot3$, поэтому уточним спецификацию модели, включив в нее квадрат указанной переменной.

Построим регрессию $lf3$ на переменные $ltot3$, $ltot3s$, $nahm$, $nch06$, $nch711$, scl , $durb$ и константу (переменная $ltot3s$ получена как $ltot3s=ltot3 \cdot ltot3$). Результаты представлены в таблице 5.

Таблица 5

Результаты оценивания регрессионной модели зависимости логарифма расходов на питание от логарифма полных расходов, его квадрата, социального класса, числа членов семьи младше 6 лет, от 7 до 11 лет, старше 11 лет и степени урбанизации

Dependent Variable: LF3
 Method: Least Squares
 Sample: 1 427
 Included observations: 427

	Coefficient	Std. Error	t-Statistic	Prob.
LTOT3	3.498138	1.066525	3.279938	0.0011
LTOT3S	-0.150076	0.051452	-2.916794	0.0037
SCL	0.031979	0.016510	1.936963	0.0534
NCH711	0.109975	0.028963	3.797111	0.0002
NCH06	0.124841	0.026531	4.705400	0.0000
NAHM	0.183839	0.020944	8.777831	0.0000
DURB	-0.013064	0.005714	-2.286442	0.0227
C	-11.81793	5.512742	-2.143749	0.0326
R-squared	0.507322	Mean dependent var	8.741329	
Adjusted R-squared	0.499091	S.D. dependent var	0.475891	
S.E. of regression	0.336812	Akaike info criterion	0.679973	
Sum squared resid	47.53227	Schwarz criterion	0.755978	
Log likelihood	-137.1742	Hannan-Quinn criter.	0.709993	
F-statistic	61.63619	Durbin-Watson stat	2.011364	
Prob(F-statistic)	0.000000			

Воспользуемся полученными результатами для оценивания эластичности при различных уровнях дохода $tot3$. Эластичность расходов на питание будет равна:

$$\frac{dlf3}{dltot3} (3,49 \cdot ltot3 - 0,15 \cdot ltot3^2) = 3,49 - 2 \cdot 0,15 \cdot ltot3.$$

Для среднего дохода (35762 гульдена) она составит $3,49 - 2 \cdot 0,15 \cdot \ln(35762) = 0,34$, что меньше единицы, т.е. спрос на продукты питания неэластичен и еда является необходимым товаром.

Построим график зависимости остатков регрессии (переменная e) от $ltot3$ (рис. 7).

График не дает права предполагать независимость ошибок регрессии и регрессора $ltot3$, что свидетельствует о гетероскедастичности остатков в построенной модели регрессии.

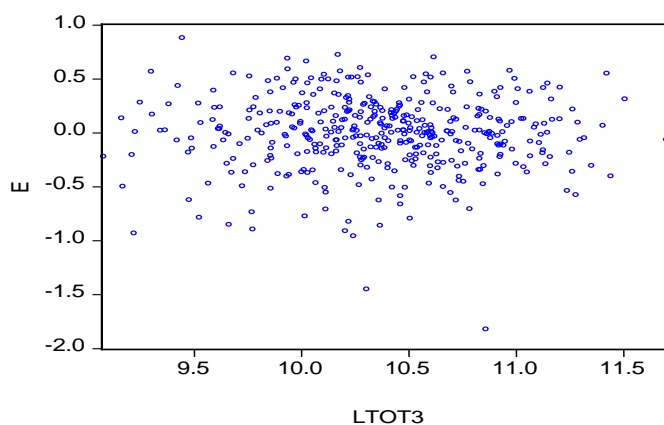


Рис. 7. График остатков регрессионной модели

Этот факт ожидаем, так как ясно, что вариация расходов на питание для семей с различным уровнем дохода будет непостоянна, а именно: для семей с низким уровнем доходов дисперсия расходов на питание больше, а для семей с высоким уровнем доходов дисперсия расходов на питание – меньше.

Тестирование на гомоскедастичность тестом Уайта (в окне уравнения **View\Residual Tests\Heteroskedasticity Tests...\White** оставив флажок для опции **Include White cross terms**) показывает (табл. 6), что дисперсия случайной ошибки модели непостоянна и, следовательно, стандартные ошибки построенного уравнения регрессии нуждаются в коррекции.

Таблица 6

Результаты теста Уайта

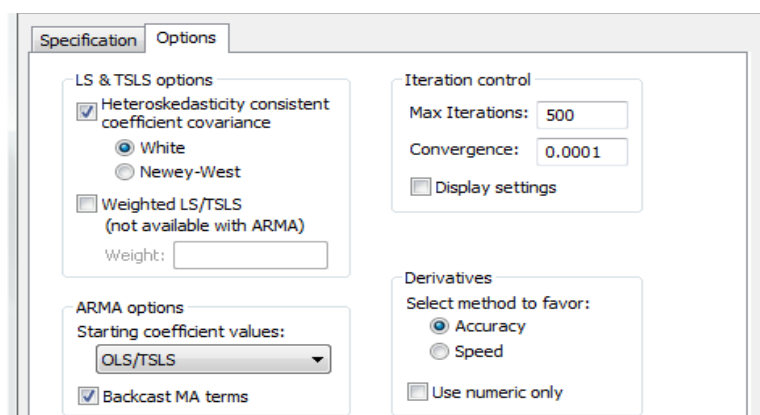
Heteroskedasticity Test: White			
F-statistic	1.632281	Prob. F(33,393)	0.0171
Obs*R-squared	51.47072	Prob. Chi-Square(33)	0.0212
Scaled explained SS	110.9721	Prob. Chi-Square(33)	0.0000

Результаты статистики Дарбина-Уотсона не дают повода усомниться в отсутствии автокорреляции остатков ($DW = 2,01$). Вывод не изменяется и в случае непосредственного тестирования в окне уравнения путем **View\Residual Tests\Serial Correlation LM Test...** и, указав в окне для **Lags to include** количество лагов 1:

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.034320	Prob. F(1,418)	0.8531

Obs*R-squared 0.035056 Prob. Chi-Square(1) 0.8515

Для состоятельного оценивания стандартных ошибок коэффициентов (с учетом гетероскедастичности остатков) воспользуемся процедурой Уайта. Нажав кнопку **Estimate** в окне уравнения и выбрав затем вкладку **Options** необходимо отметить **Heteroskedasticity consistent coefficient covariance**:



Результаты представлены в табл. 7.

