

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ФОНД ПОДГОТОВКИ КАДРОВ
ИННОВАЦИОННЫЙ ПРОЕКТ РАЗВИТИЯ ОБРАЗОВАНИЯ

Программа «Совершенствование преподавания социально-экономических дисциплин в вузах»

МИЭФ ГУ-ВШЭ

(наименование вуза)

Пособие для студентов по курсу

«Анализ временных рядов»

Москва

2003

Часть I.

Руководство по эконометрическому пакету EViews

Введение.....	3
1. Общие принципы работы в EViews	3
1.1. Создание рабочего файла EViews	4
Задания для самостоятельной работы.....	9
1.2. Общая структура рабочего файла EViews.....	9
Задания для самостоятельной работы.....	12
2. Анализ одномерных временных рядов	12
2.1. Структура окна временного ряда	12
Задания для самостоятельной работы.....	16
2.2. Построение графика временного ряда	17
Задания для самостоятельной работы.....	20
2.3. Описательные статистики временного ряда.....	20
Задания для самостоятельной работы.....	23
2.4. Тесты для описательных статистик временного ряда	23
Задания для самостоятельной работы.....	31
2.5. Построение кореллограммы временного ряда (выборочных автокорреляционной и частной автокорреляционной функций).....	31
Задания для самостоятельной работы.....	33
2.6. Моделирование процессов типа $ARIMA(p, d, q)$	33
Задания для самостоятельной работы.....	45
2.7. Тесты на единичные корни (Тест Дикки-Фуллера).....	45
Задания для самостоятельной работы.....	47
2.8. Прогнозирование при помощи моделей $ARMA(p, q)$	48
Задания для самостоятельной работы.....	51
3. Анализ многомерных временных рядов	51
3.1. Создание группы временных рядов (многомерного временного ряда).....	51
Задания для самостоятельной работы.....	51
3.2. Построение графика многомерного временного ряда.....	52
Задания для самостоятельной работы.....	54
3.3. Векторная авторегрессия.....	54
3.4. Тест на причинность по Грэнджеру	57
3.5. Тест Йохансена на коинтеграцию	58
Задания для самостоятельной работы к пунктам 3.3-3.5	62

Введение

Данное руководство представляет собой пособие по использованию эконометрического пакета EViews 4.1 применительно к анализу временных рядов. Руководство построено таким образом, что может быть использовано и при первом ознакомлении с данным пакетом в курсе эконометрики: многие его возможности традиционно применяются для оценки различных эконометрических моделей.

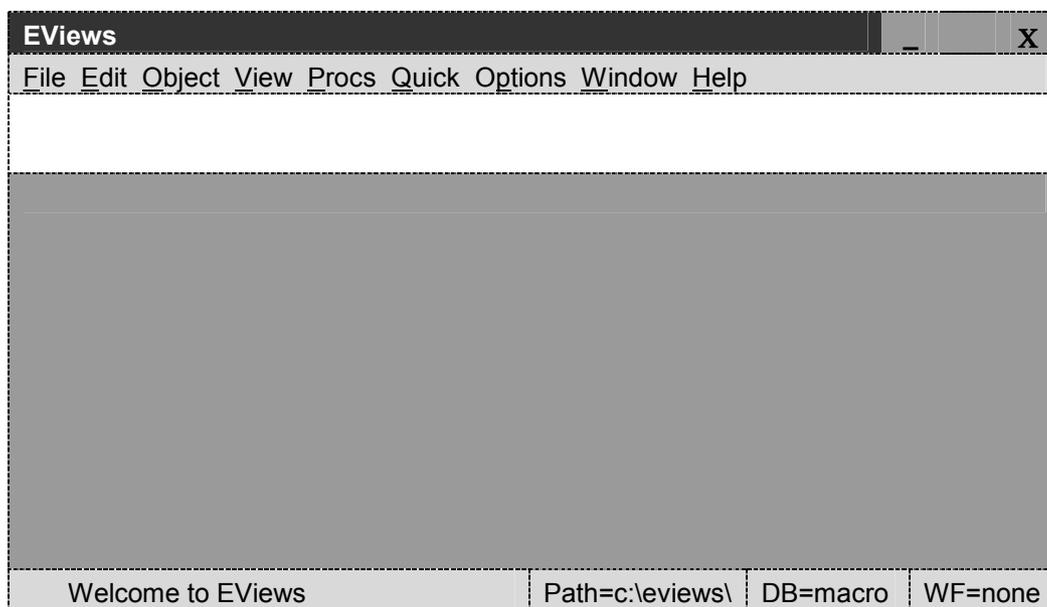
Руководство можно условно разделить на три части, касающиеся общих принципов работы в EViews (1), анализа одномерных временных рядов (2) и анализа многомерных временных рядов (3).

В первой части показано, как создается рабочий файл EViews и описана его общая структура. Во второй части рассмотрены основные принципы работы с одномерными временными рядами – построение графиков временных рядов, работа с их описательными статистиками, моделирование процессов типа $ARIMA(p, d, q)$, тесты на единичные корни и прогнозирование при помощи моделей временных рядов.

И, наконец, в последней части описаны общие принципы работы с многомерными временными рядами – создание многомерного временного ряда, построение его графика, оценка моделей векторной авторегрессии, проверка причинности по Грэнджеру и наличия коинтеграции.

1. Общие принципы работы в EViews

Окно EViews состоит из пяти основных зон:



- ✓ Титульная строка. Самая верхняя строка окна Eviews – когда Eviews активирован, она выделяется синим цветом.
- ✓ Главное меню – следующая строка окна Eviews – содержит 9 различных «выскакивающих» меню (**F**ile, **E**dit, **O**bject, **V**iew, **P**rocs,

Quick, Options, Window, Help), каждое из которых можно активировать, подведя курсор к названию конкретного меню и кликнув левой кнопкой мышки.

- ✓ **Командное окно.** Все команды EViews могут быть, с одной стороны активированы, как элементы главного меню, а с другой, могут быть набраны в командной строке. Для выполнения конкретной команды необходимо нажать клавишу **Enter** после ее ввода.
- ✓ **Рабочая область** отображает различные объекты (окна) EViews, открытые в текущий момент. К таким объектам относятся, например, рабочие файлы, программные файлы, конкретные временные ряды, оцененные уравнения и т.д.
- ✓ Область статусного состояния – показывает текущие характеристики Eviews.

1.1. Создание рабочего файла EViews

Шаг 1. Чтобы создать рабочий файл, содержащий данные, с которыми Вы собираетесь работать, необходимо в Главном меню выбрать опции **File/New/Workfile...**, в результате чего откроется диалоговое окно:

Frequency	
<input type="checkbox"/> Annual	<input type="checkbox"/> Weekly
<input type="checkbox"/> Semi-annual	<input type="checkbox"/> Daily [5 day weeks]
<input type="checkbox"/> Quarterly	<input type="checkbox"/> Daily [7 day weeks]
<input checked="" type="checkbox"/> Monthly	<input type="checkbox"/> Undated or irregular

Range	
Start date	End date
94:1	2003M6

В этом окне необходимо задать информацию о типе имеющихся у Вас данных: выбрать тип данных, с которыми Вы собираетесь работать (годовые, полугодовые, квартальные, месячные, недельные, дневные или нерегулярные (недатированные)) и определить соответствующий временной интервал.

- ✓ Годовые данные (Annual). Начальная (Start date) и конечная (End date) даты интересующего Вас интервала могут быть заданы двумя способами, если это даты типа 19**. Например, если у Вас есть данные с 1994 по 2002 г., то задать начальную дату интервала можно как 1994 либо как 94. В то же время конечную дату данного интервала необходимо задавать только с использованием всех четырех цифр, то есть как 2002.
- ✓ Квартальные данные (Quarterly): границы интервала могут быть заданы тремя способами для дат типа 19** и двумя для всех остальных. Например, 3-й квартал 1998 года можно задать как 1998:3 или как 98:3 или же как 1998Q3. Все остальные даты можно задавать, записывая год (4

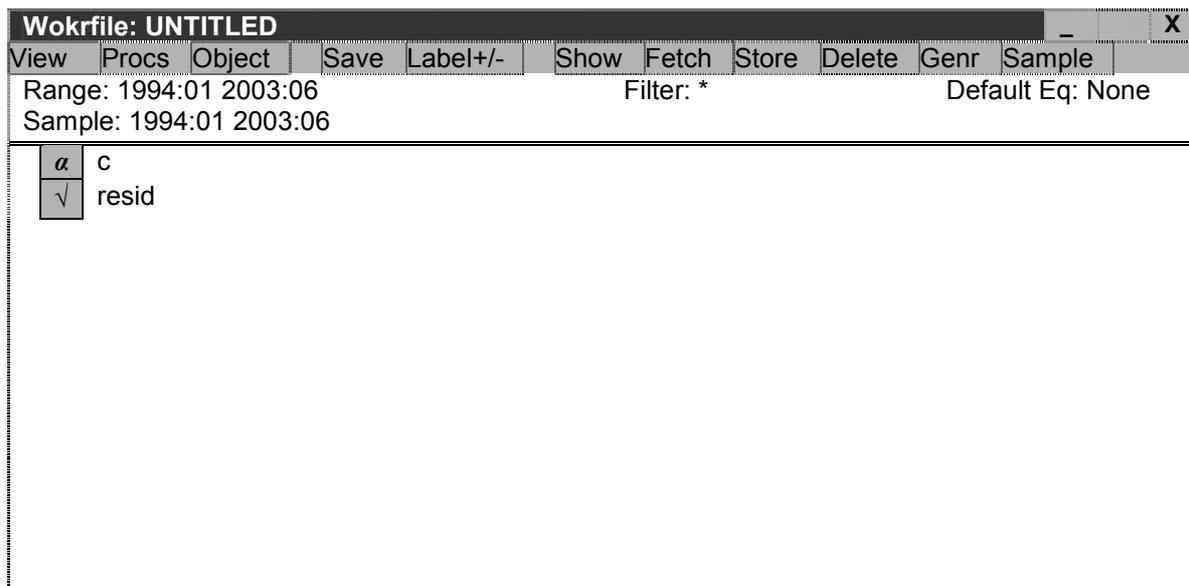
цифры), за которым через двоеточие или букву Q следует номер квартала (от 1 до 4).

- ✓ **Месячные данные (Monthly).** Задаются аналогично квартальным данным: за номером года через двоеточие или букву M следует номер месяца (от 1 (или 01) до 12). Например, август 1998 года может быть задан как 1998:08, 1998:8, 98:08 или 1998M8.
- ✓ **Полугодовые данные (Semi-annual):** как и в предыдущих случаях за номером года следует написать номер полугодия через двоеточие или букву S, например, 94:1 (первое полугодие 1994 года) или 2003S1 (первое полугодие 2003 года).
- ✓ **Недельные или дневные данные (Weekly или Daily).** В отличие от данных вышеперечисленных типов, в этом случае по умолчанию необходимо записать начальную (конечную) дату рассматриваемого интервала в формате `месяц:день:год (mm:dd:yyyy)`. Тогда если Вы ввели в качестве начальной даты `07:10:1974`, то в дневном формате такая запись означает, что Вы рассматриваете промежуток времени с 10 июля 1974 года, а в недельном, что неделя у Вас начинается со среды 10 июля 1974.
- ✓ **Нерегулярные или недатированные данные (Undated or irregular).** В этом случае просто определите номера наблюдений (первого и последнего), которые Вы будете использовать.

Таким образом, если в качестве начальной и конечной границ интервала Вы ввели числа 94 и 2002 соответственно, то в зависимости от типа данных это означает, что Вы рассматриваете следующие интервалы:

1994-2002	Для годовых данных
1994:1-2002:2	Для полугодовых данных
1994:1-2002:4	Для квартальных данных
1994:01-2002:12	Для месячных данных
01/01/1994-12/28/2002	Для недельных данных
01/01/1994-12/31/2002	Для дневных данных
94-2002	Для нерегулярных или недатированных данных

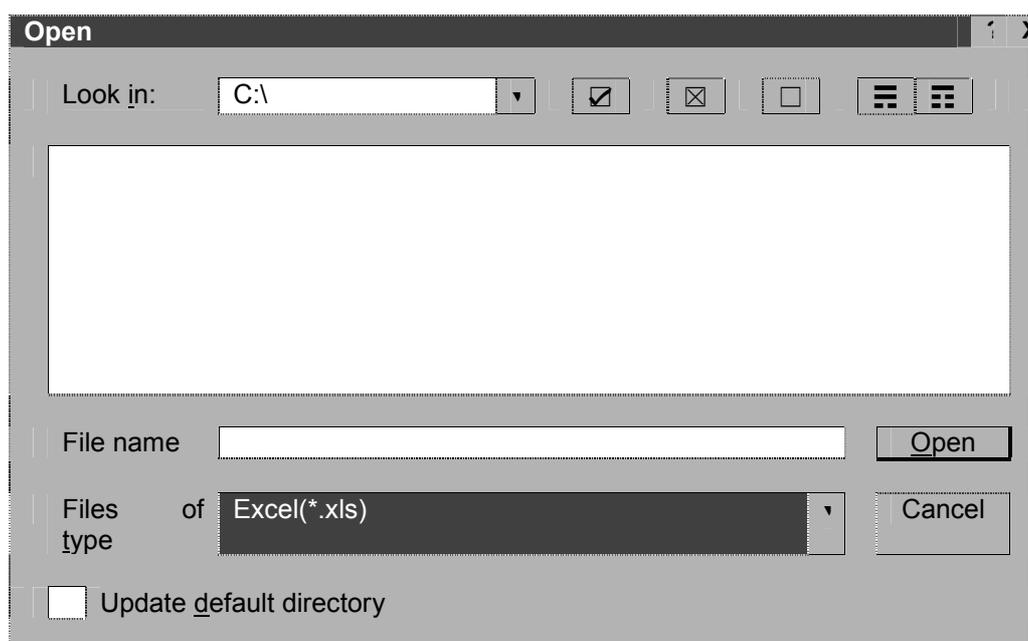
После того как Вы определили тип данных и промежуток времени, на котором Вы собираетесь работать, и нажали кнопку **OK**, EViews создаст рабочий файл (без названия) заданного Вами типа:



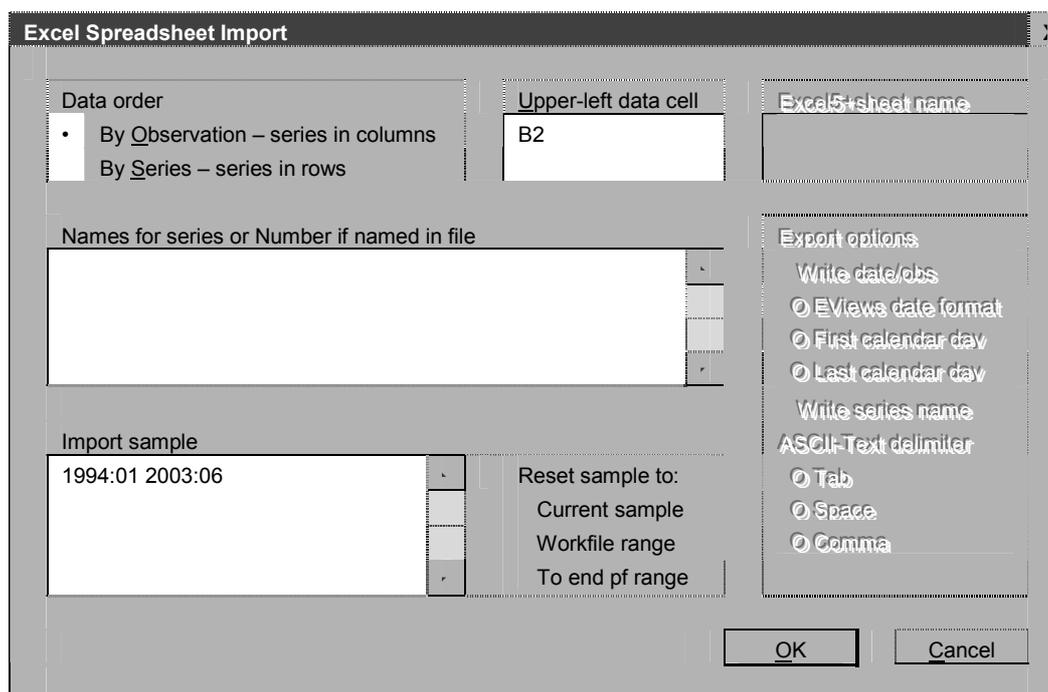
Отметим, что интервал «Range» отображает весь период, заданный при создании рабочего файла, в то время как интервал «Sample» показывает размер текущей выборки. В разделе 1.2 мы более подробно остановимся на описании структуры рабочего окна EViews.

Шаг 2. Следующим шагом при создании рабочего файла EViews является импорт данных в рабочий файл. В качестве файла, из которого импортируются данные, можно использовать любой файл формата Microsoft Excel с данными.

В Главном меню EViews выберите меню **File/Import/Read Text-Lotus-Excel...** После этого в открывшемся окне



выберете файл Microsoft Excel, из которого Вы будете импортировать данные, и откройте его. В результате откроется окно,



в котором по умолчанию отмечен порядок импортирования данных, верхняя левая ячейка листа Microsoft Excel, с которой начинается копирование данных в EViews, и границы интервала, которые Вы задали при создании рабочего файла. Прежде чем приступить к описанию основных опций окна импорта данных, остановимся на некоторых особенностях процедуры импорта данных, знание которых может оказаться полезным при работе с EViews. Во-первых, для большего удобства лучше сохранять файл с данными в формате «файл Microsoft Excel 4.0», то есть как лист, а не книгу Microsoft Excel. В этом случае не потребуется указывать название листа в опции Excel5+sheet name. Если же Вы сохранили файл с данными как книгу Microsoft Excel, то листу, с которого Вы будете импортировать данные, лучше присвоить имя, написанное латиницей. Кстати, было бы лучше, если название самого файла с данными тоже было написано латиницей.

Во-вторых, если Вы присвоите названия временным рядам, которые Вы собираетесь импортировать, непосредственно в файле Microsoft Excel (первая строка или первый столбец могут быть строкой или столбцом с названиями рядов), то это значительно упростит процедуру импорта данных, особенно, если Вы собираетесь копировать большие массивы с данными. Кроме того, удобно формировать лист Microsoft Excel только из тех временных рядов, которые Вы собираетесь импортировать в рабочий файл EViews либо импортировать подряд все ряды, имеющиеся на листе Microsoft Excel, а потом удалять ненужные ряды уже из рабочего файла EViews.

В-третьих, импортировать данные в рабочий файл EViews можно последовательно из различных файлов Microsoft Excel. И, наконец, файл Microsoft Excel, из которого Вы копируете данные, **должен** быть закрыт в момент осуществления процедуры импорта данных.

- ✓ Порядок импорта данных (Data order). Если в файле Microsoft Excel данные представлены в столбцах, то Вам необходимо выбрать опцию By Observation – series in columns (задается по умолчанию). В противном случае, если данные представлены в строках, выберите опцию By Series – series in rows.
- ✓ Ячейка, с которой начинается копирование данных – верхняя левая ячейка (Upper-left data cell). В этом окне необходимо указать название верхней левой ячейки, с которой будет начинаться импорт данных. По умолчанию указывается ячейка B2, что означает, что будет скопирована вся указанная Вами информация, расположенная ниже и правее этой ячейки, причем длина временных рядов будет соответствовать указанной.
- ✓ Название листа книги Microsoft Excel (Excel5+sheet name). В эту ячейку необходимо ввести название листа книги Microsoft Excel, с которого Вы импортируете данные, если файл с данными сохранен в формате «книга Microsoft Excel...». Если файл с данными является файлом Microsoft Excel 4.0, то данная ячейка не активирована.
- ✓ Названия (имена) временных рядов или количество импортируемых временных рядов, если им присвоены имена в файле Microsoft Excel (Names for series or Number if named in file). Если временным рядам не были присвоены имена непосредственно в файле Microsoft Excel, то в данном окне необходимо указать через пробел имена (латиницей) временных рядов, которые Вы импортируете: в этом случае будет скопировано столько временных рядов, сколько имен Вы напишите в данном окне. Если Вы уже присвоили имена временным рядам в файле Microsoft Excel, то достаточно указать количество импортируемых временных рядов, которые и будут скопированы.
- ✓ Размер выборки (Import sample). По умолчанию в данной ячейке указывается период, определенный на первом шаге, но Вы можете изменить (уменьшить) его, если это необходимо (в этом случае вместо значений наблюдений, не попавших в указанный подинтервал, будут оставлены пропуски, которые в EViews обозначаются NA).

