

Лабораторная работа № 5

«Временные ряды»

Целью лабораторной работы является изучение стационарных и нестационарных временных рядов, определения типа процесса, которому относится временной ряд, построение качественной модели *ARMA* в EViews.

- умение работать в EViews;
- умение определять тип процесса;
- умение строить качественные модели *ARMA*;
- умение проводить диагностику построенной *ARMA*;
- умение проводить корректировку построенной модели.

Часть 1

Определение типа процесса

Теоретические сведения:

Временной ряд [time-series] - ряд последовательных значений, характеризующих изменение показателя во времени.

Временные ряды могут быть стационарными и нестационарными. Временной ряд называется стационарным, если его вероятностные характеристики (математическое ожидание, дисперсия) постоянны. Временной ряд называется нестационарным, если хотя бы одна из вероятностных характеристик непостоянна. На практике часто встречаются нестационарные временные ряды.

Два разных типа нестационарных по отношению к среднему временных рядов могут быть приведены к стационарному виду с помощью взятия последовательных разностей. Это ряды с детерминированным трендом (TS – trend stationary) и ряды, имеющие стохастический процесс (DS – difference stationary).

Принципиальное различие между этими двумя типами рядов выражается в том, что TS ряд можно привести к стационарному виду с помощью выделения линейного тренда, тогда как вычитание детерминированной составляющей из DS ряда оставляет его нестационарным.

Выделим типы процессов, к которым временные ряды часто относятся на практике:

1. TS (trend stationary) – детерминированный тренд.
2. DS (difference stationary) - стохастический процесс. Выделяют несколько видов DS процесса, а именно DS I (0) - ряд сразу является стационарным, I (1) - ряд становится стационарным на 1-ых разностях, DS I(2) – ряд становится стационарным на 2-ых разностях.
3. TS+DS – стохастический тренд.

Вид уравнения в зависимости от типа процесса:

- TS+DS: $\Delta^2 y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + bt + \sum_{i=1}^n \beta_i \Delta^2 y_{t-i}$
- TS: $\Delta y_t = \alpha_0 + bt + \sum_{i=1}^n \beta_i \Delta^2 y_{t-i}$
- DS (I_1): $\Delta^2 y_t = \alpha_1 \Delta y_{t-1} + \sum_{i=1}^n \beta_i \Delta^2 y_{t-i}$
- DS (I_2): $\Delta^3 y_t = \alpha_1 \Delta^2 y_{t-1} + \sum_{i=1}^n \beta_i \Delta^3 y_{t-i}$
- DS (I_0): $\Delta y_t = \alpha_1 y_{t-1} + \sum_{i=1}^n \beta_i \Delta y_{t-i}$

Цель: определить к какому типу процесса относится временной ряд для возможности построения в дальнейшем качественного прогноза.

Задачи для достижения цели:

1. На основе анализа графика исходного ряда сделать предположение о типе процесса, к которому относится исследуемый временной ряд.
2. На основе анализа коррелограмм автокорреляционной функции (ACF/ АКФ) и частной автокорреляционной функции (PACF/ ЧКФ) для исходного ряда, его 1-ых и 2-ых разностей сделать предварительный вывод о типе процесса.
3. На основе процедуры Доладо, Дженкинса и Сосвилло-Ривера с применением расширенного теста Дики – Фуллера определить к какому типу процесса относится временной ряд.
4. Сделать заключение о типе процесса временного ряда.

Практическая реализация определения типа процесса, к которому относится временной ряд в пакете EViews.

Приложение EViews не имеет модульный системы, однако содержит так называемое окно рабочего файла, где можно сохранять ряд объектов.

Объектная структура рабочего окна позволяет работать одновременно с различными типами данных. Управление объектами осуществляется с помощью процедур (procs), которые в свою очередь могут сами создавать новые объекты. Каждый объект содержит конкретный вид информации: ряд данных, коэффициенты, графики и диаграммы, модели, результаты вычислений. Кроме того, EViews имеет командную строку, куда вводятся определенные команды, позволяющие производить статистический анализ данных. Команды могут быть сохранены в отдельном файле, что позволяет запомнить ретроспективу выполняемых действий.

Для создания рабочего файла выбираем команду меню File-New-Workfile (рис.1).