Таблица 7

Результаты регрессионного анализа с использованием процедуры Уайта

Dependent Variable: LF3

Method: Least Squares

Sample: 1 427

Included observations: 427

White Heteroskedasticity-Consistent Standard Errors & Covariance

	Coefficient	Std. Error	t-Statistic	Prob.
LTOT3	3.498138	1.153464	3.032723	0.0026
LTOT3S	-0.150076	0.055281	-2.714797	0.0069
SCL	0.031979	0.016716	1.913049	0.0564
NCH711	0.109975	0.020076	5.477827	0.0000
NCH06	0.124841	0.021937	5.690881	0.0000
NAHM	0.183839	0.023273	7.899304	0.0000
DURB	-0.013064	0.005688	-2.296756	0.0221
C	-11.81793	5.995256	-1.971214	0.0494
R-squared	0.507322	Mean dependent var		8.741329
Adjusted R-squared	0.499091	S.D. dependent var		0.475891
S.E. of regression	0.336812	Akaike info criterion		0.679973
Sum squared resid	47.53227	Schwarz criterion		0.755978
Log likelihood	-137.1742	Hannan-Quinn criter.		0.709993
F-statistic	61.63619	Durbin-Watson stat		2.011364
Prob(F-statistic)	0.000000			

Результаты практически соответствуют предыдущим, однако стандартные ошибки увеличились.

Задание для самостоятельной работы.

В файле flat.wfl содержатся данные о стоимости квартир в Москве. Постройте модель стоимости квартиры (или стоимости квадратного метра жилой площади квартиры) в зависимости от имеющихся параметров. Тестируйте гипотезу, что уравнения для цены

различны для 1, 2, 3-4 комнатных квартир, т.е. гипотезу, что рынок распадается на рынки однокомнатных, двухкомнатных и трех-четырекомнатных квартир.

4. Фиктивные переменные в уравнении регрессии

Изучим взаимосвязь между зарплатой преподавателей (y), количеством лет опыта работы на должности преподавателя (x_1) и гендерной принадлежностью преподавателя (x_2). Файл с данными ex2.wf1.

Построим линейную регрессионную модель вида

$$y = \beta_0 + \beta_1 x_1 + u.$$

Получим результаты (табл. 8) с помощью метода наименьших квадратов и пакета прикладных программ Eviews.

Таблица 8

Результаты оценивания уравнения регрессии

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 12				
Included observations: 12				
	Coefficient	Std. Error	t-Statistic	Prob.
X1	1223.258	202.6682	6.035768	0.0001
C	18655.24	714.1479	26.12238	0.0000
R-squared	0.784624	Mean dependent var		22630.83
Adjusted R-squared	0.763087	S.D. dependent var		1964.066
S.E. of regression	955.9841	Akaike info criterion		16.71437
Sum squared resid	9139055.	Schwarz criterion		16.79519
Log likelihood	-98.28623	Hannan-Quinn criter.		16.68445
F-statistic	36.43050	Durbin-Watson stat		2.400780
Prob(F-statistic)	0.000126			

Таким образом, с увеличением опыта работы на 1 год, зарплата увеличивается на 1223\$.

Учтем гендерную принадлежность преподавателей, т.е. оценим регрессию вида $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ (табл. 9). Поскольку переменная x_2 принимает значение либо 1, либо 0, то при $x_2 = 1$ получим $y = (\beta_0 + \beta_2) + \beta_1 x_1$, а при $x_2 = 0$ получим $y = \beta_0 + \beta_1 x_1$. Т.е. две линии регрессии имеют одинаковый наклон, но различные смещения и представимы двумя параллельными прямыми на графике.

Таблица 9

Результаты оценивания уравнения регрессии

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 12				
Included observations: 12				
	Coefficient	Std. Error	t-Statistic	Prob.
X1	1108.140	59.55298	18.60762	0.0000
X2	-1707.597	162.1840	-10.52877	0.0000
C	19883.18	236.9680	83.90660	0.0000
R-squared	0.983827	Mean dependent var		22630.83
Adjusted R-squared	0.980233	S.D. dependent var		1964.066
S.E. of regression	276.1358	Akaike info criterion		14.29198

Sum squared resid	686258.9	Schwarz criterion	14.41321
Log likelihood	-82.75188	Hannan-Quinn criter.	14.24710
F-statistic	273.7462	Durbin-Watson stat	2.440495
Prob(F-statistic)	0.000000		

График прогнозируемых значений для мужчин и женщин отдельно представлен на рис. 7.

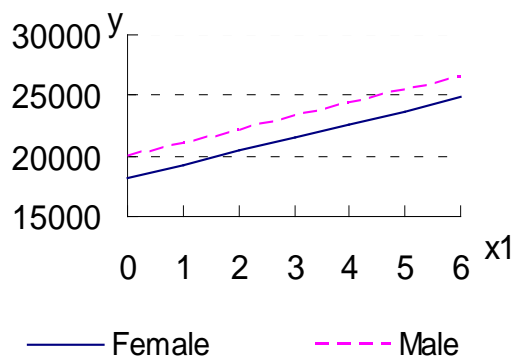


Рис. 7. Прогнозируемые значения зарплаты для мужчин и женщин

3. Учет эффект взаимодействия пола преподавателя и его опыта работы, для этого введем новую переменную $x_1x_2=x_1 \cdot x_2$ в модель регрессии: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1x_2+u$.

Интерпретация полученной регрессии может быть выполнена с учетом того факта, что переменная x_2 является фиктивной. При $x_2=1$, т.е. для женщин уравнение регрессии примет вид:

$$y=\beta_0+\beta_1x_1+\beta_2 \cdot 1+\beta_3x_1 \cdot 1=(\beta_0+\beta_2)+(\beta_1+\beta_3)x_1.$$

При $x_2=0$, т.е. для мужчин уравнение регрессии примет вид: $y=\beta_0+\beta_1x_1+\beta_2 \cdot 0+\beta_3x_1 \cdot 0=\beta_0+\beta_1x_1$.