После того как Вы заполните все данные поля окна импорта данных, и нажмете кнопку **ОК**, временные ряды будут скопированы в рабочий файл EViews, и окно рабочего файла примет, например, вид:

Wokrfile: UNTITLED		View	Procs	Object	Save	Label+/-	Show	Fetch	Store	Delete	Genr	Sample	
Range: 1994:01 2003:06		Filter: *					Default Eq: None						
Sample: 1994:01 2003:06													
α	c												
$\sqrt{}$	resid												
\surd	y1												
\surd	y2												
\surd	y3												

где y1, y2, y3– некоторые временные ряды.

Задания для самостоятельной работы

1. Загрузите пакет EViews 3.1 или EViews 4.1
2. Создайте рабочий файл EViews, состоящий из 1000 недатированных наблюдений

1.2. Общая структура рабочего файла EViews

Изменение имени и сохранение рабочего файла.

Поскольку эконометрический пакет EViews создан под оболочку Windows, ему присущи те же свойства, что и всем продуктам, совместимым с Windows, например, Microsoft Word, Excel и т.д. Поэтому на некоторых очевидных свойствах пакета EViews мы не будем останавливаться слишком подробно.

Существует два основных способа сохранения рабочего файла EViews. С одной стороны, Вы можете сохранить рабочий файл EViews, воспользовавшись опциями Главного меню **File/Save...** или **File/Save as...** С другой стороны, можно воспользоваться опцией меню рабочего файла Save (соответствующая «кнопка» в меню рабочего файла).

Отметим, что открыть уже существующий рабочий файл можно, воспользовавшись опцией **File/Open...** Главного меню EViews и указав путь к нужному файлу.

Изменение границ временного интервала.

В пакете EViews существуют специальные функции¹, позволяющие изменять (уменьшать) длину рассматриваемого интервала в процессе работы:

- ✓ @all – обозначает весь, используемый в рабочем файле интервал;
- ✓ @first – обозначает начальную дату интервала;
- ✓ @last – обозначает конечную дату интервала.

Существует три способа изменения границ временного интервала.

- ✓ При помощи специальной команды, набранной в командной строке. Для того чтобы изменить длину интервала, необходимо написать

```
smpl дата-1 дата-2
```

Пусть Вы создали файл с месячными данными, изменяющимися на интервале с января 1993 по июль 1998 г. Тогда если Вы написали в командной строке

```
smpl @all  
smpl @first 1996:12  
smpl 1995:08 @last,
```

то данные записи означают, соответственно, что Вы хотите использовать данные за весь имеющийся в Вашем распоряжении период, либо Вы собираетесь использовать данные с начала рассматриваемого периода до декабря 1996 г., либо, в последнем случае, Вы будете использовать данные на интервале с августа 1995 г. до конца исходного периода. Если Вы хотите исключить из рассмотрения какие-либо подпериоды в начале и в конце всего интервала времени, то необходимо в командной строке написать, например, следующее:

```
smpl 1993:09 1997:10.
```

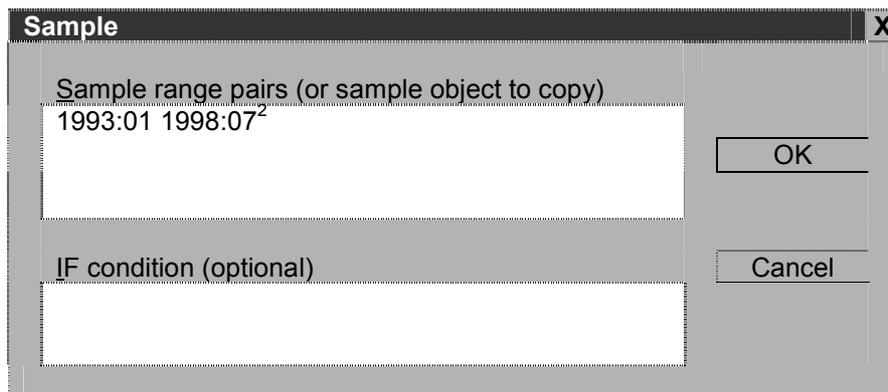
В этом случае Вы сократите исходный интервал до подинтервала с сентября 1993 по октябрь 1997 г.

Кроме того, в пакете EViews существует возможность использовать математические выражения при работе со встроенными функциями. Например, подинтервал с сентября 1993 по октябрь 1997 г. Вы можете также записать как

```
smpl @first+9 @last-10.
```

- ✓ При помощи специальной опции меню рабочего файла. Чтобы изменить границы временного интервала Вам необходимо выбрать опцию **Sample** в меню окна рабочего файла, после чего в появившемся окне вместо введенных при создании рабочего файла границ

¹ Отметим, что все функции, используемые в пакете EViews, обозначаются символом @.



набрать через пробел границы того подпериода, который Вы планируете рассматривать в дальнейшем.

- ✓ При помощи меню **Quick** Главного меню EViews. Чтобы изменить границы временного интервала Вам необходимо в меню **Quick** Главного меню EViews выбрать опцию **Sample**, после чего нужно действовать также как и в предыдущем случае.

Типы данных пакета EViews

Каждый объект рабочего файла EViews имеет свой специфический формат и свое обозначение («кнопку»), которое отображается в окне рабочего файла после создания объекта:

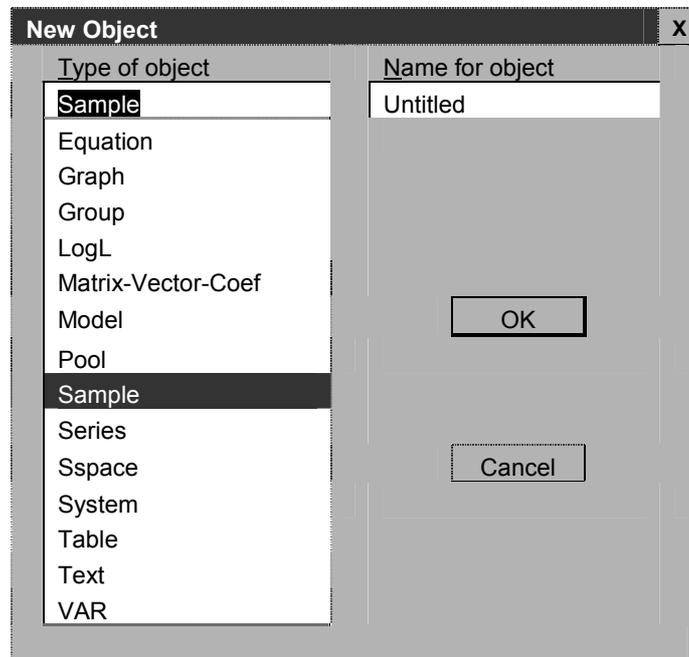
	– вектор коэффициентов (Coefficient Vector) – задается по умолчанию при создании рабочего файла
	– уравнение (Equation)
	– график (Graph)
	– группа (Group)
	– логарифм функции правдоподобия (LogL)
	– скаляр (число) (Scalar)
	– временной ряд (серия) (Series)
	– фазовое пространство (State space)
	– система (System)
	– симметричная матрица (SYM – Symmetric Matrix)
	– матрица (Matrix)
	– модель (Model)
	– панель данных (Pool)
	– размер выборки (Sample)
	– таблица (Table)
	– текст (Text)

² При появлении данного окна по умолчанию задаются текущие границы временного интервала, которые Вам и необходимо изменить.

– векторная авторегрессия (VAR – Vector Autoregression)

 – вектор (столбец или строка) (Vector/Row Vector)

Чтобы создать любой из перечисленных объектов, воспользуйтесь меню **Object/New Object...** Главного меню EViews



Вы можете выбрать объект нужного Вам типа, указать его имя и затем кликнуть ОК, после чего появится окно, соответствующее объекту того типа, который Вы выбрали, и в котором Вам нужно написать названия тех временных рядов (одного, нескольких или их арифметических выражений), для которых Вы строите данный объект, или другие необходимые данные.

Задания для самостоятельной работы

1. Сохраните, созданный в предыдущем разделе файл, присвоив ему имя DEMO.wf.
2. Создайте в этом файле несколько временных рядов с названиями y1, y2, y3, y4, y5.

2. Анализ одномерных временных рядов

2.1. Структура окна временного ряда

Окно каждого объекта рабочего файла EViews по своей структуре напоминает и окно рабочего файла, и окно EViews. Чтобы открыть окно любого объекта можно, например, дважды кликнуть левой кнопкой мышки на названии открываемого объекта.

Окно с данными³ любого временного ряда имеет следующий вид:

Series: y1 Workfile: CP-2														
View	Procs	Object	Print	Name	Freeze	Transform	Edit+/-	Smpl+/-	Label+/-	Wide+/-	InsDel	Title	Sample	Genr
Y1														
Last updated: 03/20/03 - 12:31														
Modified: 1 1000 // y1=-2														
Modified: 2 1000 // y1=0.5*y1(-1)+nrnd														
1				-2										
2				-0.9566483										
3				1.3842802										
4				1.0306127										
5				-0.3917438										
6				-1.6905064										
7														

- ✓ Меню View окна с данными временного ряда содержит следующие опции:
 - SpreadSheet** – позволяет вернуться в окно с данными временного ряда;
 - Graph** ► – строит график временного ряда одного из пяти типов (более подробно см. пункт 2.2);
 - Descriptive Statistics** ► – дает информацию об описательных статистиках временного ряда: математическом ожидании, дисперсии и т.д. (более подробно см. пункт 2.3);
 - Tests for Descriptive Statistics** ► – позволяет проводить простейшие тесты для описательных статистик (более подробно см. пункт 2.4);
 - Distribution** ► – отображает различные графики, характеризующие эмпирическое распределение временного ряда;
 - One-Way Tabulation...** – дает информацию о распределении данных на интервалах одинаковой длины, сгруппированных по возрастанию: отображает информацию о рассматриваемом подинтервале, количестве наблюдений временного ряда, значения которых попадают в этот подинтервал, и %, который составляют эти наблюдения от всей выборки, а также кумулятивные статистики (количество наблюдений и %);
 - Correlogram...** – отображает кореллограмму временного ряда (более подробно см. пункт 2.5);
 - Unit Root Test...** – позволяет исследовать временной ряд на наличие единичных корней (более подробно см. пункт 2.7);
 - BDS Independence Test...** – дает возможность проводить тестирование на наличие временной зависимости во временных рядах. Данный тест может быть применен как к самим временным, так и к остаткам оцененных

³ Окно временном ряда, в котором представлены результаты какой-либо обработки исходных данных имеет несколько иной вид (содержит иные опции меню), о чем будет подробнее сказано ниже.

моделей. Позволяет проверить наличие различных типов зависимостей: линейную, нелинейную и других;

Conversion Options... – при импорте данных из базы данных EViews позволяет автоматически конвертировать временные ряды согласно настройкам рабочего файла;

Label – выводит окно, содержащее информацию о временном ряде: название временного ряда, название таблицы с данными, последнее обновление данных, описание данных, источник, единицы измерения, замечания.

- ✓ Меню **Procs** окна с данными временного ряда позволяет:
 - сгенерировать новые временные ряды (**Generate by Equation** – аналог опции **Genr** меню окна временного ряда);
 - создать новый временной ряд (**Resample...**), случайным образом переставив значения текущего временного ряда;
 - провести сезонную корректировку временного ряда (**Seasonal Adjustment ►**) с использованием одного из предложенных методов: **Sensus X12...; X11 (Historical)...; Tramo/Seats...; Moving Average Methods...**;
 - построить прогнозы при помощи методов экспоненциального сглаживания (**Exponential Smoothing...**);
 - получить сглаженный временной ряд при помощи метода Ходрика-Прескотта (**Hodrick-Prescott Filter...**).
- ✓ Меню **Object** окна с данными временного ряда содержит следующие опции, многие из которых совпадают с опциями меню окна с данными временного ряда:
 - Store to DB...** – позволяет сохранить данные в базе данных EViews;
 - Update from DB...** – позволяет обновить имеющиеся данные, воспользовавшись информацией базы данных EViews;
 - Copy Object...** – создает безымянную копию временного ряда;
 - Name...** – позволяет изменить имя временного ряда (то же что и меню **Name** окна с данными временного ряда);
 - Delete** – удаляет временной ряд из рабочего файла EViews (то же что и меню **Delete** окна рабочего файла);
 - Freeze Output** – создает таблицу с данным временным рядом (то же что и меню **Freeze** окна с данными временного ряда);
 - Print** – позволяет распечатать таблицу с данными (то же что и меню **Print** окна с данными временного ряда);
 - View Options ►** – содержит ряд опций меню окна с данными временного ряда, описание которых приведено ниже:
 - Sample +/-; Edit +/-; Label +/-; Wide +/-; InsDel;**
 - Title...**
- ✓ Меню **Print** окна с данными временного ряда. Опция **Print** позволяет распечатать таблицу с данными.
- ✓ Меню **Name** окна с данными временного ряда. При помощи меню **Name** можно изменять имя временного ряда.
- ✓ Меню **Freeze** окна с данными временного ряда. Данная опция позволяет создать таблицу с данным временным рядом.

- ✓ Меню **Transform**⁴ окна с данными временного ряда. Позволяет преобразовывать исходные данные (**Level**) к следующему виду:

1 Period Difference – временной ряд представляется в первых разностях:

$$y_t - y_{t-1}$$

1 Year Difference – временной ряд представляется в виде сезонных разностей:

$$y_t - y_{t-12}$$

для месячных данных. Для квартальных и т.д. данных преобразования носят аналогичный характер.

1 Period % Change – временной ряд представляется в темпах прироста:

$$\frac{y_t - y_{t-1}}{y_t} \cdot 100\%$$

1 Period % Change (Annual Rate) – временной ряд представляется в виде темпа прироста в годовом исчислении:

$$\left(\left(\frac{y_t - y_{t-1}}{y_t} + 1 \right)^{12} - 1 \right) \cdot 100\%$$

1 Year % Change – временной ряд представляется в виде «сезонного» темпа прироста:

$$\frac{y_t - y_{t-12}}{y_t} \cdot 100\% ,$$

если Вы используете месячные данные. Для квартальных и т.д. данных преобразования носят аналогичный характер.

- ✓ Меню **Edit**+/- окна с данными временного ряда. Включает/выключает возможность непосредственного редактирования данных в окне с данными временного ряда. Изменив данные в ячейке, не забудьте нажать клавишу Enter.
- ✓ Меню **Smpl**+/- окна с данными временного ряда. В зависимости от того активирована данная возможность или нет, в окне с данными временного ряда показаны лишь данные на текущем интервале времени (подинтервале), либо на всем интервале.
- ✓ Меню **Label**+/- окна с данными временного ряда. Опция **Label**+/- выводит на экран (либо убирает с экрана) информацию о времени создания временного ряда, времени его изменения и т.д.
- ✓ Меню **Wlde**+/- окна с данными временного ряда. Опция **Wlde**+/- позволяет представить временной ряд более компактно (в виде матрицы).
- ✓ Меню **InsDel** окна с данными временного ряда. При помощи данной опции можно добавить/удалить одно наблюдение. Если Вы хотите добавить наблюдение к временному ряду, в появившемся окне выберите опцию Insert obs and move subsequent obs down one. Помните, что в этом случае все наблюдения, следующие за наблюдением (Вы также должны его выбрать), на место которого Вы вставляете новое, сместятся на одно наблюдение вниз.

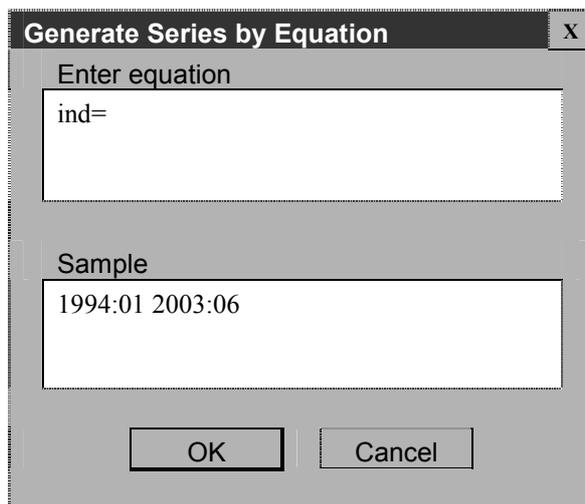
⁴ Данной опции нет в EViews 3.1.

Размер выборки при этом не меняется, и последнее наблюдение может быть потеряно.

Если Вы хотите удалить какое-либо наблюдение, выберите опцию Delete obs and move subsequent obs up one. В этом случае все следующие за удаляемым наблюдения сместятся на одну позицию вверх.

- ✓ Меню Title окна с данными временного ряда. Выбрав эту опцию меню окна с данными временного ряда можно изменить название таблицы (не имя временного ряда) с данными временного ряда: в появившемся окне наберите нужное вам название.
- ✓ Меню Sample окна с данными временного ряда. Данная опция позволяет менять размер текущей выборки (более подробно см. выше – п.1.2).
- ✓ Меню Genr окна с данными временного ряда. При помощи данной опции можно сгенерировать новые временные как с использованием уже существующих временных рядов, так и с использованием генератора случайных чисел EViews. Выбрав данную опцию, введите в появившемся окне выражение для вычисления нового временного ряда или для перекодирования старого:

✓



В окне Sample измените границы временного интервала, если это необходимо.

Задания для самостоятельной работы

1. В файле DEMO.wf сгенерируйте следующие случайные процессы:

$$\begin{aligned}wn &= \text{nrnd}^* \\ y1 &= a0 + a1 * y1(-1) + wn \\ y2 &= a0 + a2 * y2(-1) + wn,\end{aligned}$$

где a_0 , a_1 , a_2 – некоторые константы, которые Вы можете выбрать самостоятельно так, чтобы сгенерированные случайные процессы оказались слабо стационарными.

* `nrnd` – встроенная в EViews функция, позволяющая генерировать случайный процесс, имеющий стандартное нормальное распределение

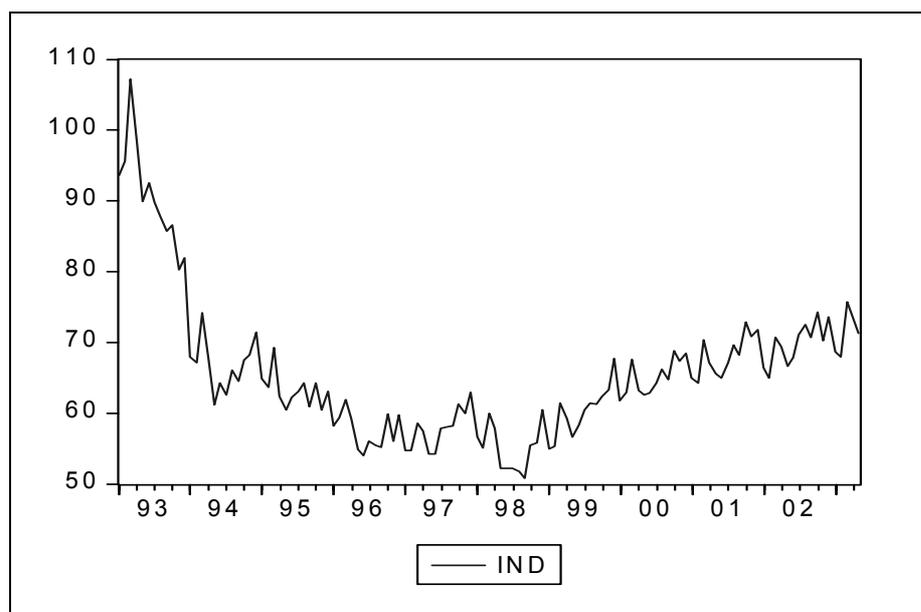
Запись $y1(-1)$ означает, что Вы включаете в число объясняющих переменных первое запаздывание временного ряда. Помните, что прежде чем сгенерировать временные ряды $y1$ и $y2$, во-первых, необходимо задать начальные условия (это можно сделать, воспользовавшись, например, меню **Edit** окна временного ряда). Во-вторых, после задания спецификации случайного процесса в окне **Generate Series by Equation** не забудьте поменять размер выборки на 2 1000.