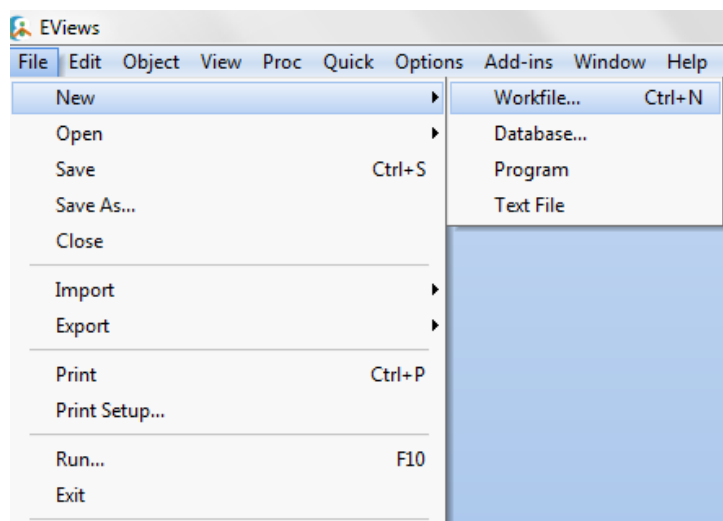


Рисунок 1 – Создание рабочего файла

В появившемся диалогом окне *Workfile Create* необходимо определить структуру типа данных (рис.2): неструктурированные (данные без дат) – кросс-секции (*Unstructured/Undated*), временные ряды (*Dated – regular frequency*), сбалансированные панельные данные, которые являются объединением кросс-секций и временных рядов (*Balanced Panel*).

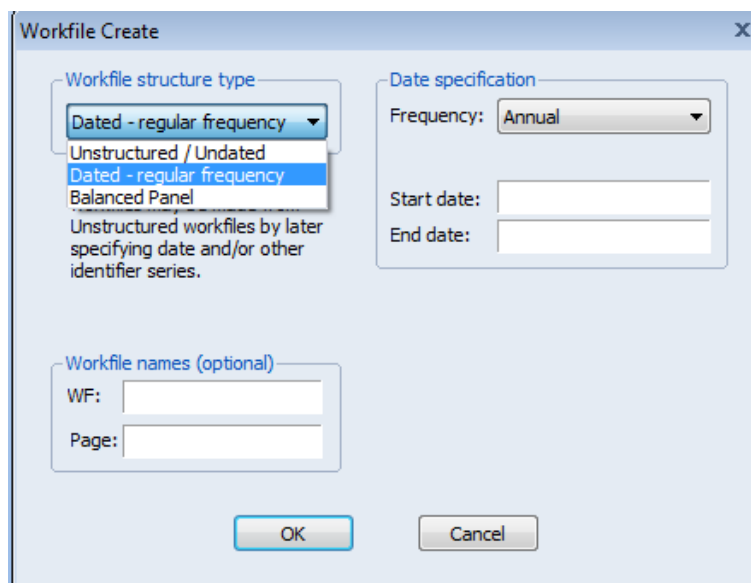


Рисунок 2 – Определение структуры данных

Следующим шагом является определение типа данных, в пакете допускается работать с четырнадцатью типами данных (рис.3), например:

- Мульти-годовые (Multi-year)
- Годовые (Annual) – годы 20 века идентифицируются по последним двум цифрам (98 эквивалентно 1998), а для годов 21 века необходима полная идентификация (например, 2015)

- Полугодовые (Semi-annual) – 1998:1, 2001:2 (формат – год и номер полугодия)
- Квартальные (Quarterly) – 1998:1, 2005:4 (формат – год и номер квартала)
- Ежемесячные (Monthly) – 1956:2, 1990:12 (формат – год и номер месяца)
- Двумесечные (Bimonthly)
- Декадные (Ten-day)
- Недельные (Weekly) и дневные (5/7 day weeks) – допускаются формата Месяц/День/Год (по умолчанию) – 8:10:98 (август, 10, 1998г.)