Результаты расчетов представлены в таблице 10:

Таблица 10

Результаты оценивания уравнения регрессии				
Dependent Variable: Y	Method: Least Squares			
Sample: 1 12	Included observations: 12			
	Coefficient	Std. Error	t-Statistic	Prob.
X1	1229.130	59.37305	20.70182	0.0000
X2	-866.7101	305.2568	-2.839282	0.0218
X1*X2	-260.1304	87.05798	-2.988014	0.0174
C	19459.71	223.4720	87.07895	0.0000
R-squared	0.992357	Mean dependent var		22630.83
Adjusted R-squared	0.989491	S.D. dependent var		1964.066
S.E. of regression	201.3438	Akaike info criterion		13.70911
Sum squared resid	324314.6	Schwarz criterion		13.87074
Log likelihood	-78.25464	Hannan-Quinn criter.		13.64926
F-statistic	346.2381	Durbin-Watson stat		2.345922
Prob(F-statistic)	0.000000			

График прогнозируемых значений для мужчин и женщин отдельно представлен на рис. 8.

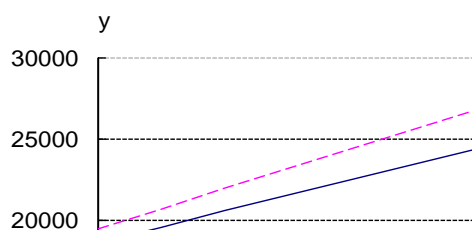


Рис. 8. Прогнозируемые значения зарплаты для мужчин и женщин

Линия уравнения регрессии для преподавателей мужчин имеет больший угол наклона, чем для женщин, и начальный уровень заработной платы также выше у мужчин.

Коэффициент детерминации, показывающий, насколько хорошо модель подогнана под имеющиеся данные, равен 0,992, т.е. 99,2% суммы квадратов отклонений у от среднего значения \bar{y} объясняется переменными модели.

Гипотеза $H_0: \beta_1=\beta_2=\beta_3=0$, согласно значению F -критерия 346,2 отвергается, поскольку табличное значение F -критерия для 3 и 8 степеней свободы и уровне значимости $\varepsilon=0,05$, будет равно $F_{0,05}(3;8)=4,07$. Таким образом, построенное уравнение регрессии значимо.

Уравнение имеет вид: $\hat{y} = 19459 + 1229x_1 - 866x_2 - 260x_1x_2$.

Тестирование каждого из параметров модели с помощью t критерия Стьюдента показывает значимость каждого из оцененных коэффициентов. Например, для коэффициента β_2 : $t = \text{оценка параметра}/\text{стандартная ошибка} = -866/305 = -2,84$. Табличное значение $t_{0,05}(8)=2,306$ и так как $2,84 > 2,306$, то $H_0: \beta_2=0$ отвергается. Отметим, что в таблице результатов содержится оцененное значение уровня значимости $\varepsilon=0,0218$. Поскольку это значение меньше 0,05, то мы при проверке гипотезы $H_0: \beta_2=0$ при уровне значимости 0,05 должны отвергнуть H_0 .

Оцененные стандартные ошибки также даны в таблице результатов, например, для переменной x_1 она составляет 59,4. Вычислим $100(1-\varepsilon)\%$ доверительный интервал для оценки указанного коэффициента по формуле $\hat{\beta}_1 \pm t_{\varepsilon/2} \sqrt{V(\hat{\beta}_1)}$. Таким образом 95% доверительный интервал для β_1 будет:

$$1229 \pm 2,306 \cdot 59,4 \text{ или } 1092 \leq \beta_1 \leq 1366.$$

Задание для самостоятельной работы. Имеются данные о стоимости колготок на московских оптовых торговых фирмах (осень, 1997 г.), всего 74 наблюдения в файле *tights.wfl*. Доступны следующие переменные: N – номер по порядку, $Price98$ – цена колготок в ценах 1997 г. с учетом деноминации, Den – плотность в DEN, $Polyamid$ – процент содержания полиамида, $lykra$ – процент содержания лайкры, $cotton$ – процент содержания хлопка, $wool$ – процент содержания шерсти, $firm$ – фирма-производитель: 0 – Levante, 1 – Golden Lady. Постройте регрессию для объяснения цены колготок.

5. Анализ временных рядов

Рассмотрим ряд помесечной динамики объема промышленного производства в РФ (млрд. руб.). Ряд содержит 66 наблюдений в файле данных example 3_1.wf1.

Выбирая переменную y (двойной щелчок мышью), получим диалоговое окно для этой переменной, содержащее строку меню. Выбирая, например, **View/Graph/Line&Symbol** (опция **General** в положении по умолчанию **Basic Graph**) можно получить график временного ряда. Задавая опции **View/Descriptive Statistics&Tests/Histogram and Stats**, получим гистограмму распределения уровней ряда и описательные статистики. Опция **View/Correlogram** позволяет получить (табл. 11) для заданного числа лагов графики и значения автокорреляционной (АКФ) и частной автокорреляционной функций (ЧАКФ), причем автоматически рассчитываются значения Q статистики.

Таблица 11

Автокорреляционная и частная автокорреляционная функции
Sample: 1997M01 2002M06
Included observations: 66

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *****	. *****	1	0.678	0.678	31.767	0.000
. *****	. .	2	0.485	0.046	48.270	0.000
. ***	. **	3	0.464	0.220	63.589	0.000
. **	** . .	4	0.223	-0.340	67.205	0.000
. *	. .	5	0.086	0.013	67.751	0.000
. .	. .	6	0.062	-0.027	68.034	0.000
. * . .	** . .	7	-0.139	-0.231	69.504	0.000
. *	8	-0.194	0.064	72.409	0.000
. * . .	. * . .	9	-0.192	-0.108	75.306	0.000
** . .	. * . .	10	-0.305	-0.067	82.752	0.000

Графики ряда и АКФ и ЧАКФ позволяют предположить наличие трендовой составляющей. Оценим параметры линейного тренда. В главном меню выберем опции **Quick/Estimate Equation** и получим диалоговое окно оценивания уравнения регрессии. В окне спецификации уравнения **Equation specification** запишем сначала зависимую переменную, затем константу и переменную времени: y **c @trend**. Функция **@trend** задает линейный тренд. По умолчанию уравнение оценивается МНК по всей выборке, нажимая ОК, получим таблицу результатов (табл. 12).