2. Откройте окна с созданными временными рядами.

2.2 Построение графика временного ряда

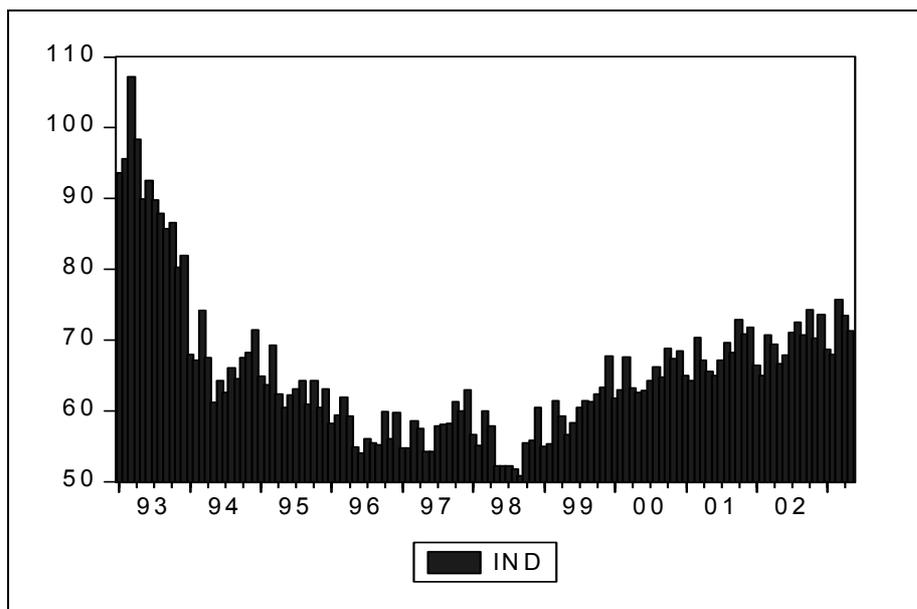
Для построения графика временного ряда, в первую очередь, откройте окно данного временного ряда. Затем в меню **View** окна временного ряда выберите опцию **Graph**. В открывшемся меню выберите нужный Вам тип графика:

- ✓ Линейный график⁵ (**Line**) – отображает обычный линейный график временного ряда. По оси абсцисс отложено время (или номер наблюдения), по оси ординат – значения временного ряда.

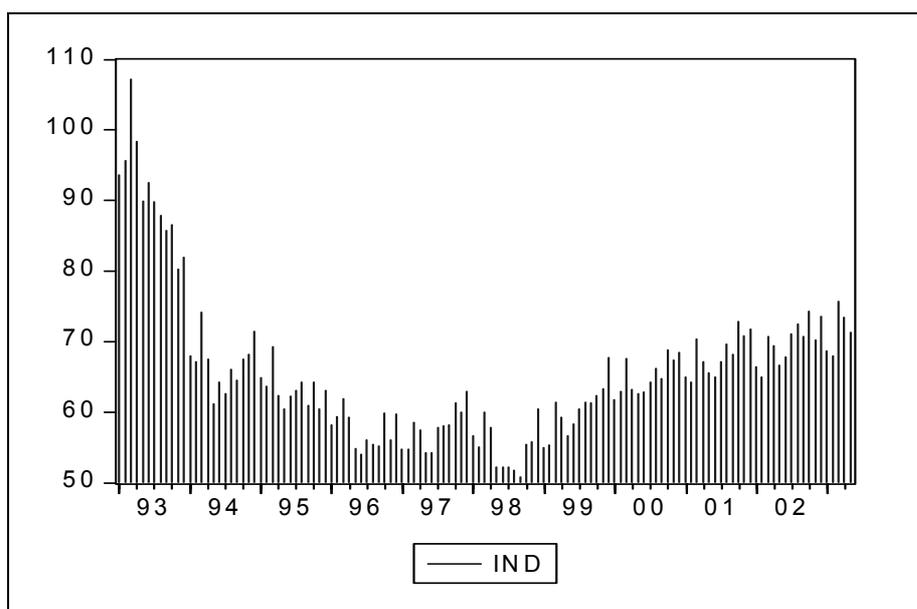


- ✓ Столбчатая диаграмма (**Bar**) – отображает столбчатую диаграмму временного ряда. Является наиболее полезным, если временной ряд содержит небольшое количество наблюдений.

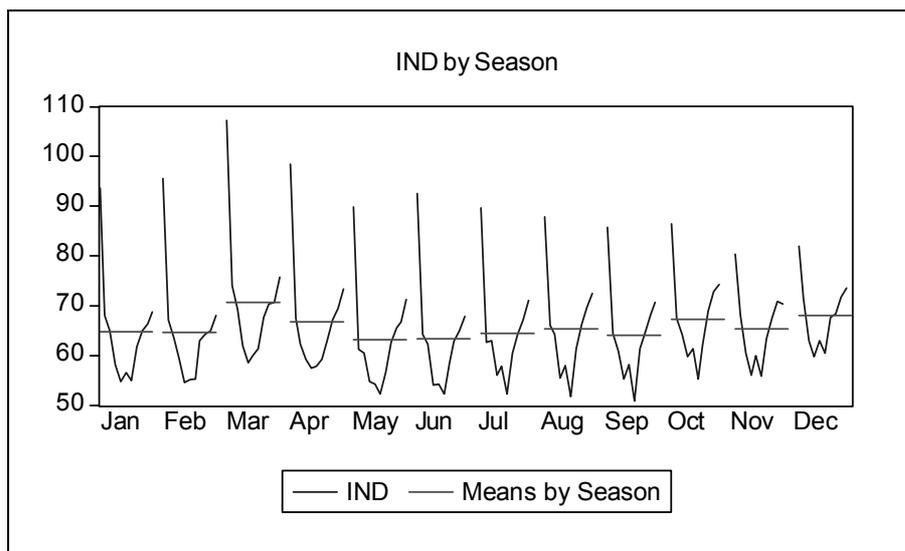
⁵ В качестве примера рассмотрены различные типы графиков индекса промышленного производства Центра экономической конъюнктуры при Правительстве РФ по промышленности в целом (IND) за период с января 1993 г. по май 2003 (скорректированное значение января 1993 г. = 100)



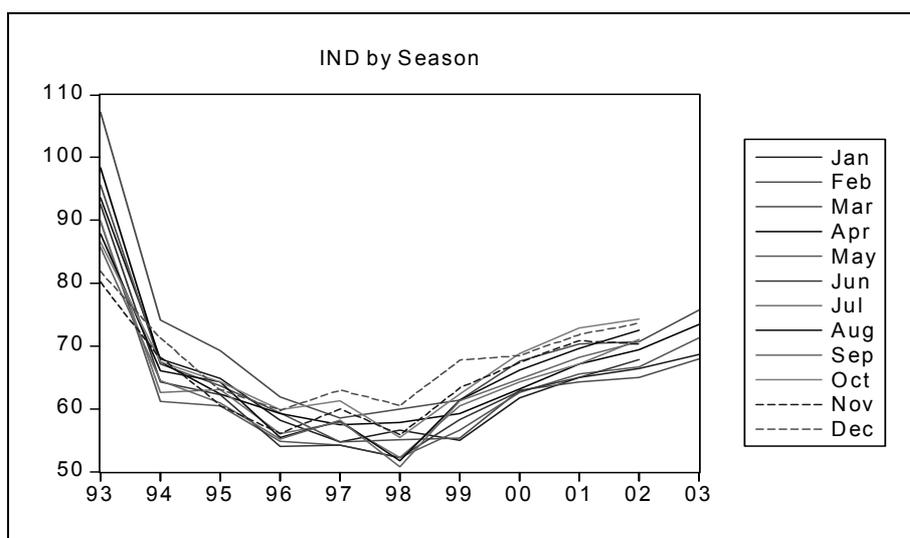
- ✓ «Игольчатый» график (Spike) – отображает каждое значение временного ряда в виде вертикальной линии, длина которой равна этому значению (в выбранном масштабе).



- ✓ Сезонно упорядоченный график (Seasonal Stacked Line) – отображает график временного ряда для каждого квартала или месяца в отдельности – значения временного ряда сгруппированы по номерам квартала или месяца и упорядочены по году. Графики следуют друг за другом: сначала график первого квартала или месяца, затем второго, и т.д. На графиках также представлены средние значения временного ряда для каждого сезона. Заметьте, что данная опция доступна лишь для квартальных либо месячных данных.



- ✓ Сезонно упорядоченный график (Seasonal Split Line) – отображает график временного ряда для каждого квартала или месяца в отдельности – значения временного ряда сгруппированы по номерам квартала или месяца и упорядочены по году. Графики располагаются на одной годовой оси. Данная опция также доступна только для квартальных либо месячных данных.



Если Вы хотите сохранить график как отдельный объект, выберите опцию **Freeze** в меню окна временного ряда – появится окно графика данного временного ряда. Изменив название данного объекта при помощи опции **Name** окна графика временного ряда, Вы сохраните его (график временного ряда) в окне рабочего файла EViews.

Для построения графика временного ряда можно также воспользоваться меню **Object/New Object/Graph...** Главного меню EViews. В появившемся окне необходимо набрать название временного ряда, график которого Вы хотите построить. Отметим, что в этом случае Вы можете построить лишь линейный график, создав при этом новый объект (график) EViews.

Воспользовавшись командной строкой и командой `graph`, Вы также можете создать новый объект – график. Для этого в командной строке необходимо записать (латиницей) строку

`graph <название графика>.<тип графика (опции)> <название временного ряда (временных рядов)>`

и нажать `Enter`.

В EViews доступные следующие типы графиков и опции:

bar	Столбчатая диаграмма
errbar	Для двух временных рядов отображает вертикальные отрезки, соединяющие одномоментные значения первого и второго временных рядов. При этом предполагается, что значение первого ряда превышает соответствующее значение второго. В случае, когда это условие не выполняется, на графике пунктирной линией отображаются лишь значения самих временных рядов, не соединенные вертикальным отрезком.
hilo	На этом графике можно отобразить до 4-х временных рядов. Первые два временных ряда изображаются аналогично случаю опции <code>errbar</code> . Третий и четвертый – отображаются пунктирными линиями.
line	Линейный график (будет построен по умолчанию в случае, если Вы не зададите тип графика в командной строке)
pie	Круговая диаграмма (опция доступна лишь для многомерных временных рядов и положительных значений этих многомерных временных рядов)
scat	Аналогично опции <code>xy</code> , только точечная диаграмма (диаграмма рассеяния)
spike	«Игольчатый» график
xy	График в координатах (x, y) . По оси абсцисс (Ox) откладываются значения первого из перечисленных в командной строке временных рядов, по оси ординат – всех остальных. Таким образом, отображается несколько графиков $y(x)$
xyline	Аналогично опции <code>xy</code>
xupair	Аналогично опции <code>xy</code>

Задания для самостоятельной работы

1. Постройте линейные графики случайных процессов w_n , y_1 , y_2 , которые были сгенерированы в пункте 2.2. Сравните графические свойства этих случайных процессов

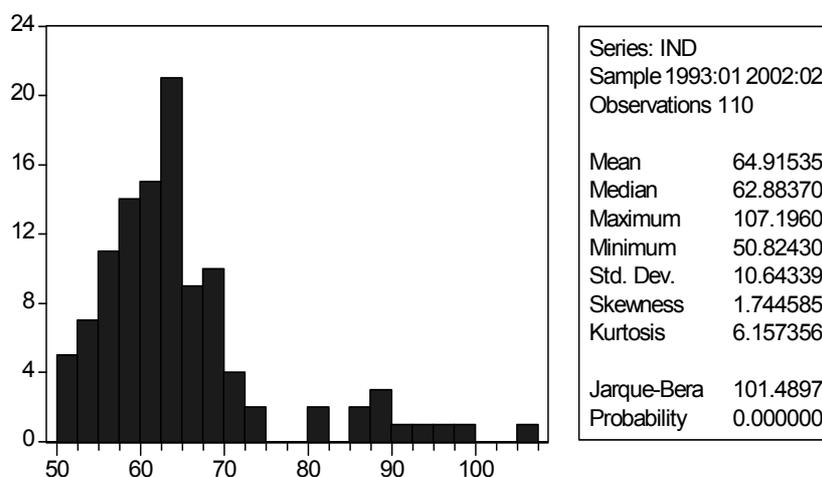
2.3. Описательные статистики временного ряда

Воспользовавшись опцией **Descriptive Statistics** меню **View** окна временного ряда, Вы можете получить простейшие описательные статистики временного ряда. Для этого в появившемся окне

Histogram and Stats
Stats Table
Stats by Classification...

выберите необходимую опцию.

✓ Гистограмма и статистики (Histogram and Stats) – отображает гистограмму временного ряда с информацией о нем (название ряда, размер выборки, количество наблюдений)



и соответствующие выборочные статистики: математическое ожидание (Mean), медиану (Median), максимальное значение (Maximum), минимальное значение (Minimum), стандартное отклонение (Std. Dev.), коэффициенты асимметрии (Skewness) и эксцесса (Kurtosis), а также значение статистики Харке-Бера (Jarque-Bera) с соответствующим P -значением (Probability).

Выборочное математическое ожидание $\hat{\mu}$ рассчитывается как сумма всех наблюдений y_t , деленная на количество наблюдений T . Медиана – среднее наблюдение упорядоченного по возрастанию (или убыванию) временного ряда, если количество наблюдений не четно, и среднее арифметическое двух центральных наблюдений упорядоченного по возрастанию/убыванию временного ряда, если он содержит четное число наблюдений. Стандартное отклонение и коэффициенты асимметрии и эксцесса рассчитываются по следующим формулам:

стандартное отклонение:

$$\hat{\sigma} = \sqrt{\frac{\sum_{t=1}^T (y_t - \hat{\mu})^2}{T - 1}} -$$

несмещенная оценка стандартного отклонения;

коэффициент асимметрии:

$$S = \frac{1}{T} \sum \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}'} \right)^3,$$

где

$$\hat{\sigma}' = \hat{\sigma} \sqrt{\frac{T - 1}{T}} -$$

оценка стандартного отклонения, рассчитанная на основе смещенной оценки дисперсии. Коэффициент асимметрии симметричного распределения, например, нормального, равен нулю. Положительное значение

коэффициента асимметрии означает, что распределение имеет длинный правый хвост, а отрицательное – что длинный левый;

коэффициент эксцесса:

$$K = \frac{1}{T} \sum_{t=1}^T \left(\frac{y_t - \mu}{\sigma'} \right)^4.$$

Коэффициент эксцесса нормального распределения равен 3. Коэффициент эксцесса островершинного распределения превышает 3, а плосковершинного – меньше 3.

И, наконец, тест Харке-Бера позволяет проверить гипотезу о том, что временной ряд нормально распределен. Статистика Харке-Бера вычисляется, исходя из свойств нормального распределения, на основе коэффициентов асимметрии и эксцесса:

$$JB = \frac{T-k}{6} \left(S^2 + \frac{(K-3)^2}{4} \right),$$

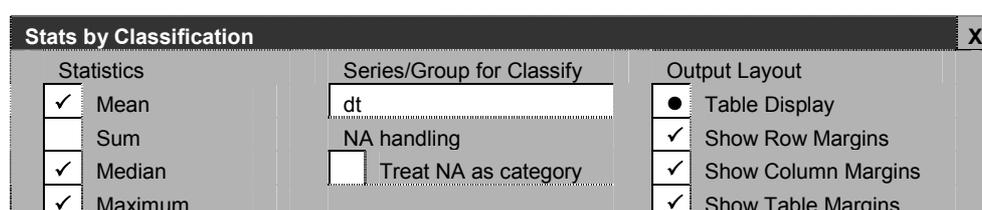
где S – значение коэффициента асимметрии, K – значение коэффициента эксцесса, k – количество оцениваемых в модели параметров (в данном случае равно нулю, т.к. никакая модель не оценивается). Статистика Харке-Бера распределена как $\chi^2(2)$.

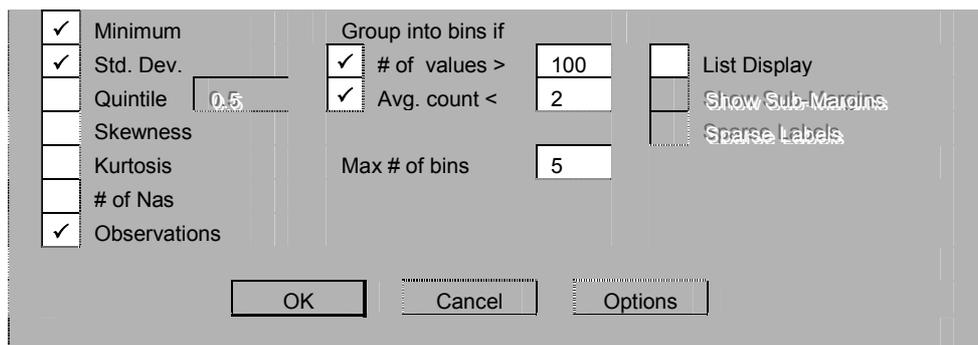
✓ Статистическая таблица (Stats Table). Представляет в форме таблицы информацию о временном ряде (название ряда и количество наблюдений) и соответствующие выборочные статистики: математическое ожидание (Mean), медиану (Median), максимальное значение (Maximum), минимальное значение (Minimum), стандартное отклонение (Std. Dev.), коэффициенты асимметрии (Skewness) и эксцесса (Kurtosis), значение статистики Харке-Бера (Jarke-Bera) с соответствующим P -значением (Probability), а также сумму всех наблюдений (Sum) и сумму квадратов отклонений (Sum Sq. Dev.).

✓ Описательные статистики временного ряда по специально выделенным подгруппам (Stats by Classification...). Данная опция позволяет вычислять описательные статистики временного ряда для его различных подгрупп. Например, допустим, Вы хотите сравнить статистические характеристики (математическое ожидание, медиану, максимальное и минимальное значения, стандартное отклонение) временного ряда базисного индекса объема промышленного производства на двух подинтервалах с января 1995 г. по август 1998 г. и с сентября 1998 г. по май 2003 г. Для этого можно создать фиктивную переменную

$$DT_t = \begin{cases} 0, & \text{если } t \leq 1998:08, \\ 1, & \text{в противном случае.} \end{cases}$$

Затем выберите меню **View/Descriptive Statistics/Stats by Classification...** и в появившемся окне выделите необходимые Вам статистики:





После того как Вы выбрали все необходимое, нажмите клавишу ОК – в окне временного ряда появится таблица с требуемой информацией:

Descriptive Statistics for IND						
Categorized by values of DT						
Date: 08/08/03 Time: 13:04						
Sample: 1995:01 2003:05						
Included observations: 101						
DT	Mean	Median	Max	Min.	Std. Dev.	Obs.
0	58.59149	58.39150	69.26350	51.70410	4.019442	44
1	65.57914	66.40950	75.67090	50.83030	5.516713	57
All	62.53502	62.33490	75.67090	50.83030	6.009558	101

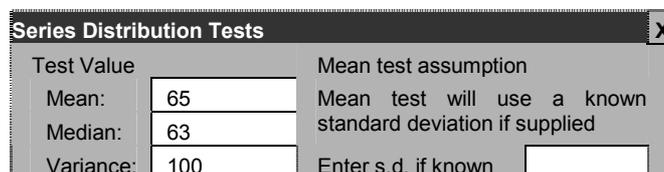
Очевидно, что можно проводить разбиение на подгруппы по нескольким признакам, если ввести в окошке Series/Group for Classify через пробел несколько временных рядов, по которым Вы хотите провести разбиение.