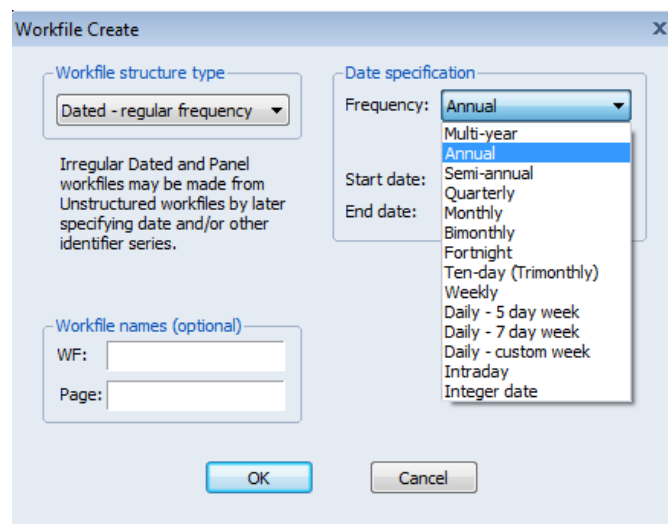


Рисунок 3 – Определение типа данных

Важным является указание начальной (start date) и конечной (end date) даты наблюдений.

Имя рабочего файла необходимо задать в рамке *Workfile names (optional)* в поле *WF* (рис.3) и допускается только латинское написание.

После ввода всех необходимых сведений создаем рабочий файл (кнопка *OK*). Рабочее окно содержит вектор коэффициентов *c* и ряд *RESID* (*остатки*). Для импорта данных необходимо выбрать в рабочем файле опцию *Procs/Import* (рис.4).

При формировании ряда данных непосредственно в пакете EViews необходимо создать объект вида ряд. Для этого в меню рабочего файла выбирают команду *Object/New object*. В появившемся диалоговом окне *New Object* выбирают объект типа – ряд (*Series*), в рамке *Name of Object* задают имя (рис.5).