Таким образом, уравнение линейного тренда $y_t = 51,6 + 7,9t$. В окне уравнения с помощью опций можно посмотреть (**View/Representation**) вид получившегося уравнения, опция **View/Actual, Fitted, Residual** позволяет получить таблицу результатов подгонки уравнения, а также графики фактических и расчетных уровней ряда и график ошибки.

Таблица 12

Результаты оценивания линейного тренда

Dependent Variable: Y
 Method: Least Squares
 Sample: 1997M01 2002M06
 Included observations: 66

	Coefficient	Std. Error	t-Statistic	Prob.
C	51.62569	9.069274	5.692373	0.0000
@TREND	7.872590	0.240744	32.70106	0.0000
R-squared	0.943531	Mean dependent var		307.4848
Adjusted R-squared	0.942648	S.D. dependent var		155.5816
S.E. of regression	37.25900	Akaike info criterion		10.10350
Sum squared resid	88846.93	Schwarz criterion		10.16985
Log likelihood	-331.4155	Hannan-Quinn criter.		10.12972
F-statistic	1069.359	Durbin-Watson stat		0.290161
Prob(F-statistic)	0.000000			

Опция **View/Coefficient Tests** содержит подменю для тестов на коэффициенты регрессии, **View/Residual Tests** – для тестов на ошибки. Другие опции доступные в окне уравнения позволяют, в частности, выполнять прогноз по полученной модели **Proc/Forecast**; получать переменную, содержащую ошибки регрессии **Proc/Make Residual Series**; переоценивать уравнение регрессии **Estimate**.

Оценим кубический тренд (табл. 13).

Таблица 13

Результаты оценивания кубического тренда

Dependent Variable: Y
 Method: Least Squares
 Sample: 1997M01 2002M06
 Included observations: 66

	Coefficient	Std. Error	t-Statistic	Prob.
C	151.5753	8.964383	16.90862	0.0000
@TREND	-7.969950	1.203529	-6.622150	0.0000
@TREND^2	0.539372	0.043213	12.48163	0.0000
@TREND^3	-0.005028	0.000437	-11.51035	0.0000
R-squared	0.985396	Mean dependent var		307.4848
Adjusted R-squared	0.984690	S.D. dependent var		155.5816
S.E. of regression	19.25075	Akaike info criterion		8.811669
Sum squared resid	22976.67	Schwarz criterion		8.944375
Log likelihood	-286.7851	Hannan-Quinn criter.		8.864108
F-statistic	1394.519	Durbin-Watson stat		1.085465
Prob(F-statistic)	0.000000			

Уравнение тренда имеет вид $y_t = 151,6 - 7,9t + 0,5t^2 - 0,005t^3$. Сравнивая с предыдущим уравнением (табл.12) убеждаемся, что кубический тренд адекватнее отражает тенденцию уровней временного ряда. В частности, выбирая **View/Actual, Fitted, Residual/Graph**, получим график фактических и расчетных значений и график ошибок (рис. 9).

Временной ряд ошибок полученного уравнения обнаруживает явную автокорреляцию уровней. В этой связи необходима коррекция построенной модели, например, с помощью подхода Бокса-Дженкинса.

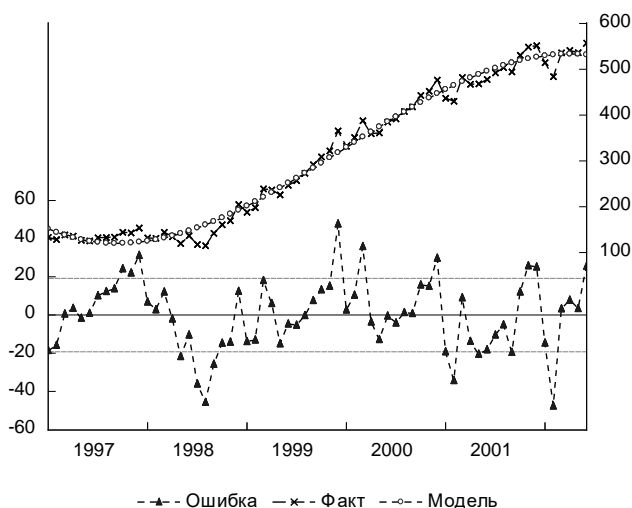


Рис. 9. График фактических, расчетных значений уровней временного ряда и ошибок

Построим уравнение показательного тренда. Прологарифмируем зависимую переменную и, задавая, в меню **Equation specification**

$$\log(y) \text{ с } @trend \ @trend^2 \ @trend^3$$

получим табл. 14.

Таблица 14

Результаты оценивания показательного тренда

Dependent Variable: LOG(Y)

Method: Least Squares

Included observations: 66

	Coefficient	Std. Error	t-Statistic	Prob.
C	4.909793	0.045252	108.4979	0.0000
@TREND	-0.022034	0.006075	-3.626662	0.0006
@TREND^2	0.002029	0.000218	9.299422	0.0000
@TREND^3	-2.14E-05	2.21E-06	-9.688444	0.0000
R-squared	0.971445	Mean dependent var		5.583225
Adjusted R-squared	0.970063	S.D. dependent var		0.561652
S.E. of regression	0.097178	Akaike info criterion		-1.765847
Sum squared resid	0.585504	Schwarz criterion		-1.633141
Log likelihood	62.27296	Hannan-Quinn criter.		-1.713409
F-statistic	703.0837	Durbin-Watson stat		0.608604
Prob(F-statistic)	0.000000			

Уравнение тренда имеет вид $\ln y_t = 4,9 - 0,02t + 0,002t^2 - 2,1 \cdot 10^{-5}t^3$. Сравнивая с предыдущим уравнением, убеждаемся, что последняя модель имеет, в частности, лучшие показатели информационных критериев Акейка и Шварца и, таким образом, может быть выбрана как наилучшая из трех построенных. Однако и для этой модели график остатков (рис. 10) и тесты на ошибки показывают необходимость моделирования ошибок. Для последующего анализа создадим ряд ошибок, выбирая в окне полученного уравнения **Proc/Make Residual Series** и задавая имя *E* для ряда ошибок.