Задания для самостоятельной работы

1. Постройте гистограммы случайных процессов w_t , y_1 , y_2 . Можно ли сказать, данные случайные процессы имеют нормальное распределение?
2. Выведите статистические таблицы для каждого из перечисленных выше временных рядов. Совпадают ли их средние значения и стандартные отклонения с теоретическими?

2.4. Тесты для описательных статистик временного ряда

Эконометрический пакет Eviews дает возможность проводить простейшие тесты для описательных статистик временных рядов. Выбрав меню **View/Tests for Descriptive Stats/Simple Hypothesis Tests**, Вы можете проверить простейшие гипотезы о равенстве математического ожидания, медианы и дисперсии временного ряда конкретным значениям, введя интересующие Вас значения в появившееся окно:





Допустим, Вы хотите проверить гипотезы о том, что математическое ожидание базисного индекса объема промышленного производства (база – январь 1993 г.) на интервале с января 1993 г. по май 2003 г. равно 65, медиана – 63, а дисперсия – 81. Тогда:

Математическое ожидание. Для математического ожидания Вы проверяете гипотезу о том, что оно равно $m=65$, против соответствующей альтернативной двусторонней гипотезы:

$$H_0: \mu = m$$

$$H_A: \mu \neq m$$

Если Вы не задаете значение стандартного отклонения y_t , то в качестве тестовой статистики вычисляется t -статистика:

$$t = \frac{\hat{\mu} - m}{\hat{\sigma} / \sqrt{T}},$$

где $\hat{\mu}$ – выборочное среднее y_t , $\hat{\sigma}$ – несмещенное выборочное стандартное отклонение. Если y_t нормально распределен, то вычисленная t -статистика соответствует распределению Стьюдента с $T-1$ степенью свободы.

Если Вы специфицировали значение стандартного отклонения σ , EViews также вычисляет и z -статистику:

$$z = \frac{\hat{\mu} - m}{\sigma / \sqrt{T}}$$

где $\hat{\mu}$ – выборочное среднее y_t , $\hat{\sigma}$ – несмещенное выборочное стандартное отклонение, σ – заданное Вами значение стандартного отклонения. Если y_t нормально распределен, то вычисленная z -статистика соответствует стандартному нормальному распределению.

Значение вероятности для соответствующей статистики интерпретируется стандартным образом: если это значение меньше некоторого уровня значимости, например, 0,05, то можно говорить о том, что на 5%-м уровне значимости гипотеза о равенстве математического ожидания значению m отвергается; если значение полученной вероятности превышает конкретный уровень значимости, то можно говорить о том, что математическое ожидание значимо не отличается от заданного значения.

Медиана. Для медианы Вы проверяете гипотезу о том, что ее значение равно $m=63$, против соответствующей альтернативной двусторонней гипотезы:

$$H_0: med(y_t) = m$$

$$H_A: med(y_t) \neq m$$

EViews вычисляет несколько статистик, позволяющих проверить данную нулевую гипотезу.

Binomial sign test. Данный тест основан на идее о том, что если выборка является случайной подвыборкой биномиального распределения, то половина наблюдений выборки должна располагаться выше медианного значения, а вторая половина – ниже него. EViews представляет двусторонние P -значения как для биномиального теста, так и для его нормального приближения (для случая наличия большого количества наблюдений с соответствующей коррекцией на непрерывность).

Wilcoxon signed ranks test. Предположим, что Вы вычислили абсолютное значение разности между каждым наблюдением и выборочным математическим ожиданием временного ряда, а затем упорядочили все наблюдения по убыванию (или возрастанию). Тогда тест Вилкоксона основан на идее о том, что сумма рангов, превышающих медиану, и сумма рангов, меньших медианы, должны быть одинаковыми. EViews вычисляет P -значение для асимптотического нормального приближения t -статистики Вилкоксона (скорректированное на непрерывность и ties) (correcting for both continuity and ties).

Van der Waerden (normal scores) test. Данный тест основан на той же идее, что и тест Вилкоксона, но базируется на сглаженных рангах. smoothed ranks.

Отметим, что результаты этих тестов могут противоречить друг другу, как видно из приведенного ниже примера.

Дисперсия. Для дисперсии проверяется гипотеза о том, что ее значение равно $\sigma=81$, против соответствующей альтернативной двусторонней гипотезы

$$H_0: \text{var}(y_t) = \sigma^2$$

$$H_A: \text{var}(y_t) \neq \sigma^2$$

EViews вычисляет следующую статистику:

$$\chi^2 = \frac{(T-1)\hat{\sigma}^2}{\sigma^2},$$

где T – число наблюдений, $\hat{\sigma}$ – выборочное стандартное отклонение. Если y_t нормально распределен, статистика соответствует распределению $\chi^2(T-1)$. В данном случае, если значение соответствующей вероятности больше некоторого уровня значимости, то на этом уровне значимости мы не можем отвергнуть нулевую гипотезу.

В нашем случае в результате проверки описанных выше гипотез получились следующие результаты:

Hypothesis Testing for IND		
Date: 08/08/03 Time: 17:18		
Sample: 1993:01 2003:05		
Included observations: 125		
Test of Hypothesis: Mean = 65.00000		

Sample Mean = 65.66241		
Sample Std. Dev. = 10.21339		

<u>Method</u>	<u>Value</u>	<u>Probability</u>
t-statistic	0.725120	0.4697

Test of Hypothesis: Variance = 81.000000		

Sample Variance = 104.3133		

<u>Method</u>	<u>Value</u>	<u>Probability</u>
Variance Ratio	159.6894	0.0170

Test of Hypothesis: Median = 63.00000

Sample Median = 64.22410

Method	Value	Probability
Sign (exact binomial)	68	0.3712
Sign (normal approximation)	0.894427	0.3711
Wilcoxon signed rank	1.749391	0.0802
van der Waerden (normal scores)	2.172257	0.0298

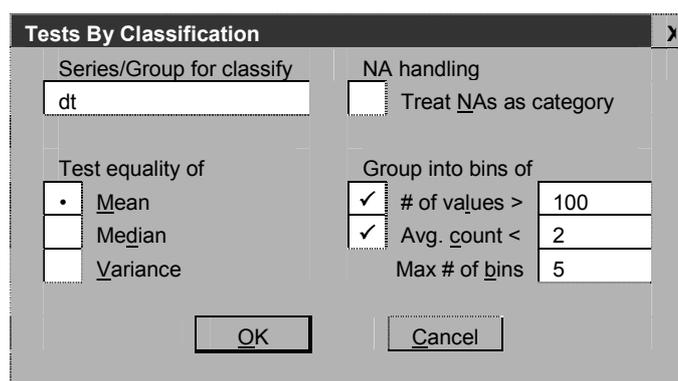
Median Test Summary

Category	Count	Mean Rank
Obs > 63.00000	68	68.3529412
Obs < 63.00000	57	56.6140351
Obs = 63.00000	0	
Total	125	

Таким образом, на 5%-м уровне значимости мы не можем отвергнуть гипотезу о том, что среднее значение базисного индекса объемов промышленного производства за период с января 1993 г. по май 2003 г. равно 65 не отвергается; дисперсия равна 81 – отвергается; а в случае с медианой – результаты противоречат друг другу.

Далее, выбрав меню **View/Tests for Descriptive Stats/Equality Tests by Classification...**, Вы можете проверить гипотезы о равенстве математических ожиданий, медиан и дисперсий различных подвыборок имеющейся у Вас выборки.

Допустим, Вы хотите проверить гипотезы о равенстве математических ожиданий, медиан и дисперсий базисного индекса объемов промышленного производства на двух подпериодах: с января 1995 г. по август 1998 года и с сентября 1998 г. по май 2003 г. Выбрав соответствующие опции меню, в появившемся окне укажите название(я) временного(ых) ряда(ов) (Series/Group for classify), а также отметьте равенство каких статистических характеристик временного ряда Вы собираетесь проверить: математического ожидания (Mean), медианы (Median) или дисперсии (Variance) –



Тест на равенство математических ожиданий. Данный тест основан на подходе, связанным с анализом дисперсий (ANOVA) различных подвыборок имеющейся выборки: считается, что если математические ожидания подгрупп совпадают, то сумма квадратов разностей между математическим ожиданием любой подгруппы и математическим ожиданием, рассчитанным по всей выборке,

должна совпадать с суммой квадратов разностей между математическим ожиданием всей выборки и каждым наблюдением.

Обозначим i -е наблюдение в группе g как $y_{g,i}$, где $i=1, \dots, n_g$ для групп $g=1, 2, \dots, G$. Тогда межгрупповые (Between) и внутригрупповые (Within) суммы квадратов равны:

$$SS_B = \sum_{g=1}^G n_g (\hat{\mu}_g - \hat{\mu})^2$$

$$SS_W = \sum_{g=1}^G \sum_{i=1}^{n_g} (y_{g,i} - \hat{\mu}_g)^2,$$

где $\hat{\mu}$ – выборочное математическое ожидание, рассчитанное для выборки в целом, а $\hat{\mu}_g$ – выборочное математическое ожидание, рассчитанное для подгруппы $g=1, \dots, G$. F -статистика для проверки гипотезы о равенстве математических ожиданий вычисляется по формуле:

$$F = \frac{SS_B / (G - 1)}{SS_W / (T - G)},$$

где T – общее количество наблюдений. F -статистика имеет F -распределение со степенями свободы $(G-1, T-G)$ в предположении, что процесс одинаково, независимо и нормально распределено.

Test for Equality of Means of IND				
Categorized by values of DT				
Date: 08/11/03 Time: 13:33				
Sample: 1995:01 2003:05				
Included observations: 101				
Method	df	Value	Probability	
t-test	99	7.073519	0.0000	
Anova F-statistic	(1, 99)	50.03467	0.0000	
Analysis of Variance				
Source of Variation	df	Sum of Sq.	Mean Sq.	
Between	1	1212.464	1212.464	
Within	99	2399.015	24.23248	
Total	100	3611.479	36.11479	
Category Statistics				
DT	Count	Mean	Std. Dev.	Std. Err. of Mean
0	44	58.59149	4.019442	0.605954
1	57	65.57914	5.516713	0.730707
All	101	62.53502	6.009558	0.597973

Если $G=2$, то EViews вычисляет t -статистику, которая представляет собой квадратный корень из F -статистики с $T-G$ степенями свободы.

В итоговой таблице также приведены результаты анализа дисперсий: общая сумма квадратов описанных выше разностей разложена на межгрупповую (Between) и внутригрупповую (Within) составляющие –

$$Between = \frac{SS_B}{df_B}$$

$$Within = \frac{SS_W}{df_W},$$

где df_B и df_W – соответствующее число степеней свободы. Тогда F -статистика – это отношение

$$F = \frac{Between}{Within}.$$

Тесты на равенство медиан. EViews вычисляет несколько различных тестов, позволяющих проверить гипотезу о том, что подгруппы имеют одинаковое распределение, против альтернативной о том, что по крайней мере одна из подгрупп имеет распределение, отличающееся от остальных. В случае наличия двух подгрупп нулевая гипотеза состоит в том, что обе подвыборки являются независимыми выборками одного и того же общего распределения.

Необходимо отметить, во избежание ошибок, что в данном случае проверяется равенство не самих медиан, а различных статистик, вычисленных для подгрупп.

Wilcoxon/Mann-Whitney signed ranks test. Данный тест проводится для случая, когда выборка разбита по некоторому признаку на две подвыборки, и рассчитывается аналогично тесту Вилкоксона для проверки гипотезы о равенстве медианы какому-то конкретному заданному значению (см. раздел 2.3).

Chi-square test for the median. Данный тест относится к классу ANOVA тестов и основан на сравнении количества наблюдений, расположенных выше и ниже медианы, рассчитанной для всей выборки. Иногда этот тест называют «медианным тестом» («the median test»). Медианная χ^2 -статистика асимптотически распределена как $\chi^2(G-1)$. EViews также вычисляет статистику Ята (Yate), скорректированную на непрерывный случай, хотя использование данной корректировки является довольно спорной.

Kruskal-Wallis one-way ANOVA by ranks. Данный тест является обобщением теста Манна-Уитни (Mann-Whitney test) для случая наличия более чем двух подгрупп. Основная идея состоит в том, что временной ряд ранжируется по возрастанию (наименьшему значению присваивается ранг 1, наибольшему – ранг T). После этого сравниваются суммы рангов подгрупп 1 и 2. Если медианы этих подгрупп совпадают, то значения данных сумм должны быть одинаковыми.

Test for Equality of Medians of IND			
Categorized by values of DT			
Date: 08/12/03 Time: 13:02			
Sample: 1995:01 2003:05			
Included observations: 101			
Method	df	Value	Probability
Wilcoxon/Mann-Whitney		5.948351	0.0000
Wilcoxon/Mann-Whitney (tie-adj.)		5.948351	0.0000
Med. Chi-square	1	30.60075	0.0000
Adj. Med. Chi-square	1	28.42072	0.0000
Kruskal-Wallis	1	35.42363	0.0000

Kruskal-Wallis (tie-adj.)	1	35.42363	0.0000
van der Waerden	1	31.91995	0.0000
<hr/>			
Category Statistics			
DT	Count	Median	> Overall Median Mean Rank Mean Score
0	44	58.39150	8 31.25000 -0.617909
1	57	66.40950	42 66.24561 0.476983
All	101	62.33490	50 51.00000 -9.51E-17

EViews вычисляет асимптотическое нормальное приближение U -статистики (с корректировкой на непрерывность и ...tie) и P -значения для двустороннего теста. Тест основан на односторонний анализ дисперсии, использующий только ранги данных. EViews также вычисляет χ^2 -приближение теста Крускала-Уоллеса (с tie корректировкой). Данная статистика приближенно распределена как $\chi^2(G-1)$.

Van der Waerden (normal scores) test. Данный тест является аналогом теста Крускала-Уоллеса за исключением того, что ранги сглаживаются при помощи обращения в нормальные квантили. EViews вычисляет статистику, которая приближенно распределена как $\chi^2(G-1)$.

Кроме того, в итоговой таблице представлены различные тестовые статистики для каждой подгруппы наблюдений: количество наблюдений, как в отдельно подгруппе, так и по выборке в целом (Count); медианы по подгруппам и по всей выборке (Median); количество наблюдений к подгруппам и выборке в целом, превышающих медиану по всей выборке (> Overall Median); средние значения рангов⁶ по подгруппам и по всей выборке (Mean Rank); и сглаженные средние значения ван дер Вардена для каждой подгруппы и выборки в целом (Mean Score).

Тесты на равенство дисперсий. Данная опция позволяет проверять гипотезу о равенстве между собой дисперсий различных подгрупп имеющейся выборки против альтернативной гипотезы о том, что дисперсия по крайней мере одной подгруппы отличается от остальных.

F -test – вычисляется только для случая наличия двух подгрупп ($G=2$). Затем вычисляются дисперсии каждой подгруппы s_L^2 и s_S^2 – большая и меньшая дисперсия соответственно. Тогда F -статистика – это отношение большей дисперсии к меньшей:

$$F = \frac{s_L^2}{s_S^2}.$$

Данная F -статистика имеет F -распределение с $(T_L - 1, T_S - 1)$ степенями свободы в предположении, что подвыборки нормально и независимо распределены.

⁶ Данные значения вычисляются следующим образом: сначала все наблюдения выборки упорядочиваются по возрастанию и каждому значению присваивается свой ранг (минимальному – 1, максимальному – T), после чего ранги наблюдений, принадлежащих данной подгруппе, суммируются и делятся на количество наблюдений, составляющих эту подгруппу.

Siegel-Tukey test. Аналогично предыдущему тесту, данный тест проводится для случая наличия двух подгрупп. Предполагается, что эти подвыборки независимы и имеют одинаковые медианы. Тестовая статистика вычисляется с использованием процедуры, аналогичной процедуре вычисления теста Крускала-Уоллеса (см. выше), но с иным распределением рангов. Сначала все наблюдения упорядочиваются по возрастанию, и наименьшему значению присваивается ранг 1. Наибольшему значению присваивается ранг 2, а значению, которое предшествует наибольшему – ранг 3. Значению, следующему за наименьшим, присваивается ранг 4, а следующему за ним – 5, и так далее. EViews вычисляет нормальное приближение статистики Сигеля-Таки с корректировкой на непрерывность.

Bartlett test. Данный тест основан на сравнении логарифма weighted average variance с взвешенной суммой логарифмов дисперсий. Нулевая гипотеза: дисперсии всех подгрупп совпадают. В предположении, что выборка нормально распределена, тестовая статистика имеет приближенное распределение $\chi^2(G-1)$. Заметим, однако, нулевая гипотеза предполагает чувствительность к отклонениям распределения выборки от нормального. EViews вычисляет скорректированную статистику Бартлета.

Levene test – основан на анализе дисперсий (ANOVA) абсолютных отклонений от среднего. Соответствующая F -statistic имеет приближенное F -распределение с $(G-1, T-G)$ степенями свободы.

Brown-Forsythe (modified Levene) test – модификация предыдущего теста, в которой абсолютные отклонения от математического ожидания заменяются на абсолютные отклонения от медианы. Данный тест превосходит предыдущий с точки зрения чувствительности (robustness) и мощности.

Test for Equality of Variances of IND						
Categorized by values of DT						
Date: 08/12/03 Time: 18:07						
Sample: 1995:01 2003:05						
Included observations: 101						
Method	df	Value	Probability			
F-test	(43, 56)	1.883776	0.0263			
Siegel-Tukey		0.715720	0.4742			
Bartlett	1	4.623970	0.0315			
Levene	(1, 99)	4.142851	0.0445			
Brown-Forsythe	(1, 99)	3.613701	0.0602			
Category Statistics						
DT	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.	Mean Tukey-Siegel Rank	
0	44	4.019442	3.291818	3.291564	53.38636	
1	57	5.516713	4.448680	4.411904	49.15789	
All	101	6.009558	3.944700	3.923835	51.00000	
Bartlett weighted standard deviation: 4.922649						

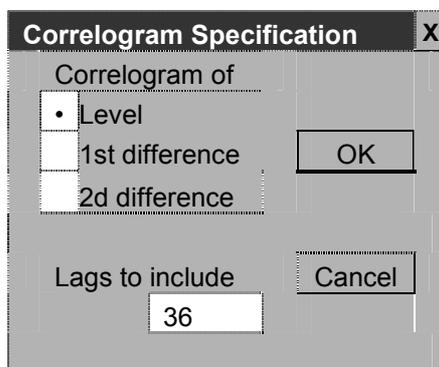
В таблице также приведены промежуточные статистики⁷ (Category Statistics), необходимые при вычислении различных тестовых статистик.

Задания для самостоятельной работы

1. Для процессов w_n , y_1 , y_2 проверьте гипотезы о том, что выборочные средние значения и стандартные отклонения этих процессов совпадают с теоретическими значениями.

2.5. Построение кореллограммы временного ряда (выборочных автокорреляционной и частной автокорреляционной функций)

Для построения кореллограммы временного ряда выберите опцию **Correlogram...** в меню **View** окна временного ряда – **View/Correlogram...**:



В появившемся окне выберите, будете ли Вы строить кореллограмму (Correlogram of) самого временного ряда (Level), либо его первых (1st difference) или вторых разностей (2d difference), а также количество лагов⁸, для которых Вы хотите построить кореллограмму⁹ (Lags to include). После того как Вы отметите все необходимые поля, нажмите ОК.

Series: Y1 Workfile: CP-2						
View	Procs	Object	Print	Name	Freeze	Sample
Genr	Sheet	Stats	Ident	LineBar		
Correlogram of Y1						
Date: 07/31/03 Time: 15:38						
Sample: 1 1000						
Included observations: 1000						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. ****	. ****	1	0.548	0.548	301.56	0.000
. **	.	2	0.313	0.017	399.80	0.000
. *	.	3	0.162	-0.023	426.25	0.000
. *	.	4	0.094	0.013	435.08	0.000
.	.	5	0.045	-0.012	437.14	0.000
.	.	6	0.027	0.006	437.85	0.000
.	.	7	-0.002	-0.025	437.86	0.000
.	.	8	0.012	0.031	438.00	0.000
.	.	9	-0.003	-0.022	438.01	0.000

⁷ Методики вычисления этих промежуточных статистик приведены в описании тестов.