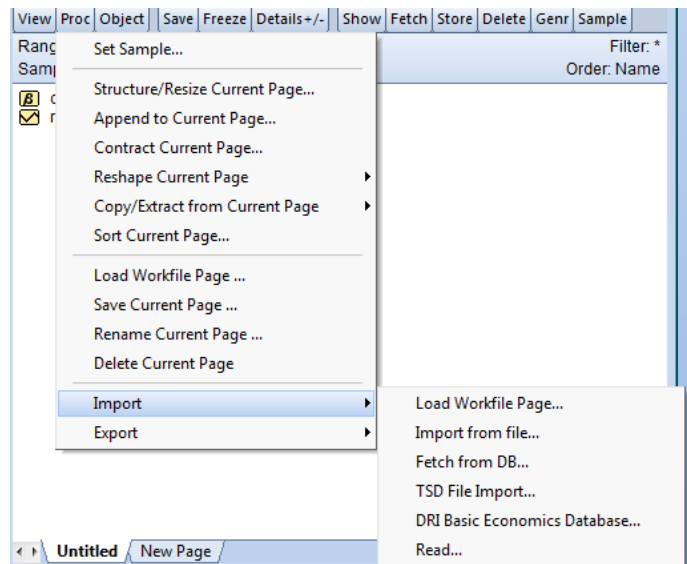


Рисунок 4 – Импорт данных

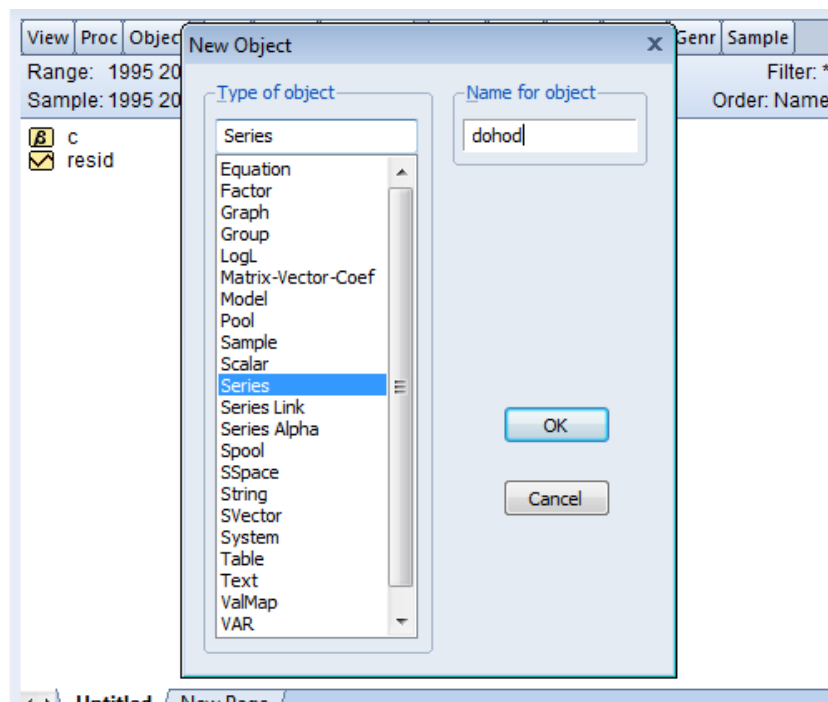


Рисунок 5 – Создание нового ряда

Для непосредственного ввода данных в окне редактирования ряда *Series*: необходимо выбрать команду Edit+/- . Появится строка ввода данных. Принцип ввода информации как в любой электронной таблице (рис. 6).

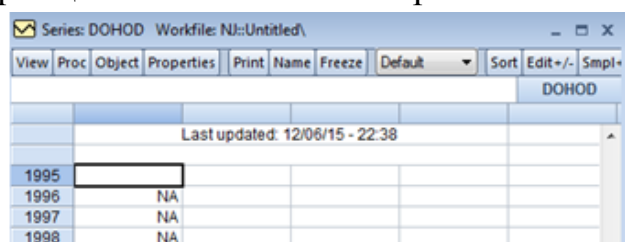
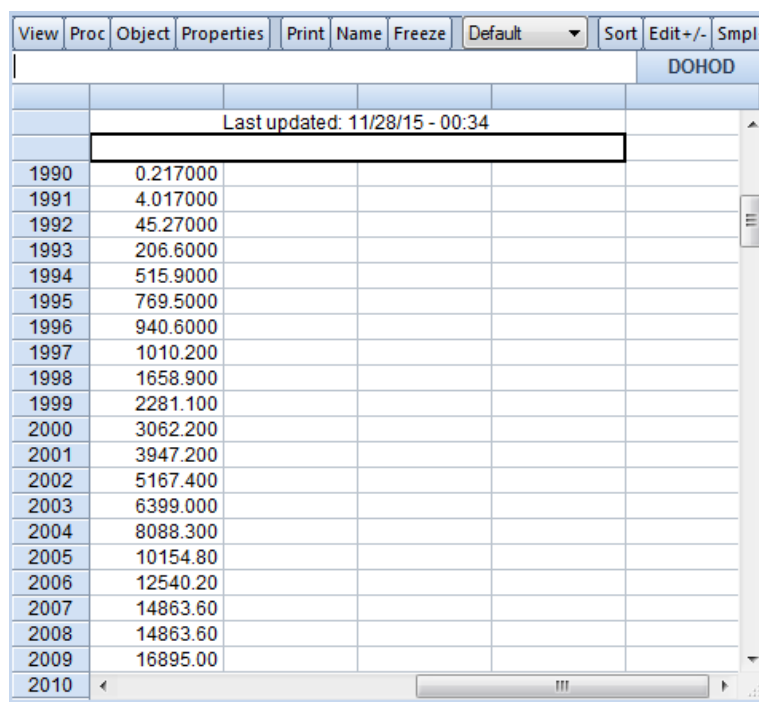


Рисунок 6 – Формирование ряда данных

После формирования ряда данных (рис.7) переходим к определению типа процесса временного ряда.



Year	Value
1990	0.217000
1991	4.017000
1992	45.27000
1993	206.6000
1994	515.9000
1995	769.5000
1996	940.6000
1997	1010.200
1998	1658.900
1999	2281.100
2000	3062.200
2001	3947.200
2002	5167.400
2003	6399.000
2004	8088.300
2005	10154.80
2006	12540.20
2007	14863.60
2008	14863.60
2009	16895.00
2010	

Рисунок 7 – Временной ряд «Доход» за период с 1990 по 2015

1. Анализ графика исходного временного ряда.

Для построения графика исходного временного ряда воспользуемся командой *View-Graphs* (рис. 8).