Выполним проверку адекватности и точности модели в случае показательного тренда. В окне для переменной ошибок E с помощью опций **View/Graph** и, выбирая в окне **Specific** вид графика **Quantile-Quantile**, получим квантиль-квантильный график (рис. 11), который показывает, что за исключением первых и последних нескольких значений совокупность остальных ошибок близка к линии нормального распределения.

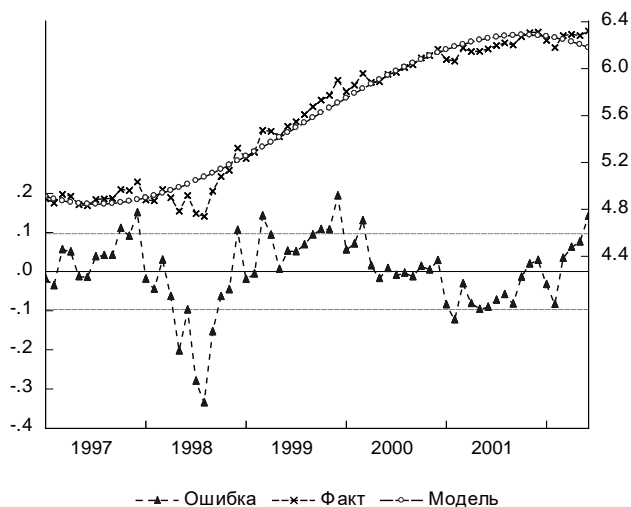


Рис. 10. График фактических, расчетных значений уровней временного ряда и ошибок

Проверка на нормальность с помощью критериев согласия осуществляется выбором в окне для переменной ошибок E опции **View/Descriptive Statistics&Tests/Empirical Distribution Tests**. В появившемся диалоговом окне по умолчанию в **Test Specification** указано нормальное (**Normal**) распределение.

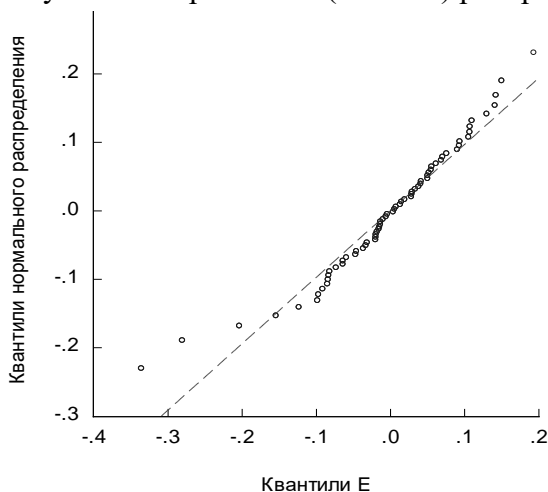


Рис. 11. Квантиль-квантильный график ошибок

Доверяя пакету получение оценок математического ожидания и дисперсии распределения (оставляя пустыми соответствующие поля в окне), после нажатия ОК, получим результаты расчета ряда критериев (табл. 15) с соответствующими вероятностями ошибки первого рода. В частности, статистики Лиллайефорса (аналог критерия Колмогорова-Смирнова в случае оценивания параметров нормального закона по выборке) и Крамера-фон Мизеса не позволяют отвергнуть гипотезу о нормальности распределения для ряда ошибок E .

При этом оцененные методом максимального правдоподобия параметры нормального закона: математическое ожидание – 0, дисперсия – 0,09.

Тестирование автокорреляции ошибок удобно выполнить как **View/Correlogram...** для уровней ряда (**Level**) и числа лагов, например, 10. Как и ожидалось значения Q теста показывают автокоррелированность уровней ряда ошибок.

Таблица 15

Результаты проверки на нормальность ошибок E

Empirical Distribution Test for E
Hypothesis: Normal
Sample: 1997M01 2002M06
Included observations: 66

Method	Value	Adj. Value	Probability
Lilliefors (D)	0.099197	NA	> 0.1
Cramer-von Mises (W2)	0.085026	0.085670	0.1750
Watson (U2)	0.067720	0.068233	0.2595
Anderson-Darling (A2)	0.663863	0.671750	0.0794

Method: Maximum Likelihood - d.f. corrected (Exact Solution)

Parameter	Value	Std. Error	z-Statistic	Prob.
MU	2.16E-16	0.011683	1.85E-14	1.0000
SIGMA	0.094909	0.008324	11.40175	0.0000

Log likelihood	62.26914	Mean dependent var.	2.14E-16
No. of Coefficients	2	S.D. dependent var.	0.094909

Для определения качества прогноза рассчитаем значения средней абсолютной процентной ошибки, средней процентной ошибки и суммы квадратов ошибок. Последний показатель имеется в результатах расчета регрессии (табл. 14) и он равен $SSE=0,59$. Остальные могут быть получены с помощью команд в командной строке пакета: **series mape=abs(e/y)*100** и **series mse=e/y*100**. После каждой команды нажимается <Enter>. Затем вычисляются средние значения для созданных переменных $MAPE$ (средняя абсолютная процентная ошибка) и MSE (средняя процентная ошибка). Получаем $MAPE = 0,03\%$, $MSE = -0,01\%$, что означает отличное качество прогноза.

Рассмотрим этот же ряд из 66 наблюдений помесечной динамики объема промышленного производства в РФ. Задавая в пакете EViews в командной строке опцию по созданию новой переменной: **series ys**, и после нажатия <Enter> осуществляя присваивание этой переменной значений скользящего среднего, полученного по формуле средней хронологической (предполагаем линейный тренд в локальной окрестности точек сглаживания) с длиной периода сглаживания по пяти точкам: **ys=@movavc(y,5)**, получим значения скользящей средней **ys**. Задавая команду **series e=y-ys**, получим значения ошибки в предположении аддитивной модели временного ряда.