⁸ Простое эмпирическое правило говорит, что достаточно взять $T/4$ запаздываний, где T – длина временного ряда (количество наблюдений).

⁹ При большом количестве наблюдений, как правило, по умолчанию предлагается построить кореллограмму для 36 запаздываний.

.	.	10	-0.012	-0.009	438.15	0.000
.	.	11	-0.018	-0.006	438.47	0.000
.	.	12	-0.022	-0.011	438.98	0.000
.	.	13	0.007	0.038	439.02	0.000

В первой строке появившегося окна представлена информация о том, с каким временным рядом Вы работаете (Y1), и из какого рабочего файла он взят (SP-2). Следующая строка – меню окна временного ряда:

View	Procs	Object	Print	Name	Freeze	Sample	Genr	Sheet	Stats	Ident	Line	Bar
------	-------	--------	-------	------	--------	--------	------	-------	-------	-------	------	-----

Строка Correlogram of Y1 показывает, что в окне представлена кореллограмма временного ряда Y1. Далее представлена информация о дате и времени открытия данного окна (Date: ... Time: ...), текущем размере выборки (Sample: ...) и количестве включенных наблюдений (Included observations: ...).

В столбцах Autocorrelation и Partial Correlation представлены графики выборочных автокорреляционной и частной автокорреляционной функций с соответствующими доверительными интервалами (пунктирные линии).

Напомним, что k -е значение выборочной автокорреляционной функции слабостационарного процесса y_t рассчитывается по формуле

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \mu)(y_{t-k} - \mu)}{\sum_{t=1}^T (y_t - \mu)^2},$$

где T – длина временного ряда y_t , μ – выборочное среднее временного ряда y_t . k -е значение выборочной частной автокорреляционной функции рассчитывается по формуле

$$\phi_{kk} = \phi_k = \begin{cases} r_1, & k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}, & k > 1 \end{cases}$$

где r_k – k -е значение выборочной автокорреляционной функции и

$$\phi_{k,j} = \phi_{k-1,j} - \phi_k \phi_{k-1,k-j},$$

и является состоятельной оценкой частной автокорреляционной функции.

Пунктирные линии на графиках выборочных автокорреляционной и частной автокорреляционной функций удалены от оси соответствующего графика на расстояние в два стандартных отклонения и вычисляются как $\pm 2/\sqrt{T}$. Если k -е значение выборочной автокорреляционной (либо частной автокорреляционной) функции находится внутри данного интервала, то можно говорить о том, что это значение (приблизительно) на 5%-м уровне значимости незначимо отличается от нуля.

В столбцах АС и РАС приведены численные значения выборочных автокорреляционной и частной автокорреляционной функций соответствующего порядка, информация о котором приведена в третьем столбце таблицы.

И, наконец, в столбцах Q-Stat и Prob приведены значения Q -статистики Льюнга-Бокса (Ljung-Box) и P -значения для этой статистики. Статистика Льюнга-Бокса порядка k вычисляется по формуле

$$Q_{LB}(k) = T(T+2) \sum_{j=1}^k \frac{r_j^2}{T-j}$$

и позволяет проверить нулевую гипотезу об отсутствии автокорреляции порядка меньшего либо равного k :

$$H_0 : \sum_{j=1}^k r_j^2 = 0$$

$$H_A : \sum_{j=1}^k r_j^2 > 0$$

Если тест Льюнга-Бокса применяется непосредственно к временному ряду y_t , то $Q_{LB}(k)$ асимптотически распределена как $\chi^2(k)$. Если же данный тест применяется к остаткам моделей типа $ARIMA(p, d, q)$, то $Q_{LB}(k)$ асимптотически распределена как $\chi^2(k - p - q)$.

Задания для самостоятельной работы

1. В файле DEMO.wf сгенерируйте случайные процессы авторегрессии второго порядка (y3), скользящего среднего первого порядка ($y4 = a_0 + wn + b_1 * wn(-1)$) и смешанный процесс авторегрессии скользящего среднего порядка (1, 1) (y5).
2. Для случайных процессов wn, y1-y5, постройте кореллограммы. Совпадают ли выборочные кореллограммы с теоретическими? В том случае, если они не совпадают, почему, по Вашему мнению, возникают такие расхождения?

2.6. Моделирование процессов типа $ARIMA(p, d, q)$

Эконометрический пакет EViews позволяет довольно легко моделировать случайные процессы типа $ARIMA(p, d, q)$, поскольку в пакете запрограммированы специальные команды, позволяющие оценивать соответствующие модели.

Авторегрессионные модели порядка $p - AR(p)$.

В качестве общего замечания остановимся на одной специфической особенности оценки авторегрессионных моделей в пакете EViews. В теории временных рядов под авторегрессионной моделью порядка p обычно понимается модель вида:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t, \quad (*)$$

где $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Оценить такую модель в пакете EViews можно несколькими способами (подробнее об этом будет сказано чуть ниже). Например, допустим, у Вас есть некий временной ряд y_1^{10} , и Вы хотите оценить для него авторегрессионную модель первого порядка. Тогда наберите в командной строке окна EViews

ls y5 c y5(-1).

В результате появится окно уравнения EViews,

Equation: UNTITLED Workfile: CP-2									
View	Procs	Objects	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: Y5									
Method: Least Squares									
Date: 08/13/03 Time: 17:18									
Sample(adjusted): 2 1000									
Included observations: 999 after adjusting endpoints									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	5.055556	0.096585	52.34281	0.0000					
Y5(-1)	-0.503753	0.027201	-18.51944	0.0000					
R-squared	0.255953	Mean dependent var		3.363014					
Adjusted R-squared	0.255207	S.D. dependent var		1.144243					
S.E. of regression	0.987497	Akaike info criterion		2.814714					
Sum squared resid	972.2257	Schwarz criterion		2.824538					
Log likelihood	-1403.950	F-statistic		342.9695					
Durbin-Watson stat	1.994097	Prob(F-statistic)		0.000000					

в котором отображены результаты оценивания модели. Меню окна уравнения частично совпадает с меню окна временного ряда, а частично содержит свои специфические опции:

- ✓ Estimate – данная опция меню позволяет оценить/переоценить модель;
- ✓ Forecast – позволяет строить прогнозы и вычислять соответствующие статистики;
- ✓ Stats – отображает в окне уравнения таблицу с оценками текущей модели;
- ✓ Resids – отображает в окне уравнения график остатков оцененной модели (Residual), график временного ряда (Actual) и график оценок временного ряда, полученных на основе оцененной модели (Fitted).

В пакете EViews содержатся встроенные функции – $ar(1)$, $ar(2)$, ..., $ar(p)$, – позволяющие оценивать авторегрессионные модели следующего типа (применительно к моделям временных рядов):

$$y_t = \alpha_0 + u_t, \quad (**)$$

где

$$u_t = \rho_1 u_{t-1} + \dots + \rho_p u_{t-p} + \varepsilon_t.$$

¹⁰ y_5 – случайный процесс авторегрессии первого порядка с константой, равной 5, и коэффициентом при авторегрессионном члене, равном (-0,5).

Можно показать, что оценки параметров α_i и ρ_i моделей (*) и (**) совпадают с точностью до свободного члена. Например, в рассмотренном выше примере, если оценить модель:

$$ls \ y5 \ c \ ar(1),$$

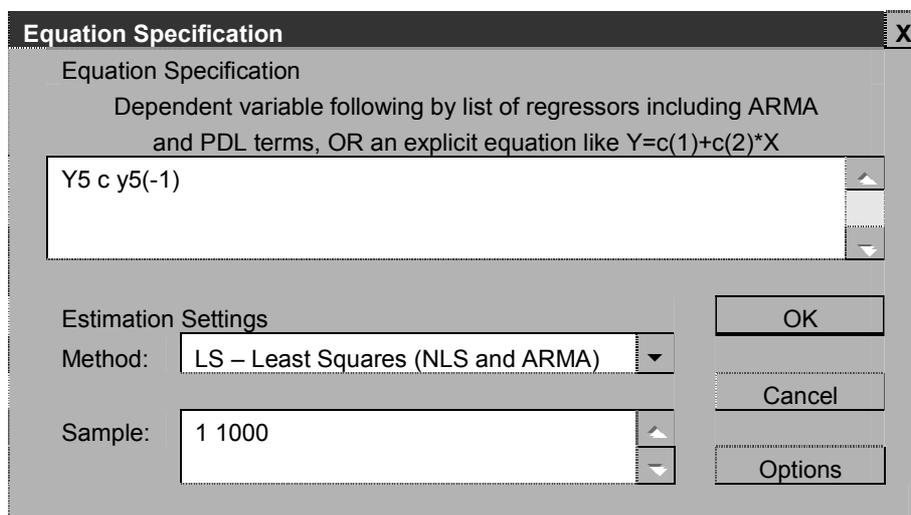
то оценки будут следующие:

Dependent Variable: Y5				
Method: Least Squares				
Date: 08/13/03 Time: 19:24				
Sample(adjusted): 2 1000				
Included observations: 999 after adjusting endpoints				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.361958	0.020777	161.8136	0.0000
AR(1)	-0.503753	0.027201	-18.51944	0.0000
R-squared	0.255953	Mean dependent var		3.363014
Adjusted R-squared	0.255207	S.D. dependent var		1.144243
S.E. of regression	0.987497	Akaike info criterion		2.814714
Sum squared resid	972.2257	Schwarz criterion		2.824538
Log likelihood	-1403.950	F-statistic		342.9695
Durbin-Watson stat	1.994097	Prob(F-statistic)		0.000000
Inverted AR Roots	-.50			

Данные результаты практически не отличаются от результатов оценки предыдущей модели: как уже говорилось, оценки моделей, а также полученные статистики, характеризующие качество моделей, совпадают с точностью до оценок константы.

Отметим, что модель (**) оценивается при помощи итерационных методов: фактически, информация об этом содержится в строке «Convergence achieved after 3 iterations» - «сходимость достигнута после 3 итераций». Кроме того, в данном случае приводится информация о характеристических корнях, вычисленных исходя из полученных оценок (Inverted AR Roots).

Получить оценку любой модели (в том числе и авторегрессионной) можно и при помощи различных опций Главного меню окна EViews. Например, если Вы выберете меню **Object/New Object/Equation**, то появится окно



В результате Вы получите оценку модели (*). Отметим, что в строке Method можно выбрать метод оценивания уравнения:

LS – метод наименьших квадратов;

TOLS – двухшаговый метод наименьших квадратов;

ARCH – авторегрессионная модель с условной гетероскедастичностью;

GMM – обобщенный метод моментов;

BINARY – бинарная модель (logit, probit, extreme value);

ORDERED – модель **порядкового выбора**;

CENSORED – модель, позволяющая оценить цензурированные данные;

COUNT – модель, предназначенная для оценки данных, являющихся натуральными числами.

В контексте данного пособия мы не будем останавливаться подробно на описании вышеперечисленных моделей: данную информацию можно найти практически в любом учебнике по эконометрике.

Оценить модель в EViews также можно, выбрав опцию **Quick/Estimate Equation...** в Главном меню основного окна EViews. Выбрав эту опцию, Вы получите окно, аналогичное предыдущему, в которое необходимо ввести нужную Вам спецификацию модели.

Прежде, чем приступить к дальнейшему изложению способов оценивания моделей $ARMA(p, q)$, сделаем несколько замечаний.

Замечание 1. Поскольку описанными выше способами можно оценить любое эконометрическое уравнение, мы будем использовать лишь один из них, например, на первый.

Замечание 2. Если Вы хотите оценить модель авторегрессии более высокого порядка, чем первый, например, второго

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t,$$

то в спецификации оцениваемого уравнения необходимо указывать оба авторегрессионных члена, т.е. необходимо написать в командной строке, например, следующее

$$\text{ls } y \text{ c } y(-1) \ y(-2)$$

(или

$$\text{ls } y \text{ c } \text{ar}(1) \ \text{ar}(2)).$$

Если же Вы напишите, например,

$$\text{ls } y \text{ c } y(-2),$$

то получите результат оценки модели

$$y_t = c + a_2 y_{t-2} + \varepsilon_t,$$

что само по себе является самостоятельным результатом, который Вы также имеет право рассматривать.

Аналогичное замечание относится и к авторегрессионным моделям более высоких порядков.

Замечание 3. В качестве общего замечания, касающегося оценки всех рассматриваемых моделей, остановимся на описании статистик, характеризующих качество модели, и присутствующих в стандартном окне уравнения EViews.

Статистики, характеризующие качество модели:

- ✓ Коэффициент детерминации R^2 (R-squared).

$$R^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(y - \hat{\mu})'(y - \hat{\mu})},$$

где $\hat{\mu}$ – выборочное среднее зависимой переменной y_t , $\hat{\varepsilon}$ – вектор-столбец случайных ошибок регрессии.

При стандартных предположениях коэффициент детерминации может быть интерпретирован как доля дисперсии зависимой переменной y_t , которая объясняется при помощи данного набора экзогенных переменных. Если регрессия оценивается методом наименьших квадратов, значения коэффициента детерминации изменяются от 0 до 1. В некоторых случаях (например, при использовании других методов оценивания) коэффициент детерминации может быть отрицательным.

- ✓ Скорректированный коэффициент детерминации R_{adj}^2 (Adjusted R-squared).

$$R_{adj}^2 = 1 - (1 - R^2) \frac{T-1}{T-k},$$

где k – число оцениваемых параметров.

Поскольку обычный коэффициент детерминации R^2 не уменьшается при включении в оцениваемую модель дополнительных переменных, то он не может служить хорошей мерой качества множественной регрессии. При расчете скорректированного коэффициента детерминации вводится штраф за дополнительные регрессоры, которые не способствуют увеличению объясняющей силы регрессии. Значения скорректированного коэффициента детерминации не превышают соответствующих значений обычного коэффициента детерминации, могут уменьшаться при включении в регрессию дополнительных переменных и могут быть отрицательными, если модель плохо специфицирована.

- ✓ Стандартная ошибка регрессии $s.e.regr.$ (S.E. of regression).

$$s.e.regr. = \sqrt{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{T-k}}$$

- ✓ Сумма квадратов остатков регрессии RSS (Sum squared resid). Данный показатель приводится в окне регрессии для удобства пользователя, поскольку используется для расчетов многочисленных статистических характеристик регрессии.

$$RSS = SSR = \hat{\varepsilon}'\hat{\varepsilon} = \sum_{t=1}^T (y_t - X_t' b)^2,$$

где X_t – матрица объясняющих переменных, b – вектор-столбец коэффициентов при объясняющих переменных соответствующих размерностей.

- ✓ Логарифм функции правдоподобия l (Log likelihood).

$$l = -\frac{T}{2} \left(1 + \log(2\pi) + \log \frac{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}{T} \right).$$

Логарифм функции правдоподобия вычисляется в предположении, что остатки модели нормально распределены.

- ✓ Статистика Дарбина-Ватсона DW (Durbin-Watson stat). Позволяет определить (при определенных условиях на параметры модели) наличие автокорреляции остатков первого порядка.

$$DW = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2}$$

- ✓ Среднее значение и стандартное отклонение зависимое переменная $\hat{\mu}$ и $\hat{\sigma}_y$ (Mean dependent var и S.D. dependent var). Среднее значение и стандартное отклонение зависимой переменной вычисляются с использованием стандартных формул:

$$\hat{\mu} = \sum_{t=1}^T \frac{y_t}{T}$$

и

$$\hat{\sigma}_y = \sqrt{\sum_{t=1}^T \frac{(y_t - \hat{\mu})^2}{T-1}}.$$

- ✓ Информационный критерий Акаике AIC (Akaike info criterion).

$$AIC = -2 \frac{l}{T} + 2 \frac{k}{T},$$

где l логарифм функции правдоподобия, k – количество оцениваемых в модели параметров. Информационный критерий Акаике, также как и информационный критерий Шварца, используется для выбора лучшей модели из некоторого набора альтернативных моделей – чем меньше значение критерия, тем лучше модель.

- ✓ Информационный критерий Шварца BIC или SC (Schwarz criterion).

$$BIC = SC = -2 \frac{l}{T} + \frac{k \log T}{T}.$$

Информационный критерий Шварца всегда выбирает лучшую модель с числом параметров, не превышающим число параметров в модели, которая была выбрана по критерию Акаике. Кроме того, критерий Шварца является асимптотически состоятельным, в то время как информационный критерий Акаике смещен в сторону выбора перепараметризованной модели.

- ✓ F -статистика (F-statistic). При помощи F -статистики в предположении, что остатки модели распределены нормально, проверяется гипотеза о значимости регрессии в целом, т.е. проверяется нулевая гипотеза о том, что коэффициенты при всех экзогенных переменных, включенных в модель, кроме свободного члена, значимо отличаются от нуля.

$$F = \frac{R^2 / (k-1)}{(1-R^2) / (T-k)},$$

где k – число ограничений в модели, т.е. число оцениваемых параметров (включая константу). Также в окне регрессии EViews приводится P -значение для F -статистики (Prob(F-statistic)). Если P -значение меньше, чем уровень значимости, на котором Вы проверяете нулевую гипотезу, то гипотезу о том, что все коэффициенты модели равны нулю, можно отвергнуть на этом уровне значимости. Помните, что регрессия может быть значимой, даже если каждый коэффициент в отдельности не значим.

Замечание 4. Для тестирования остатков модели необходимо воспользоваться опцией **Residual Tests** меню **View** окна уравнения. В данном меню доступны следующие опции:

- ✓ Correlogram – Q-statistics – позволяет построить кореллограмму остатков текущей модели (более подробно см. пункт 2.5);
- ✓ Correlogram Squared Residuals – позволяет построить кореллограмму квадратов остатков;
- ✓ Histogram – Normality Test – выводит гистограмму остатков и приводит результаты теста на нормальность остатков (более подробно см. пункт 2.3);
- ✓ Serial Correlation LM Test... – тест на серийную коррелированность остатков Бройша-Годфри. Данный тест позволяет проверить гипотезу о том, что остатки модели описываются моделью авторегрессии порядка p :

$$\varepsilon_t = \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_p \varepsilon_{t-p} + u_t,$$

где $u_t \sim WN(0, \sigma_u^2)$. Таким образом, проверяется следующая нулевая гипотеза:

$$H_0: \gamma_1 = \dots = \gamma_p$$

против альтернативной:

$$H_A: \gamma_1^2 + \dots + \gamma_p^2 > 0.$$

Тогда соответствующая статистика рассчитывается по формуле TR^2 , где T – число наблюдений временного ряда, а R^2 – коэффициент множественной детерминации регрессии

$$e_t = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \gamma_1 e_{t-1} + \gamma_2 e_{t-2} + \dots + \gamma_p e_{t-p} + u_t,$$

где e_t – остатки модели $y_t = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \varepsilon_t$. Статистика TR^2 имеет асимптотическое распределение $\chi^2(p)$.

- ✓ ARCH LM Test... – позволяет проверить гипотезу о том, что случайные ошибки остатки описываются моделью ARCH;
- ✓ White Heteroskedastisity (no cross terms) – проверяет гипотезу о гетероскедастичности остатков модели (без включения смешанных произведений объясняющих переменных);

- ✓ White Heteroskedastisity (cross terms) – проверяет гипотезу о гетероскедастичности остатков модели (с включением смешанных произведений объясняющих переменных).

Модели скользящего среднего порядка q – $MA(q)$

Оценка моделей скользящего среднего осуществляется при помощи специальных функций $ma(1)$, $ma(2)$..., встроенной в EViews. Например, если Вам нужно оценить модель

$$y_t = c + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \beta_3 \varepsilon_{t-3},$$

то в командной строке необходимо набрать

```
ls c ma(1) ma(2) ma(3).
```

Заметим, что аналогично авторегрессионным моделям необходимо включать в спецификацию уравнения все переменные $ma(q)$, содержащиеся в регрессионном уравнении. Результат оценки некоторого процесса скользящего среднего третьего порядка $ma3$ приведен в таблице

Dependent Variable: MA3				
Method: Least Squares				
Date: 09/09/03 Time: 17:24				
Sample(adjusted): 4 998				
Included observations: 995 after adjusting endpoints				
Convergence achieved after 30 iterations				
Backcast: 1 3				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.028819	0.017601	58.45392	0.0000
MA(1)	-1.383183	0.031186	-44.35218	0.0000
MA(2)	0.640902	0.049815	12.86569	0.0000
MA(3)	-0.186233	0.031133	-5.981893	0.0000
R-squared	0.669620	Mean dependent var		1.022033
Adjusted R-squared	0.668620	S.D. dependent var		13.37097
S.E. of regression	7.697083	Akaike info criterion		6.923572
Sum squared resid	58711.88	Schwarz criterion		6.943281
Log likelihood	-3440.477	F-statistic		669.5251
Durbin-Watson stat	1.915246	Prob(F-statistic)		0.000000
Inverted MA Roots	.90	.24+.39i	.24-.39i	

Структура этой таблицы ничем не отличается от стандартных таблиц с результатами оценки регрессий. Отметим лишь, что в последней строке «Inverted MA Roots» приведены значения соответствующих характеристических корней, исходя из абсолютных величин которых можно сделать вывод об обратимости оцененной модели.

Смешанные модели авторегрессии и скользящего среднего порядка (p, q) – $ARMA(p, q)$

В отличие от двух предыдущих случаев в пакете EViews нет общей встроенной функции, позволяющей оценивать такие модели. Поэтому, чтобы оценить любую смешанную модель необходимо указать в ее спецификации все включаемые авторегрессионные члены и все включаемые запаздывания случайного возмущения. К примеру, если Вы хотите оценить модель $ARMA(2, 1)$, то в командной строке необходимо, например, написать

ls c ar(1) ar(2) ma(1)¹¹.

Итоговая таблица с оценками соответствующей модели практически не отличается от предыдущих случаев и содержит информацию как о характеристических авторегрессионных корнях, так и о характеристических корнях, вычисленных для *МА*-части:

Dependent Variable: ARMA21
 Method: Least Squares
 Date: 09/10/03 Time: 17:11
 Sample(adjusted): 3 1000
 Included observations: 998 after adjusting endpoints
 Convergence achieved after 9 iterations
 Backcast: 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.746622	0.184022	20.35962	0.0000
AR(1)	1.495641	0.016269	91.93400	0.0000
AR(2)	-0.894885	0.014565	-61.44169	0.0000
MA(1)	-0.524807	0.032300	-16.24770	0.0000
R-squared	0.832956	Mean dependent var		3.728876
Adjusted R-squared	0.832452	S.D. dependent var		11.91745
S.E. of regression	4.878129	Akaike info criterion		6.011401
Sum squared resid	23653.37	Schwarz criterion		6.031063
Log likelihood	-2995.689	F-statistic		1652.174
Durbin-Watson stat	1.951439	Prob(F-statistic)		0.000000
Inverted AR Roots	.75+.58i	.75-.58i		
Inverted MA Roots	.52			

Сезонные авторегрессионные модели и сезонные модели скользящего среднего

Отметим в первую очередь, что, как и в случае обычных авторегрессионных моделей или моделей скользящего среднего, встроенные в EViews специальные функции, которые позволяют оценивать сезонные модели, отражают соответствующую спецификацию ошибок оцениваемых регрессий и позволяют оценить модели двух типов (в данном случае рассмотрен пример для месячных данных):

$$(1 - \alpha_1 L - \dots - \alpha_p L^p)(1 - \phi L^{12})u_t = \varepsilon_t$$

или

$$u_t = (1 + \beta_1 L + \dots + \beta_q L^q)(1 + \omega L^{12})\varepsilon_t.$$

Тогда спецификация исходной модели имеет вид

$$y_t = c + u_t,$$

где u_t – случайная ошибка регрессия одного из описанных выше типов. Чтобы оценить мультипликативную сезонную модель, необходимо в спецификации модели указать не только все включаемые авторегрессионные члены и запаздывания случайной ошибки регрессии, но также порядок и тип сезонности: $ar(12)$ ¹² для сезонных авторегрессионных моделей и $ma(12)$ – для сезонных

¹¹ Вместо $ar(1)$ и $ar(2)$ можно написать первое и второе запаздывание самого временного ряда.

¹² В случае наличия квартальных данных необходимо в скобках указывать число 4: $ar(4)$ или $ma(4)$.

моделей скользящего среднего. Например, если Вы хотите оценить сезонную авторегрессионную модель первого порядка (ряд `s_ar`), напишите в командной строке окна EViews

```
ls s_ar c ar(1) ar(12).
```

Результат оценки такой модели представлен в таблице, имеющей стандартную структуру:

Dependent Variable: S_AR
Method: Least Squares
Date: 07/15/03 Time: 14:34
Sample(adjusted): 14 1000
Included observations: 987 after adjusting endpoints
Convergence achieved after 13 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.006013	0.245655	8.165992	0.0000
AR(1)	0.574613	0.026095	22.02035	0.0000
SAR(12)	-0.620051	0.025378	-24.43280	0.0000
R-squared	0.728619	Mean dependent var		2.017917
Adjusted R-squared	0.728068	S.D. dependent var		10.19894
S.E. of regression	5.318454	Akaike info criterion		6.183277
Sum squared resid	27833.38	Schwarz criterion		6.198155
Log likelihood	-3048.447	F-statistic		1320.952
Durbin-Watson stat	1.963491	Prob(F-statistic)		0.000000
Inverted AR Roots	.93 -.25i	.93+.25i	.68 -.68i	.68+.68i
	.57	.25 -.93i	.25+.93i	-.25+.93i
	-.25 -.93i	-.68 -.68i	-.68 -.68i	-.93 -.25i
	-.93+.25i			

Если теперь Вы хотите оценить данную модель как авторегрессионную модель первого порядка с сезонным скользящим средним, то в командной строке напишите

```
ls s_ar c ar(1) sma(12).
```

Заметим, что полученные оценки

Dependent Variable: S_AR
Method: Least Squares
Date: 07/15/03 Time: 14:36
Sample(adjusted): 2 1000
Included observations: 999 after adjusting endpoints
Convergence achieved after 13 iterations
Backcast: -10 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.994558	0.302074	6.602889	0.0000
AR(1)	0.698247	0.022740	30.70528	0.0000
MA(12)	-0.485910	0.027931	-17.39699	0.0000
R-squared	0.702925	Mean dependent var		2.012653
Adjusted R-squared	0.702328	S.D. dependent var		10.13974
S.E. of regression	5.532175	Akaike info criterion		6.262037
Sum squared resid	30482.54	Schwarz criterion		6.276772
Log likelihood	-3124.888	F-statistic		1178.342
Durbin-Watson stat	1.983751	Prob(F-statistic)		0.000000
Inverted AR Roots	.70			
Inverted MA Roots	.94	.82+.47i	.82 -.47i	.47+.82i
	.47 -.82i	.00 -.94i	-.00+.94i	-.47 -.82i
	-.47+.82i	-.82+.47i	-.82 -.47i	-.94

будут фактически совпадать с оценками авторегрессионной модели первого порядка с аддитивным сезонным скользящим средним, т.е. с результатом оценки модели

$$y_t = c + u_t,$$

где $(1 - a_1 L)u_t = (1 + \omega L^{12})\varepsilon_t$.

Кроме мультипликативных сезонных моделей существуют и аддитивные сезонные модели. Например, рассмотренная выше сезонная авторегрессионная модель первого порядка является мультипликативной, т.е. она представима в виде

$$y_t = c + u_t,$$

где $(1 - a_1 L)(1 - \phi L^{12})u_t = \varepsilon_t$. Если же мы хотим рассмотреть аддитивную сезонную авторегрессионную модель первого порядка, т.е. модель вида

$$y_t = c + u_t,$$

где $(1 - a_1 L - \phi L^{12})u_t = \varepsilon_t$, то в командной строке необходимо написать

```
ls s_ar c ar(1) ar(12).
```

Оценки, полученные в этом случае, будут довольно сильно отличаться от оценок мультипликативной модели:

Dependent Variable: S_AR
Method: Least Squares
Date: 07/15/03 Time: 14:37
Sample(adjusted): 13 1000
Included observations: 988 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.007593	0.149227	13.45325	0.0000
AR(1)	0.475595	0.018907	25.15399	0.0000
AR(12)	-0.513877	0.019092	-26.91633	0.0000
R-squared	0.772290	Mean dependent var		2.023927
Adjusted R-squared	0.771827	S.D. dependent var		10.19552
S.E. of regression	4.870138	Akaike info criterion		6.007153
Sum squared resid	23362.47	Schwarz criterion		6.022019
Log likelihood	-2964.534	F-statistic		1670.336
Durbin-Watson stat	1.962955	Prob(F-statistic)		0.000000
Inverted AR Roots	.97 -.24i	.97+.24i	.71 -.66i	.71+.66i
	.28+.90i	.28-.90i	-.21 -.91i	-.21+.91i
	-.64 -.66i	-.64+.66i	-.88 -.24i	-.88+.24i

Аналогичное замечание относится и к сезонным моделям скользящего среднего.

Отметим также, что вместо встроенных функций $ar(1)$ и $ar(12)$ можно писать $s_ar(-1)$ $s_ar(-12)$ – полученные оценки будут совпадать с результатами, указанными в последней таблице с точностью до константы.

Модели $ARIMA(p, d, q)$

Модель $ARIMA(p, d, q)$ можно записать при помощи оператора запаздывания L в следующем виде:

$$(1 - a_1 L - \dots - a_p L^p) \Delta^d y_t = c + (1 + \beta_1 L + \dots + \beta_q L^q) \varepsilon_t,$$

где $\Delta = 1 - L$ – оператор разности. В EViews встроены специальные функции, позволяющие компактно записывать разности различных порядков:

$d(y)$ – используется для обозначения разности первого порядка временного ряда y , т.е. эквивалентна разности $y_t - y_{t-1} = (1 - L)y_t = \Delta y_t$;

$d(y, n, s)$ – используется для обозначения сезонной разности порядка (n, s) временного ряда y , т.е. равносильна записи $(1 - L)^n (1 - L^s)y_t = \Delta^n (1 - L^s)y_t$;

$dlog(y, n, s)$ – используется для обозначения сезонной разности порядка (n, s) логарифма временного ряда y , т.е. равносильна записи $(1 - L)^n (1 - L^s)\log y_t = \Delta^n (1 - L^s)\log y_t$.

Чтобы получить оценки модели $ARIMA(p, d, q)$ необходимо оценить модель в разностях порядка d , т.е. модель вида

$$\Delta^d y_t = c + (a_1 L + \dots + a_p L^p) \Delta^d y_t + (1 + \beta_1 L + \dots + \beta_q L^q) \varepsilon_t.$$

Например, чтобы оценить модель $ARIMA(1, 1, 0)$

$$(1 - a_1 L) \Delta y_t = c + \varepsilon_t,$$

в командной строке рабочего окна EViews введите:

```
ls d(y) c d(y(-1))
```

или

```
ls d(y) c ar(1).
```

Соответствующие оценки будут совпадать с точностью до константы:

Dependent Variable: D(Y)
 Method: Least Squares
 Date: 07/15/03 Time: 18:53
 Sample(adjusted): 3 1000
 Included observations: 998 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.106930	0.190197	11.07765	0.0000
D(Y(-1))	0.473341	0.027914	16.95697	0.0000
R-squared	0.224020	Mean dependent var		3.992921
Adjusted R-squared	0.223241	S.D. dependent var		5.530322
S.E. of regression	4.874090	Akaike info criterion		6.007746
Sum squared resid	23661.73	Schwarz criterion		6.017577
Log likelihood	-2995.865	F-statistic		287.5388
Durbin-Watson stat	1.956180	Prob(F-statistic)		0.000000

и во втором случае:

Dependent Variable: D(Y)
 Method: Least Squares
 Date: 07/15/03 Time: 18:57
 Sample(adjusted): 3 1000
 Included observations: 998 after adjusting endpoints
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.000559	0.292955	13.65589	0.0000
AR(1)	0.473341	0.027914	16.95697	0.0000
R-squared	0.224020	Mean dependent var		3.992921

Adjusted R-squared	0.223241	S.D. dependent var	5.530322
S.E. of regression	4.874090	Akaike info criterion	6.007746
Sum squared resid	23661.73	Schwarz criterion	6.017577
Log likelihood	-2995.865	F-statistic	287.5388
Durbin-Watson stat	1.956180	Prob(F-statistic)	0.000000
Inverted AR Roots		.47	

Для оценки моделей $ARIMA(p, d, q)$, содержащие запаздывающие значения случайной ошибки необходимо включить в спецификацию уравнения соответствующие члены: $ma(1)$, $ma(2)$ и т.д.

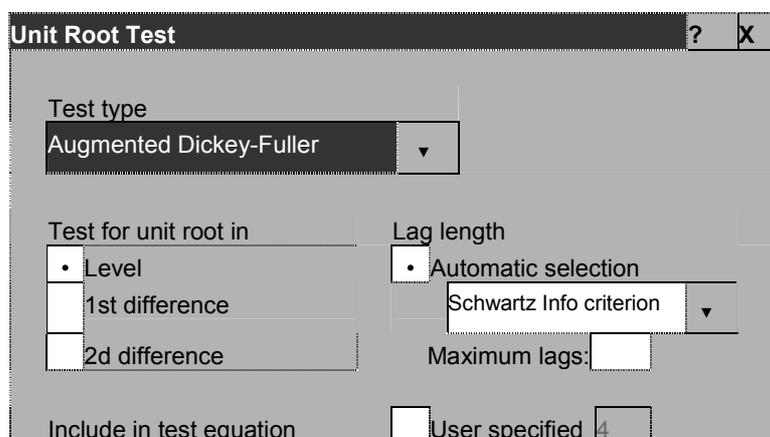
Задания для самостоятельной работы

1. Для каждого из случайных процессов w_n , y_1 - y_5 оцените модели $ARMA(p, q)$ при $p, q \leq 2$.
2. В каждом из случаев выберите наилучшую модель с точки зрения информационных критериев Акаике и Шварца. Совпадают ли выбранные модели с теоретическими.
3. Проверьте остатки выбранных моделей на нормальность и автокоррелированность.

2.7. Тесты на единичные корни (Тест Дикки-Фуллера)

EViews 4.1 предлагает несколько тестов на наличие единичных корней: расширенный тест Дикки-Фуллера (Augmented Dickey-Fuller Test), тест Филлипса-Перрона¹³ (Phillips-Perron Test), тесты Эллиота-Розенберга-Стока (Dickey-Fuller GLS Test & Elliot-Rothenberg-Stock Point-Optimal Test), тест Кватковского-Филлипса-Шмидта-Шина (Kwiatkowski-Phillips-Shmidt-Shin Test – KPSS Tests) и тест Нг-Перрона (Ng-Perron Test). В рамках настоящего пособия будет рассмотрен лишь расширенный тест Дикки-Фуллера.

Для тестирования временного ряда на наличие единичных корней с использованием расширенного теста Дикки-Фуллера в открытом окне временного ряда выберите опцию **View/Unit Root Test...** В появившемся окне



¹³ Тесты Дикки-Фуллера и Филлипса-Перрона также есть в EViews 3.1.



выберите:

- ✓ будете ли Вы тестировать на наличие единичных корней сам временной ряд (Level), либо его первые разности (1st difference), либо его вторые разности (2d difference);
- ✓ необходимо ли включать в тестируемое уравнение свободный член (Intercept), либо тренд и свободный член (Trend and intercept), либо не включать ничего (None);
- ✓ какое количество запаздывающих разностей Вы включаете (Lag length). Количество запаздывающих разностей, включаемых в тестируемое уравнение, можно выбрать как автоматически (Automatic selection) с использованием одного из информационных критериев (по умолчанию - критерий Шварца¹⁴ (Schwartz Info criterion))¹⁵, либо задать вручную при помощи опции User specified (по умолчанию при выборе данной опции, как правило, предлагается включить 4 запаздывающих разности).

После того как Вы выбрали все необходимые параметры, нажмите кнопку ОК. Результаты расширенного теста Дикки-Фуллера будут приведены в следующем окне (окно временного ряда):

Series: Y1 Workfile: CP-2				
View Procs Object Print Name Freeze Sample Genr Sheet Stats Ident LineBar				
Augmented Dickey-Fuller Unit Root Test on Y1				
Null Hypothesis: Y1 has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic based on SIC, MAXLAG=21)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-17.08816	0.0000
Test critical values:		1% level	-2.567279	
		5% level	-1.941140	
		10% level	-1.616486	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Y1)				
Method: Least Squares				
Date: 07/25/03 Time: 18:40				
Sample (adjusted): 2 1000				
Included observations: 999 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.

¹⁴ Кроме информационного критерия Шварца можно использовать информационные критерии Акаике (Akaike) и Хэннана-Квина (Hannan-Quinn), а также модифицированные информационные критерии Акаике, Шварца и Хэннана-Квина.

¹⁵ Данная возможность не реализована в пакете EViews 3.1

Y1(-1)	-0.451460	0.02641	-17.08816	0.0000
		9		
R-squared	0.226358	Mean dependent var		0.002017
Adjusted R-squared	0.226358	S.D. dependent var		1.198658
S.E. of regression	1.054303	Akaike info criterion		2.944637
Sum squared resid	1109.331	Schwarz criterion		2.949549
Log likelihood	-1469.846	Durbin-Watson stat		2.018780

В первой строке таблицы представлена информация о том, с каким временным рядом Вы работаете (Y1), и из какого рабочего файла он взят (CP-2). Следующая строка – меню окна временного ряда:

View	Procs	Object	Print	Name	Freeze	Sample	Genr	Sheet	Stats	Ident	LineBar
------	-------	--------	-------	------	--------	--------	------	-------	-------	-------	---------

Строка Augmented Dickey-Fuller Unit Root Test on Y1 показывает, что в окне отображены результаты теста Дикки-Фуллера для временного ряда Y1. Затем следует информация о том, что тестируется гипотеза о наличии единичного корня (Null Hypothesis: Y1 has a unit root) в случае, когда в тестируемое уравнение не включены ни свободный член, ни тренд (Exogenous: None), и по информационному критерию Шварца была выбрана модель без включения запаздывающих разностей (Lag Length: 0 (Automatic based on SIC, MAXLAG=21)), т.е. тестируется модель:

$$\Delta Y1_t = \gamma Y1_{t-1} + e_t,$$

а нулевая гипотеза имеет вид:

$$H_0: \gamma = 0.$$

Далее в таблице приведены значение статистики Дикки-Фуллера (Augmented Dickey-Fuller test statistic, в данном случае – τ_0) и ее P-значение, а также критические значения этой статистики на 1, 5 и 10%-м уровнях значимости:

Test critical values:	1% level	-2.567279
	5% level	-1.941140
	10% level	-1.616486

Последняя часть таблицы – результаты оценки модели – по своей структуре повторяет стандартную таблицу EViews, в которой представляются результаты оценки регрессий.

Задания для самостоятельной работы

1. В файле DEMO.wf сгенерируйте случайные процессы:

$$y6 = 0,99 * y6(-1) + wn$$

$$y7 = y7(-1) + wn$$

$$y8 = 0,7 + y8(-1) + wn$$

2. Постройте графики данных случайных процессов. Обсудите графические свойства этих временных рядов.
3. Используя тест Дикки-Фуллера, проверьте гипотезу о наличии единичного корня для случайных процессов y2, y6-y8.

4. Прокомментируйте результаты теста Дикки-Фуллера: для всех ли случаев полученные результаты совпадают с теоретическими моделями? Если не совпадают, то каковы причины таких несоответствий?