Чтобы просмотреть значения одновременно уровней исходного ряда y , сглаженных ys уровней и ошибки e надо выделить в рабочей области все три переменные и, нажав правую кнопку мыши выбрать **Open/as Group**. В окне созданной группы также доступно

меню, с помощью которого (**View/Graph...**) получим график исходных и сглаженных значений и значений ошибки на одной плоскости (рис. 12).

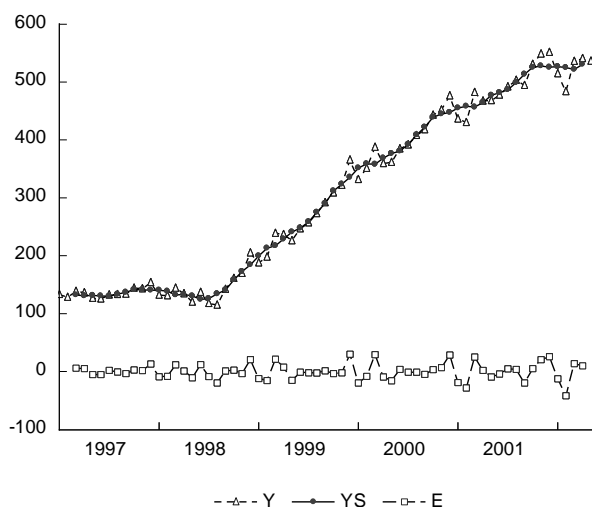


Рис. 12. График исходных, сглаженных уровней временного ряда и ошибок

По графику видно, что значения для месячной динамики объема промышленного производства неплохо аппроксимируются выбранной скользящей средней за исключением локальных пиков и спадов в конце/начале календарного года.

Рассмотрим построение модели временного ряда с сезонной компонентой. Имеются данные об индексе объема выпуска промышленной продукции в РФ – файл `example 4_1.wf1`. Визуально предполагаем наличие сезонной компоненты (рис. 13).

Поскольку амплитуда колебаний уровней ряда на графике меняется, воспользуемся мультипликативной моделью временного ряда. Применим метод скользящих средних для выделения сезонности. Для этого откроем в отдельном окне переменную с уровнями ряда `y`. Затем воспользуемся опцией **Proc\Seasonal Adjustment\Moving Average Methods...** и выберем модель **Multiplicative**. По умолчанию скорректированные на сезонность значения ряда будут сохранены в виде переменной `ysa`. Получим таблицу результатов расчета индексов сезонности для каждого месяца (табл. 16), по которой можно сделать вывод о том, что в январе индекс объема выпуска промышленной продукции меньше на 12% от уровня тренда, полученного методом скользящих средних. Также меньше, но на величину от 4% до 1% индекс для апреля, мая, июня и июля. В сторону увеличения индекс колеблется в январе – на 7% больше тренда, в марте – на 5%, в сентябре – на 4%. Очевидно, динамика индекса имеет существенную специфику, связанную с поквартальным учетом выпуска.

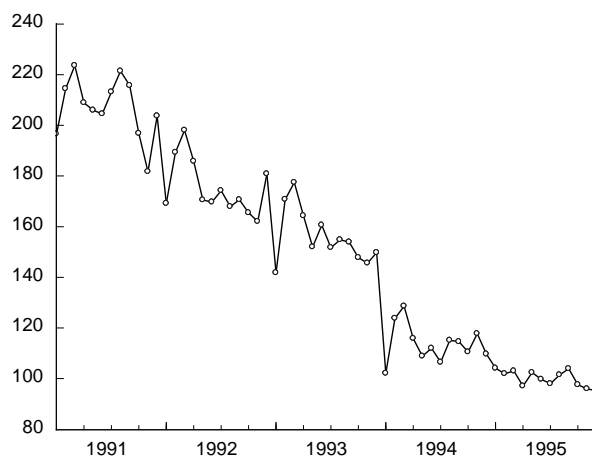


Рис. 13. График уровней временного ряда

Таблица 16

Результаты расчета индексов сезонности

Sample: 1991M01 1995M12

Included observations: 60

Ratio to Moving Average

Original Series: Y

Adjusted Series: YSA

Scaling Factors:

1	0.877655
2	1.000416
3	1.051481
4	0.989456
5	0.961017
6	0.988833
7	0.990754
8	1.029213
9	1.040182
10	1.005956
11	1.009361
12	1.069640

Динамику сезонной компоненты удобно представить наглядно в виде лепестковой диаграммы с помощью табличного процессора Excel (рис. 14). Сильно выделяются первый и последний месяцы года.

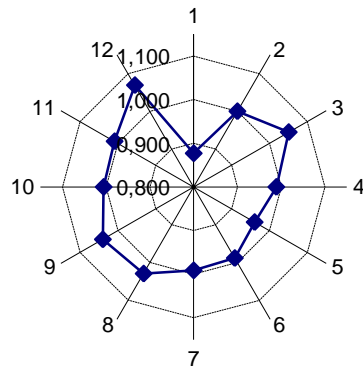


Рис. 14. Динамика сезонной компоненты

Воспользовавшись тем, что Eviews рассчитал десезонализированные уровни временного ряда, сохранив их в переменной **ysa**, построим на одном графике исходные и очищенные от сезонности уровни ряда (рис. 15). Для этого в Eviews выделим переменные **y** и **ysa** и откроем их как группу.

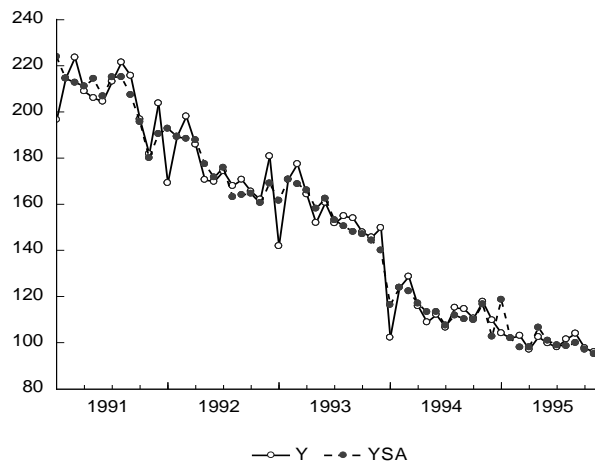


Рис.15. Исходные и десезонализированные уровни временного ряда

На следующем этапе анализа временного ряда необходимо выделить имеющийся тренд. Предполагая линейный тренд, получим с учетом коррекции стандартных ошибок по Ньюи-Весту (табл. 17), поскольку критерий Дарбина-Уотсона показывает наличие автокорреляции ошибок.

Таблица 17

Результаты оценивания линейного тренда

Dependent Variable: YSA Method: Least Squares
 Sample: 1991M01 1995M12 Included observations: 60
 Newey-West HAC Standard Errors & Covariance (lag truncation=3)

	Coefficient	Std. Error	t-Statistic	Prob.
C	218.1474	2.387174	91.38313	0.0000
@TREND	-2.294434	0.082203	-27.91164	0.0000
R-squared	0.963666	Mean dependent var		150.4616
Adjusted R-squared	0.963039	S.D. dependent var		40.81899
S.E. of regression	7.847537	Akaike info criterion		6.991042
Sum squared resid	3571.863	Schwarz criterion		7.060853
Log likelihood	-207.7313	Hannan-Quinn criter.		7.018349

F-statistic 1538.283 Durbin-Watson stat 0.819948
 Prob(F-statistic) 0.000000

Уравнение тренда $\hat{z}_t = 218,1 - 2,3t$. На рисунке 16 представлены графики исходного десеզонализированного ряда, график тренда и график, полученных вычитанием из десеզонализированных значений ряда уровней тренда. График ошибок показывает ярко выраженную автокорреляцию первого порядка.

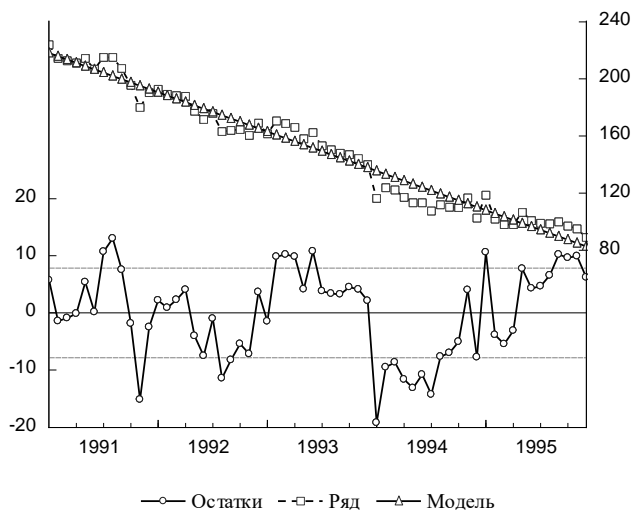


Рис. 16. График исходных уровней временного ряда, линейного тренда и остатков

В справочном виде дадим уравнение с авторегрессией первого порядка в ошибках регрессии (AR(1) модель) (табл. 18), а также график остатков модели (рис. 17).

Таблица 18

Модель авторегрессии первого порядка

Dependent Variable: YSA Method: Least Squares
 Sample (adjusted): 1991M02 1995M12
 Included observations: 59 after adjustments
 Convergence achieved after 3 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	216.8072	4.281920	50.63316	0.0000
@TREND	-2.254511	0.119919	-18.80027	0.0000
AR(1)	0.587152	0.108315	5.420792	0.0000
R-squared	0.975012	Mean dependent var		149.2170
Adjusted R-squared	0.974119	S.D. dependent var		40.00470
S.E. of regression	6.435764	Akaike info criterion		6.611127
Sum squared resid	2319.467	Schwarz criterion		6.716765
Log likelihood	-192.0283	Hannan-Quinn criter.		6.652364
F-statistic	1092.520	Durbin-Watson stat		2.213348
Prob(F-statistic)	0.000000			

Inverted AR Roots .59

Остатки полученного уравнения демонстрируют случайный характер, а значения Q статистики показывают, что исследуемый ряд остатков можно считать белым шумом.

Применяя другой способ выделения сезонности с помощью фиктивных переменных, дадим в командной строке Eviews команду:

equation eq3.ls y c @trend @expand(month, @dropfirst) ar(1)

В указанной команде **equation** означает создание нового объекта – уравнения с именем *eq3*, опция **ls** означает применение метода наименьших квадратов, *y* – зависимая переменная, *c* – константа, **@trend** – линейный тренд, **ar(1)** – авторегрессию первого порядка в остатках уравнения, опция **@expand(month, @dropfirst)** дает задание пакету создать и включить в уравнение регрессии фиктивные переменные $\lambda_1, \dots, \lambda_{12}$, которые равны 1 для выбранного месяца и 0 иначе, причем предварительно необходимо создать переменную, равную номеру месяца, к которому относится наблюдение: **series month=@month**; опция **@dropfirst** специфицирует в качестве базисной категории для набора фиктивных переменных первый месяц.

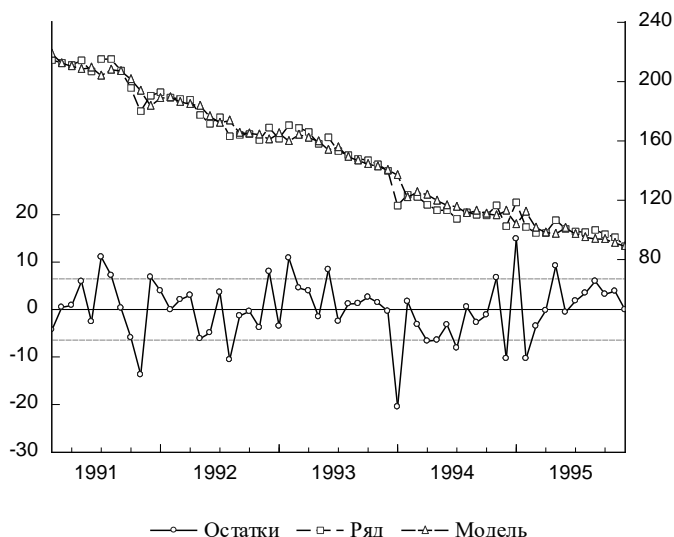


Рис. 17. График исходных уровней ряда, уравнения авторегрессии первого порядка и остатков

Получим таблицу с результатами, которые в целом достаточно похожи на представленные выше (табл. 19).