2.8. Прогнозирование при помощи моделей $ARMA(p, q)$

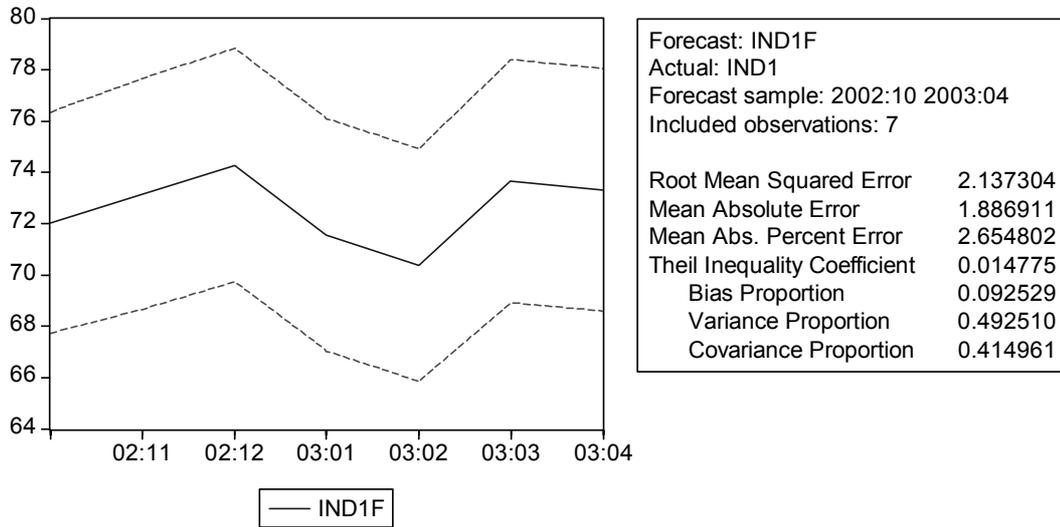
Для большей наглядности рассмотрим методы построения прогнозов в пакете EViews на примере модели $ARMA(p, q)$ индекса промышленного производства Центра экономической конъюнктуры при Правительстве РФ¹⁶ промышленности в целом (IND1). На первом этапе для данного временного ряда была оценена некоторая модель на интервале с октября 1998 г. по сентябрь 2002 г. (Допустим, что у Вас есть данные только за этот период.) Затем при помощи опции **Forecast** меню окна уравнения был построен прогноз для этого временного ряда с октября 2002 г. по апрель 2003 г. (чтобы сделать это, необходимо при создании рабочего файла указать размер выборки с октября 1998 г. по апрель 2003 г.):

Отметьте в появившемся окне необходимые опции: название ряда прогнозов (Forecast name), название ряда ошибок прогнозирования (S.E. (optional)), границы интервала, на котором строите прогнозы (Forecast sample), метод прогнозирования (Method: Dynamic, Static или Structural) и те результаты, которые Вы хотите получить при построении прогнозов (Output: Do graph и Forecast evaluation). Отметим разницу между динамическими и статическими прогнозами:

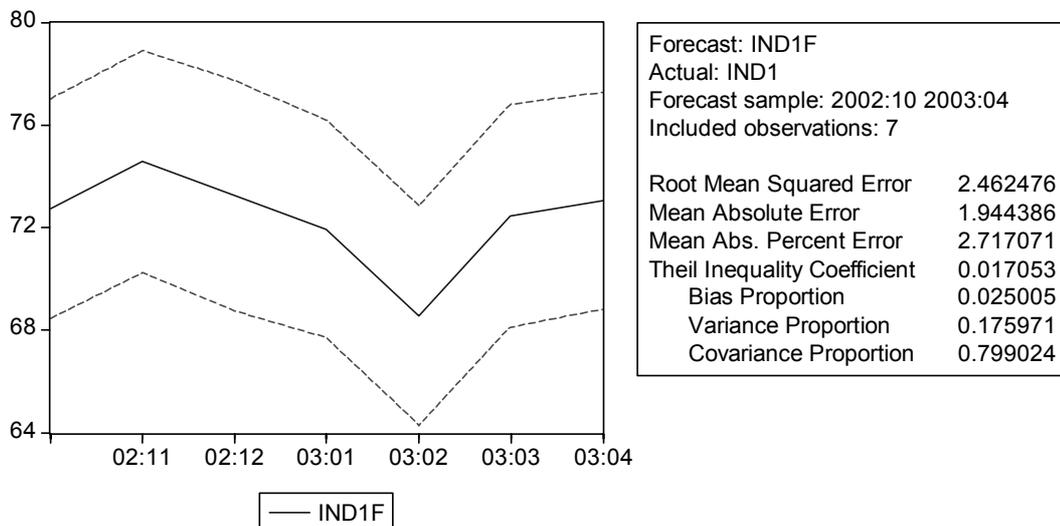
- ✓ Динамические прогнозы. Допустим, Вы оценили модель авторегрессии второго порядка для некоторого временного ряда и хотите построить соответствующие прогнозы на h шагов вперед. Тогда, строя прогноз на один шаг вперед, в качестве запаздывающих значений переменной Вы должны использовать ее истинные значения в данные моменты времени. Но при построении прогноза на 2 шага вперед вместо переменной y_{t-1} уже

¹⁶ Базисный месяц – январь 1993 г.

необходимо использовать ее прогнозное значение, полученное на предыдущем шаге, а вместо переменной y_{t-2} ее истинное значение. Начиная с прогноза на 3 шага вперед, в качестве запаздывающих значений прогнозируемых переменных необходимо использовать их прогнозные значения, полученные на предыдущих шагах. Результаты прогнозирования, полученные этим методом, приведены ниже:



- ✓ Статические прогнозы. В отличие от динамических прогнозов для прогнозирования используются только истинные значения, т.е. фактически, если у нас есть данные на интервале от 1 до T , то построить прогноз по моделям $ARMA(p, q)$ мы можем только на один шаг вперед \hat{y}_{T+1} . В нашем примере можно построить статический прогноз на 7 месяцев, поскольку у нас есть все данные:



Справа от графика в тех случаях, когда известны истинные значения временного ряда на прогнозируемом периоде, приведены статистики, характеризующие качество прогноза:

Root Mean Squared Error – квадратный корень из средней квадратичной ошибки:

$$RMSE = \sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h+1}},$$

где h – длина интервала прогнозирования, \hat{y}_t – прогнозное значение временного ряда, y_t – истинное значение временного ряда;

Mean Absolute Error – средняя абсолютная ошибка:

$$MAE = \sum_{t=T+1}^{T+h} \frac{|\hat{y}_t - y_t|}{h+1};$$

Mean Abs. Percent Error – средняя абсолютная ошибка (в %):

$$MAPE = \sum_{t=T+1}^{T+h} \frac{\left| \frac{\hat{y}_t - y_t}{y_t} \right|}{h+1};$$

Theil Inequality Coefficient – коэффициент Тэйла:

$$TIC = \frac{\sqrt{\sum_{t=T+1}^{T+h} \frac{(\hat{y}_t - y_t)^2}{h+1}}}{\sqrt{\sum_{t=T+1}^{T+h} \frac{\hat{y}_t^2}{h+1} + \sum_{t=T+1}^{T+h} \frac{y_t^2}{h+1}}};$$

Bias Proportion – показывает смещение среднего значения прогноза временного ряда относительно среднего значения реального временного ряда:

$$\frac{\left(\left(\sum \frac{\hat{y}_t}{h} \right) - \bar{y} \right)^2}{\sum \frac{(\hat{y}_t - y_t)^2}{h}},$$

где \bar{y} – среднее значение временного ряда;

Variance Proportion – показывает смещение дисперсии прогноза временного ряда относительно дисперсии реального временного ряда:

$$\frac{(s_{\hat{y}} - s_y)^2}{\sum \frac{(\hat{y}_t - y_t)^2}{h}},$$

где $s_{\hat{y}}$ и s_y – смещенные стандартные отклонения прогноза временного ряда и истинного временного ряда;

Covariance Proportion – измеряет остаточную несистематическую ошибку прогнозирования:

$$\frac{2(1-r)s_y s_y}{\sum \frac{(y_t - y_t)^2}{h}}$$

где r – коэффициент корреляции между \hat{y}_t и y_t .

Задания для самостоятельной работы

1. Для случайных процессов w_1, y_2, y_4-y_8 на подинтервале от 1 до 100 оцените наилучшую модель в терминах $ARIMA(p, d, q)$.
2. По оцененным моделям постройте прогнозы 101, ..., 110 значений для каждого случайного процесса
3. Совпадают ли полученные прогнозные значения со средним значением соответствующего временного ряда? Что Вы можете сказать о дисперсии ошибок прогнозирования в каждом из случаев?

3. Анализ многомерных временных рядов

3.1. Создание группы временных рядов (многомерного временного ряда)

Простейший способ создать группу временных рядов состоит в следующем: нажмите клавишу **Ctrl** на клавиатуре и при помощи левой кнопки мыши выделите в окне рабочего файла необходимые для создания группы ряды. После этого, не убирая указатель мыши с выделенного синим фона, нажмите правую кнопку мыши и в появившемся окне выберите опцию **Open/as Group**. В открывшемся окне в одной таблице будут собраны все выделенные ряды на указанном подинтервале. Окно группы временных рядов содержит те же опции Меню, что и окно временного ряда, за исключением опций **Label+/-** и **Wide+/-**, описание которых дано в пункте 2.1.

Создать группу временных рядов также можно воспользовавшись меню **Object/New object/Group** Главного меню EViews или при помощи команды

`group <название группы> <список временных рядов>`, которую необходимо написать в командной строке.

Задания для самостоятельной работы

1. В файле DEMO.wf сгенерируйте случайные процессы

```
rw=rw(-1)+nrnd
```

```
rw1=rw1(-1)+nrnd
```

```
rw2=rw+nrnd
```

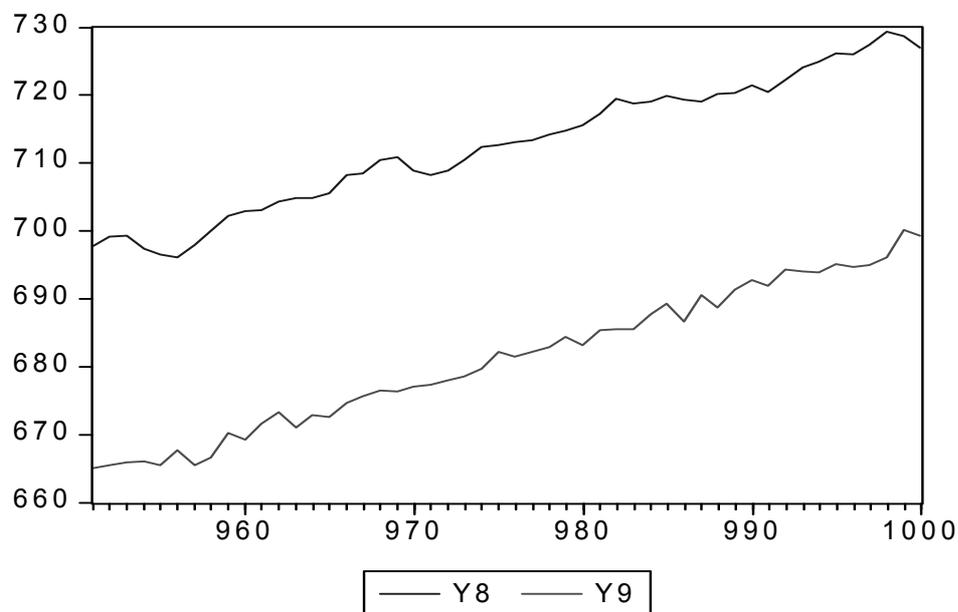
```
rw3=1+rw+nrnd
```

```
rw4=-1+0,5*@trend+rw+nrnd.
```

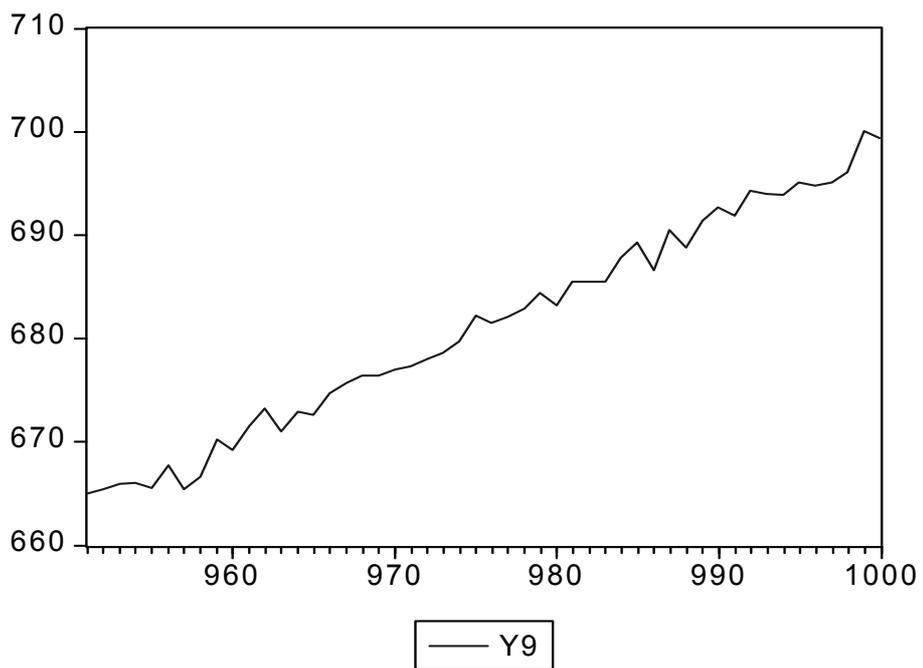
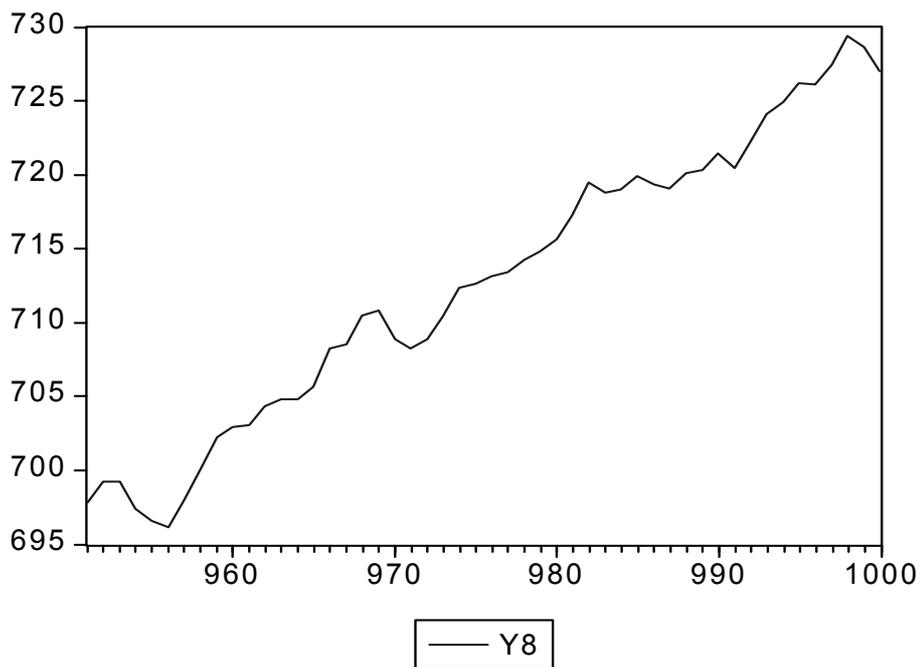
2. Создайте группу RW, состоящую из этих случайных процессов.

3.2. Построение графика многомерного временного ряда

Для построения графика многомерного временного ряда в меню **View** окна группы выберите опцию **Graph** или **Multiple Graph**. Если Вы выберите опцию **Graph**, в соответствующем окне в одних осях отобразятся графики временных рядов, например:



В случае, если Вы выберите опцию **Multiple Graph**, графики отобразятся в различных осях:



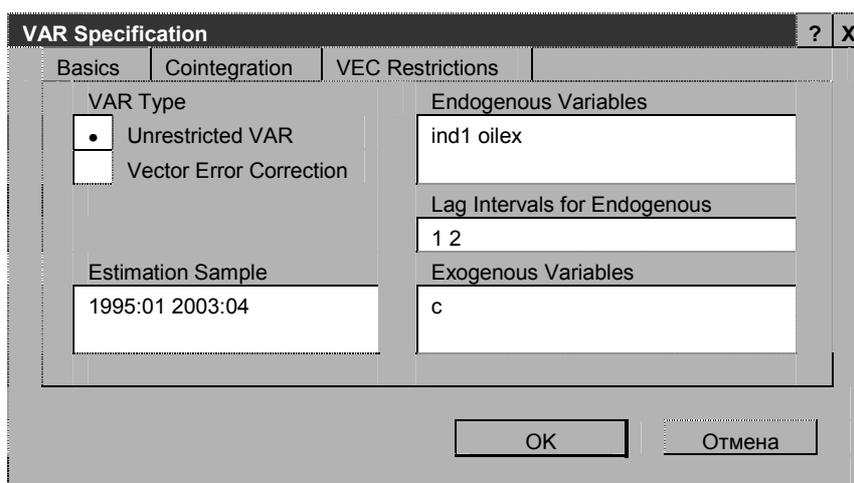
Более подробно различные типы графиков были описаны в пункте 2.3.

Задания для самостоятельной работы

1. В файле DEMO.wf постройте график многомерного временного ряда RW.
2. Сравните графические свойства случайных процессов, образующих группу RW.

3.3. Векторная авторегрессия

Чтобы оценить векторную авторегрессию в меню **Procs** окна группы выберите опцию **Make Vector Autoregression...** В появившемся окне



выберите интересующие Вас опции, т.е. введите список эндогенных переменных, входящих в векторную авторегрессию (Endogenous Variables), и экзогенных переменных (Exogenous Variables), выберите необходимое количество включаемых запаздываний эндогенных переменных, т.е. порядок векторной авторегрессии (Lag Intervals for Endogenous), а также укажите интервал, на котором Вы хотите оценить модель (Estimation Sample).

В рассмотренном примере оценивается модель векторной авторегрессии второго порядка для двух временных рядов: IND1 – индекс промышленного производства Центра экономической конъюнктуры при Правительстве РФ¹⁷ промышленности в целом, – и OILEX – индекс промышленного производства Центра экономической конъюнктуры при Правительстве РФ нефтедобывающей промышленности (на интервале с января 1995 г. по апрель 2003 г.):

Vector Autoregression Estimates		
Date: 07/22/03 Time: 19:57		
Sample: 1995:01 2003:04		
Included observations: 99		
Standard errors in () & t-statistics in []		
	IND1	OILEX
IND1(-1)	0.835523 (0.10137) [8.24255]	0.614035 (0.09675) [6.34644]

¹⁷ Базисный месяц – январь 1993 г.

IND1(-2)	-0.098109 (0.10096) [-0.97172]	-0.612509 (0.09637) [-6.35589]
OILEX(-1)	-0.362132 (0.08288) [-4.36909]	0.050454 (0.07911) [0.63776]
OILEX(-2)	0.545722 (0.08279) [6.59160]	0.981794 (0.07902) [12.4244]
C	1.127190 (3.00032) [0.37569]	-2.389568 (2.86374) [-0.83442]
R-squared	0.826194	0.917102
Adj. R-squared	0.818875	0.913611
Sum sq. resids	614.2458	559.5961
S.E. equation	2.542783	2.427032
F-statistic	112.8962	262.7452
Log likelihood	-232.6551	-227.9961
Akaike AIC	4.753102	4.659922
Schwarz SC	4.883360	4.790181
Mean dependent	62.45052	84.16891
S.D. dependent	5.974760	8.257463
Determinant Residual Covariance		33.14213
Log Likelihood (d.f. adjusted)		-458.8280
Akaike Information Criteria		9.376560
Schwarz Criteria		9.637077

В таблице приведены оценки коэффициентов модели со стандартными ошибками в (...) и t -статистиками в [...], а также стандартные МНК-статистики, характеризующие качество каждого уравнения системы. Последние 4 строки таблицы – статистики, характеризующие оцененную модель векторной авторегрессии:

Determinant Residual Covariance – определитель ковариационной матрицы случайных ошибок модели рассчитывается по формуле:

$$|\Omega| = \det\left(\frac{1}{T-p} \sum_t \varepsilon_t \varepsilon_t'\right),$$

где T – длина временного ряда, p – порядок векторной авторегрессии;

Log Likelihood (d.f. adjusted) – значения логарифмической функции максимального правдоподобия рассчитывается в предположении, что случайные ошибки модели подчиняются многомерному закону нормального распределения:

$$l = \frac{T}{2} [k(1 + \log 2\pi) + \log |\Omega|],$$

Информационные критерии Акаике и Шварца (Akaike Information Criteria и Schwarz Criteria):

$$AIC = -\frac{2l}{T} + \frac{2n}{T}$$

и

$$BIC = -\frac{2l}{T} + \frac{n \log T}{T},$$

где $n=k(d+pk)$ – число всех параметров, оцениваемых в модели векторной авторегрессии, k – число эндогенных переменных, d – число экзогенных переменных (включая константу).