Таблица 19

Результаты выделения сезонности с помощью фиктивных переменных

Dependent Variable: Y Method: Least Squares

Sample (adjusted): 1991M02 1995M12

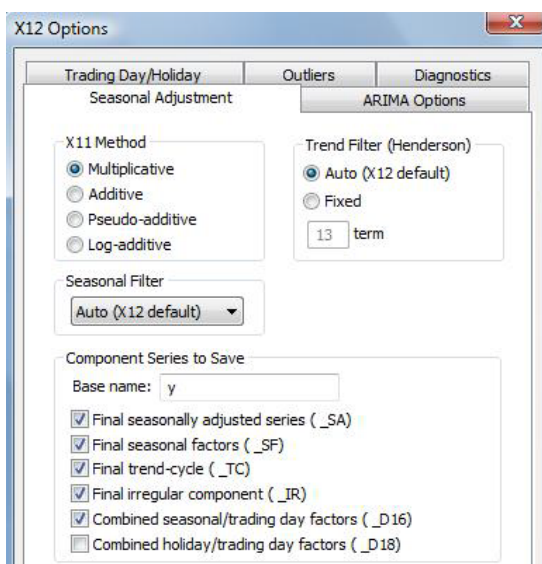
Included observations: 59 after adjustments

Convergence achieved after 3 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	198.6217	5.963277	33.30748	0.0000
@TREND	-2.299571	0.123450	-18.62751	0.0000
MONTH=2	19.26769	4.013558	4.800650	0.0000
MONTH=3	27.47454	5.006728	5.487524	0.0000
MONTH=4	17.92423	5.457690	3.284215	0.0020
MONTH=5	13.73702	5.670449	2.422563	0.0195
MONTH=6	17.39224	5.764171	3.017302	0.0042
MONTH=7	19.07915	5.790053	3.295159	0.0019
MONTH=8	24.79212	5.764681	4.300693	0.0001
MONTH=9	26.70826	5.679193	4.702827	0.0000
MONTH=10	20.90605	5.493671	3.805479	0.0004
MONTH=11	20.14469	5.109414	3.942661	0.0003
MONTH=12	29.68378	4.269068	6.953221	0.0000
AR(1)	0.520417	0.127438	4.083682	0.0002
R-squared	0.972698	Mean dependent var	149.6424	

Adjusted R-squared	0.964811	S.D. dependent var	40.61272
S.E. of regression	7.618454	Akaike info criterion	7.102725
Sum squared resid	2611.838	Schwarz criterion	7.595700
Log likelihood	-195.5304	Hannan-Quinn criter.	7.295163
F-statistic	123.3256	Durbin-Watson stat	2.214672
Prob(F-statistic)	0.000000		
Inverted AR Roots	.52		

Применим наконец процедуру Census X-12 (в окне для у выбираем **Proc\Seasonal Adjustment\Census X12...**).



В окне для опций X12 выбираем опции для декомпозиции временного ряда с учетом календарных эффектов (вкладка **Trading Day/Holiday** и опции **Adjust in X11 step** и **Flow weekday-weekend/leap year effects**) и авторегрессии (вкладка **ARIMA Options** и опции: в позиции **ARIMA Spec** выбираем **Specify in line** и ниже в окне для **In-line Specification** указываем (1 0 0), что соответствует авторегрессии для уровней ряда).

Получим результаты, принципиально совпадающие с выделенной методом скользящих средних динамикой сезонности (рис. 18).

Таким образом, в примере показано применение различных подходов для анализа тренд-сезонного временного ряда.

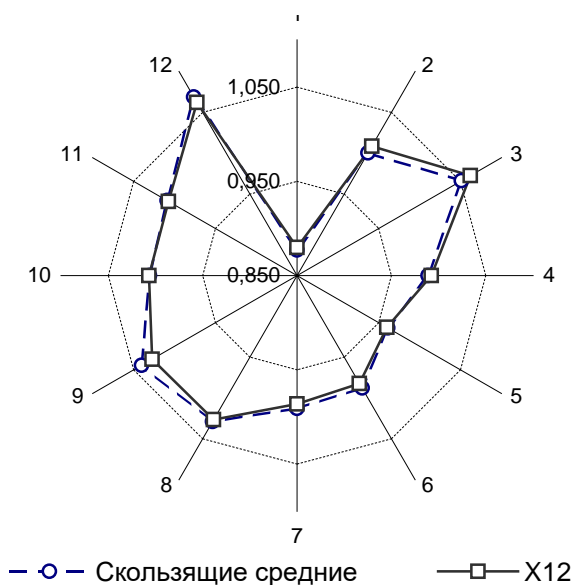


Рис. 18. Графики сезонности, полученной различными методами

Задание для самостоятельной работы. Имеются данные (информация Росстата) ежемесячной динамики производства электроэнергии в Российской Федерации в млрд. кВт-ч. Выполните анализ компонентного состава временного ряда производства электроэнергии; постройте тренд-сезонную модель производства электроэнергии и с помощью полученной модели рассчитайте прогнозную оценку производства электроэнергии в первом квартале 2002 года.

Месяц	1998	1999	2000	2001
Январь	86,6	84,7	88,9	90,6
Февраль	79	76,5	81,6	82,2
Март	79,5	81,3	81,9	83,3
Апрель	70	67,8	68,4	71,3
Май	59,6	62,3	65,2	64,7
Июнь	54,2	56,1	57,7	59,1
Июль	52,7	55,8	58,7	60,1
Август	52,9	58,2	60,4	61,7
Сентябрь	57,6	63,3	64,5	64,4
Октябрь	70,5	71,8	76,9	78,5
Ноябрь	78,4	80,8	83,4	82,5
Декабрь	85,7	87,5	90,2	92,8