Кроме приведенных выше статистик качества модели существуют и другие статистики ее качества, позволяющие выбрать наилучшую модель. Во-первых, опция **Residual Tests** меню **View** окна векторной авторегрессии позволяет получить стандартные статистики, описанные выше, характеризующие остатки модели.

Кроме того, опция **Lag Structure** меню **View** окна векторной авторегрессии дает возможность:

- ✓ вычислить характеристические авторегрессионные корни рассматриваемого многомерного временного ряда и построить их на графике (**AR Roots Table** и **AR Roots Graph**):

Roots of Characteristic Polynomial	
Endogenous variables: IND1 OILEX	
Exogenous variables: C	
Lag specification: 1 2	
Date: 07/24/03 Time: 19:07	
Root	Modulus
1.020817	1.020817
0.811875	0.811875
-0.473357 - 0.251051i	0.535811
-0.473357 + 0.251051i	0.535811
Warning: At least one root outside the unit circle. VAR does not satisfy the stability condition.	

В таблице приводятся значения характеристических корней, а в нижней части таблицы (последние две строки) указывается, удовлетворяет ли многомерный временной ряд условию устойчивости, т.е. является ли слабостационарным многомерным процессом;

- ✓ проверить гипотезы о том, что некоторые эндогенные переменные на самом деле являются экзогенными при помощи теста на причинность по Грэнджеру (**Pairwise Granger Causality Tests**).
- ✓ проверить гипотезы о значимости запаздывающих значений эндогенных переменных (**Lag Exclusion Tests**). В таблице (см. ниже) приводятся соответствующие значения статистики χ^2 для проверки совместных гипотез о значимости группы переменных на конкретном запаздывании в каждом уравнении векторной авторегрессии, а также об их значимости в векторной авторегрессии (в последнем столбце – Joint):

VAR Lag Exclusion Wald Tests			
Date: 07/24/03 Time: 18:21			
Sample: 1995:01 2003:04			
Included observations: 99			
Chi-squared test statistics for lag exclusion:			
Numbers in [] are p-values			
	IND1	OILEX	Joint
Lag 1	68.47932 [1.33E-15]	71.34502 [3.33E-16]	113.9743 [0.000000]
Lag 2	54.47254 [1.48E-12]	154.9752 [0.000000]	170.7268 [0.000000]
df	2	2	4

Как видно из таблицы, в данном примере все включенные запаздывания оказываются значимыми;

- ✓ вычислить значения различных критериев, позволяющих наилучшим образом выбрать порядок векторной авторегрессии без ограничений (**Lag Length Criteria**). Выбрав эту опцию необходимо указать количество запаздываний, для которых Вы хотите провести тестирование (в нашем случае – 13):

VAR Lag Order Selection Criteria						
Endogenous variables: IND1 OILEX						
Exogenous variables: C						
Date: 07/24/03 Time: 18:53						
Sample: 1995:01 2003:04						
Included observations: 99						
Lag	LogL	LR	FPE	AIC	SC	HQ
0	-617.0574	NA	816.8005	12.38115	12.43325	12.40223
1	-508.0099	211.5521	99.93104	10.28020	10.43651	10.34346
2	-453.6986	103.1914	36.53920	9.273973	9.534490	9.379409
3	-441.4990	22.69131	31.02164	9.109980	9.474704	9.257591
4	-431.2965	18.56853	27.41672	8.985931	9.454861	9.175715
5	-410.5607	36.90985	19.63368	8.651213	9.224351	8.883172
6	-397.0080	23.58158	16.23819	8.460160	9.137505	8.734294
7	-386.4080	18.02011	14.25299	8.328159	9.109710	8.644467
8	-382.2395	6.919574	14.23500	8.324791	9.210549	8.683273
9	-371.9166	16.72324	12.57770	8.198331	9.188296	8.598988
10	-338.1122	53.41095	6.953008	7.602243	8.696415	8.045074
11	-322.3627	24.25423	5.519332	7.367253	8.565631	7.852258
12	-312.2120	15.22602	4.904419	7.244240	8.546825	7.771419
13	-284.6735	40.20621*	3.080743*	6.773470*	8.180261*	7.342824*
14	-280.7090	5.629569	3.103997	6.774180	8.285179	7.385708

* indicates lag order selected by the criterion
 LR: sequential modified LR test statistic (each test at 5% level)
 FPE: Final prediction error
 AIC: Akaike information criterion
 SC: Schwarz information criterion
 HQ: Hannan-Quinn information criterion

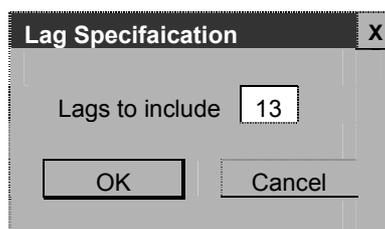
Для каждого числа запаздываний приводится 6 различных статистик, позволяющих выбрать лучшую модель, которая указывается звездочкой «*».

3.4. Тест на причинность по Грэнджеру

Для проведения теста на причинность по Грэнджеру, т.е. проверки того, что временной ряд x_t не является причиной по Грэнджеру для временного ряда y_t , в EViews есть специальная опция: в меню **View** окна группы временных рядов необходимо выбрать опцию **Granger Causality...** Например, Вы хотите провести данный тест для двух временных рядов: IND1 – индекс промышленного производства Центра экономической конъюнктуры при Правительстве РФ¹⁸ промышленности в целом, – и OILEX – индекс промышленного производства Центра экономической конъюнктуры при Правительстве РФ нефтедобывающей промышленности (на интервале с января 1995 г. по апрель 2003 г.). Для этого, во-первых, необходимо создать группу, состоящую из этих рядов (см. п. 3.1). Затем

¹⁸ Базисный месяц – январь 1993 г.

воспользуйтесь опцией **View/Granger Causality...** окна созданной группы. Прежде чем получить необходимый результат, в окне



введите необходимое количество включаемых в уравнений лагов (в данном случае их число равно 13). В результате Вы получите следующую таблицу:

Pairwise Granger Causality Tests
 Date: 07/18/03 Time: 18:13
 Sample: 1995:01 2003:04
 Lags: 13

Null Hypothesis:	Obs	F-Statistic	Probability
OILEX does not Granger Cause IND1	100	7.53840	3.6E-09
IND1 does not Granger Cause OILEX		1.32699	0.21759

Интерпретация результатов теста проста: мы можем отвергнуть на 5%-м уровне значимости нулевую гипотезу о том, что ряд OILEX не является причиной по Грэнджеру ряда IND1, и не можем отвергнуть на данном уровне значимости гипотезу о том, что ряд IND1 не является причиной по Грэнджеру для ряда OILEX. Таким образом, встроенная в EViews процедура позволяет проверять соответствующую гипотезу в обе стороны, т.е. оценить регрессии следующего вида:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \varepsilon_t$$

и

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + u_t$$

и проверить нулевую гипотезу о том, что

$$\beta_1 = \dots = \beta_p = 0$$

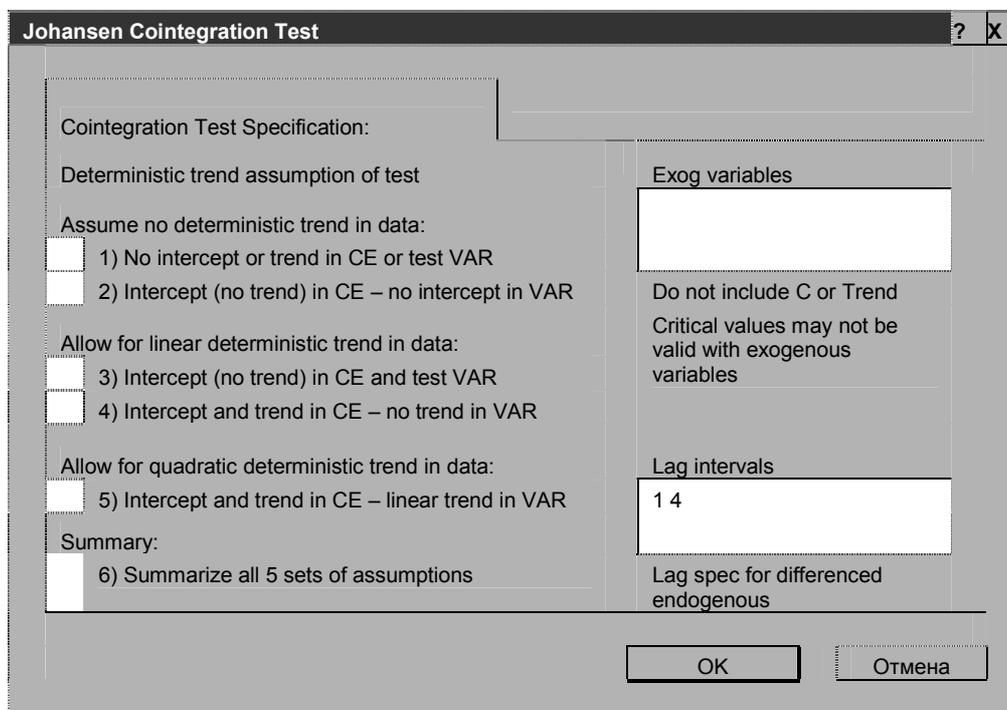
для каждого из рассмотренных уравнений.

Отметим, что если бы группа содержала больше двух временных рядов, то тест был бы проведен для каждой пары рядов, входящих в нее.

3.5. Тест Йохансена на коинтеграцию

Тест Йохансена на коинтеграцию позволяет выявить наличие стационарных линейных комбинаций временных рядов, являющихся интегрированными первого порядка и является одним из методов оценки систем, использующий метод максимального правдоподобия применительно к векторным авторегрессионным моделям. Отметим, что основными предположениями данного теста являются допущения, что переменные, входящие в векторную авторегрессионную модель, являются интегрированными процессами первого порядка и ошибки независимо и нормально распределены.

Чтобы провести тест Йохансена на коинтеграцию, выберите **View/Cointegration Test...** в окне группы или в окне, появляющемся при оценивании векторной авторегрессии. В первом случае появится окно, в котором в соответствии с тестом Йохансена представлены различные опции, касающиеся спецификации многомерного временного ряда и коинтеграционного соотношения:



Поскольку асимптотическое распределение соответствующей тестовой LR -статистики зависит от спецификации коинтеграционного соотношения и векторной авторегрессии, в EViews предусмотрены следующие опции:

- 1) No intercept or trend in CE or test VAR – и константа и тренд отсутствуют и в коинтеграционном соотношении и в векторной авторегрессии;
- 2) Intercept (no trend) in CE – no intercept in VAR – в коинтеграционное соотношение включена константа, но не тренд, а в векторной авторегрессии отсутствуют и константа и тренд;
- 3) Intercept (no trend) in CE and test VAR – только свободный член включен и в коинтеграционное соотношение и в векторную авторегрессию;
- 4) Intercept and trend in CE – no trend in VAR – в коинтеграционное соотношение включены константа и тренд, а в векторную авторегрессию только константа;
- 5) Intercept and trend in CE – linear trend in VAR – и тренд и константа включены и в коинтеграционное соотношение и в векторную авторегрессию.

Таким образом, выбор первой или второй опций предполагает отсутствие тренда в данных и наличие нулевого (в первом случае) и ненулевого (во втором) среднего; третьей или четвертой опции – наличие только линейного стохастического тренда во всех временных рядах (третий случай) или стохастического (во всех) и детерминированного (в некоторых) тренда (четвертый случай); а пятой опции – наличие квадратичного тренда в данных. Если Вы точно

не уверены в том, какова спецификация имеющихся у Вас данных, можно воспользоваться опцией, позволяющей получить таблицу со сравнительными характеристиками всех пяти возможных спецификаций – 6) Summarize all 5 sets of assumptions. Если продолжить рассмотрение примера из предыдущего пункта, то сводная таблица будет выглядеть следующим образом:

Date: 07/19/03 Time: 17:46
Sample: 1995:01 2003:05
Included observations: 100
Series: IND1 OILEX
Lags interval: 1 to 13

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or No. of CEs	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Selected (5% level) Number of Cointegrating Relations by Model (columns)					
Trace	0	0	0	0	0
Max-Eig	0	0	0	0	0
Log Likelihood by Rank (rows) and Model (columns)					
0	-288.6771	-288.6771	-286.5035	-286.5035	-281.4094
1	-285.4032	-282.5282	-281.8382	-280.3765	-276.4220
2	-283.7022	-280.7090	-280.7090	-276.4083	-276.4083
Akaike Information Criteria by Rank (rows) and Model (columns)					
0	6.813542	6.813542	6.810070	6.810070	6.748187
1	6.828063	6.790565	6.796763	6.787529	6.728440*
2	6.874044	6.854180	6.854180	6.808166	6.808166
Schwarz Criteria by Rank (rows) and Model (columns)					
0	8.168231*	8.168231*	8.216862	8.216862	8.207083
1	8.286958	8.275512	8.307762	8.324580	8.291543
2	8.437146	8.469385	8.469385	8.475475	8.475475

Как видно из таблицы, полученные результаты, с одной стороны, противоречат друг другу: по критерию Шварца лучшими оказываются модели первого и второго типа, а по критерию Акаике – пятого. Но с другой стороны, результаты вполне согласуются друг с другом: обе тестовые статистики (и λ_{trace} – Trace – и λ_{max} – Max-Eig) говорят об отсутствии коинтеграции между данными временными рядами на 5%-м уровне значимости независимо от типа модели. Действительно, если, например, оценить вторую модель, то полученный результат будет свидетельствовать об отсутствии коинтеграции:

Date: 07/22/03 Time: 19:24
Sample(adjusted): 1995:01 2003:04
Included observations: 100 after adjusting endpoints
Trend assumption: Linear deterministic trend
Series: IND1 OILEX
Lags interval (in first differences): 1 to 12

Unrestricted Cointegration Rank Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.087027	10.44561	15.41	20.04
At most 1	0.013317	1.340697	3.76	6.65

(**) denotes rejection of the hypothesis at the 5%(1%) level
Trace test indicates no cointegration at both 5% and 1% levels

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.087027	9.104913	14.07	18.63
At most 1	0.013317	1.340697	3.76	6.65

*** denotes rejection of the hypothesis at the 5%(1%) level Max-eigenvalue test indicates no cointegration at both 5% and 1% levels		
Unrestricted Cointegrating Coefficients (normalized by b*S11*b=1):		
IND1	OILEX	
0.277611	0.015040	
-0.293166	0.451230	
Unrestricted Adjustment Coefficients (alpha):		
D(IND1)	-0.221737	0.088506
D(OILEX)	0.168112	0.096102
1 Cointegrating Equation(s): Log likelihood -285.3438		
Normalized cointegrating coefficients (std.err. in parentheses)		
IND1	OILEX	
1.000000	0.054175	
	(0.43561)	
Adjustment coefficients (std.err. in parentheses)		
D(IND1)	-0.061557	
	(0.03391)	
D(OILEX)	0.046670	
	(0.03211)	

Интерпретация полученных результатов:

Unrestricted Cointegration Rank Test – в данной части таблицы приводятся результаты тестов на наличие коинтеграционных соотношений. Для определения количества коинтегрирующих векторов Йохансен предлагает две статистики:

$$\lambda_{trace} = -T \sum_{i=r+1}^n \log(1 - \lambda_i),$$

которая является тестовой статистикой для следующих нулевой и альтернативной гипотез

$$H_0 : rank \leq r, \quad H_A : r = n,$$

где n – количество случайных процессов в группе¹⁹, и

$$\lambda_{max} = -T \log(1 - \lambda_{r+1})$$

– тестовая статистика для гипотез

$$H_0 : rank \leq r, \quad H_A : rank \leq r + 1.$$

В таблице приводятся как значения данных тестовых статистик (Trace Statistic и Max-Eigen Statistic), так и их 5%-е и 1%-е критические значения. Кроме того, в таблице непосредственно указывается количество коинтеграционных соотношений на конкретном уровне значимости: Trace test indicates no cointegration at both 5% and 1% levels и Max-eigenvalue test indicates no cointegration at both 5% and 1% levels.

Вторая часть таблицы содержит информацию о коинтеграционных соотношениях. Прежде, чем приступить к описанию второй части таблицы, кратко остановимся на теории многомерных временных рядов применительно к понятию коинтеграции.

¹⁹ В EViews можно оценивать коинтеграционные модели не более чем для 10 временных рядов, т.е. $n \leq 10$.

Любая модель векторной авторегрессии порядка p может быть записана в виде:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} \dots + A_p Y_{t-p} + U_t.$$

С другой стороны, эта модель может быть записана в виде модели коррекции ошибки (*Error Correction Model*):

$$\Delta Y_t = B_1 Y_{t-1} + B_2 \Delta Y_{t-1} \dots + B_p \Delta Y_{t-p+1} + U_t,$$

где $B_1 = -I + \sum_{i=1}^p A_i$, $B_j = -\sum_{i=j}^p A_i$, $j = \overline{2, p}$. Т.к. $\Delta Y_t, \dots, \Delta Y_{t-p+1}$ – стационарные процессы, в то время как Y_{t-1} – интегрированный первого порядка. Поэтому для того чтобы уравнение было состоятельным, матрица B_1 не должна быть полного ранга. Обозначим ее ранг r . Пусть

$$B_1 = \alpha \beta^T,$$

где α – матрица размера $n \times r$ и β^T – матрица размера $r \times n$. Тогда $\beta^T Y_{t-1}$ представляет собой вектор из r коинтеграционных соотношений, β^T – матрица коинтеграционных векторов, а α можно интерпретировать как матрицу коэффициентов при корректирующих членах в модели коррекции ошибок.

Тогда в той части таблицы, которую условно можно назвать «Unrestricted Cointegrating Coefficients (normalized by b'S11*b=I)», приведены оценки матрицы коинтеграционных векторов β^T . Далее, чуть ниже, приводятся нормированные коинтеграционные векторы для моделей, содержащих r коинтеграционных соотношений при $1 \leq r \leq n-1$ (в нашем случае рассматривается лишь одно коинтеграционное соотношение – 1 Cointegrating Equation(s): Normalized cointegrating coefficients (std.err. in parentheses)). При этом в модели с одним коинтеграционным соотношением нормируется первый коинтеграционный вектор, с двумя – первый и второй, и т.д. В скобках под оценками коэффициентов в каждом из случаев приведены соответствующие стандартные ошибки.

Отметим также, что при выборе количества лагов необходимо учитывать тот факт, что Вы указываете количество запаздывающих разностей в модели коррекции ошибки, а не порядок исходной векторной авторегрессии, т.е. если Вы предполагаете, что векторная авторегрессия имеет порядок p , то необходимо включать запаздывания от 1 до $p-1$, т.е. в соответствующей опции появившегося окна написать

$$1 \quad p-1.$$

Если Вы хотите оценить модель, в которой предполагается, что векторная авторегрессия имеет порядок 1, то необходимо указать

$$0 \quad 0.$$

Задания для самостоятельной работы к пунктам 3.3-3.5

1. В файле DEMO.wf создайте группы RW1, RW2, RW3, RW4, состоящие из следующих пар случайных процессов: gw и gw1, gw и gw2, gw и gw3, gw и gw4.

2. Для данных групп временных рядов проверьте наличие коинтеграции, используя процедуры Энга-Гренджера и Йохансена, проверьте наличие причинности по Гренджеру и оцените модели векторной авторегрессии.
3. Какие результаты о наличии коинтеграции и причинности по Гренджеру можно было предположить, исходя из того, как были построены случайные процессы (см. задание к пункту 3.1)? Согласуются ли полученные результаты с предполагаемыми?