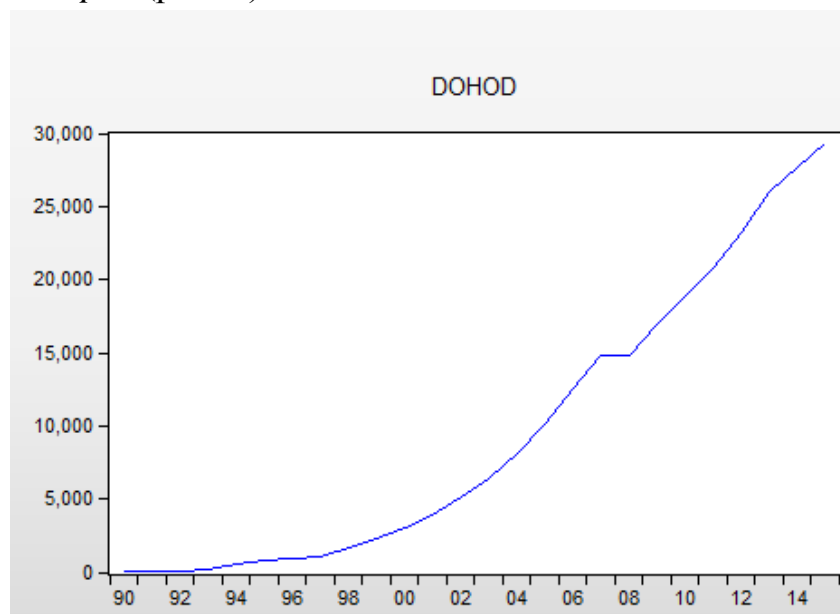


Рисунок 8 – Графическое представление исходного ряда

Если график резко возрастает, то можно сделать предварительный вывод о том, что ряд детерминированный. Если график «колеблется» - стационарный ряд.

Вывод: По графику (рис.8) ряда «Доход» можно сделать предварительный вывод, что ряд является детерминированным.

2. Анализ коррелограмм автокорреляционной функции (ACF/ АКФ) и частной автокорреляционной функции (PACF/ ЧКФ) для исходного ряда, его 1-ых и 2-ых разностей.

Для построения коррелограмм АКФ и ЧКФ воспользуемся командой *View-Correlogram* (рис.9).

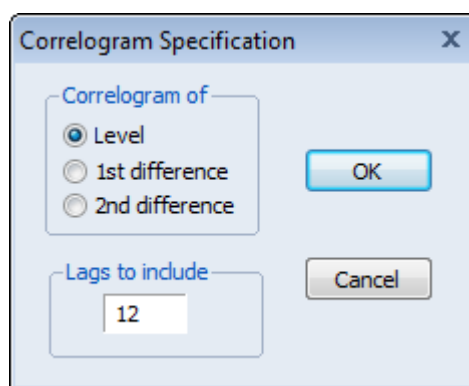


Рисунок 9 – Построение коррелограмм АКФ и ЧКФ

Необходимо построить коррелограммы исходного ряда, 1-ой и 2-ой разностей (рис.10-12).

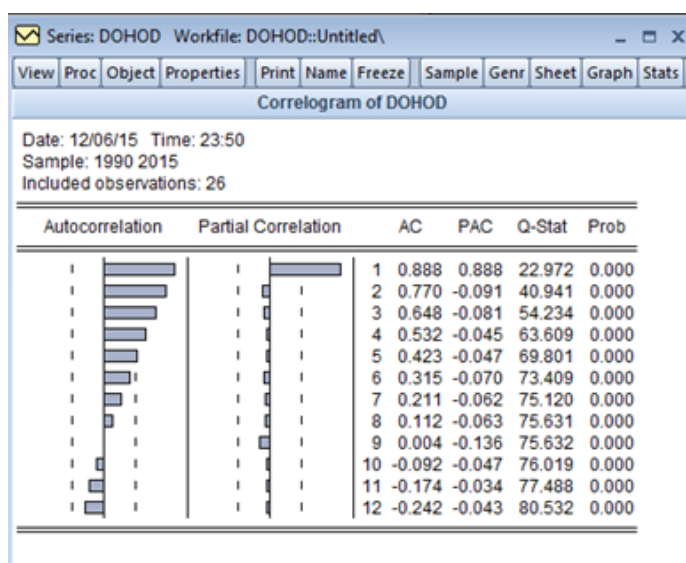


Рисунок 10 – Коррелограмма исходного ряда

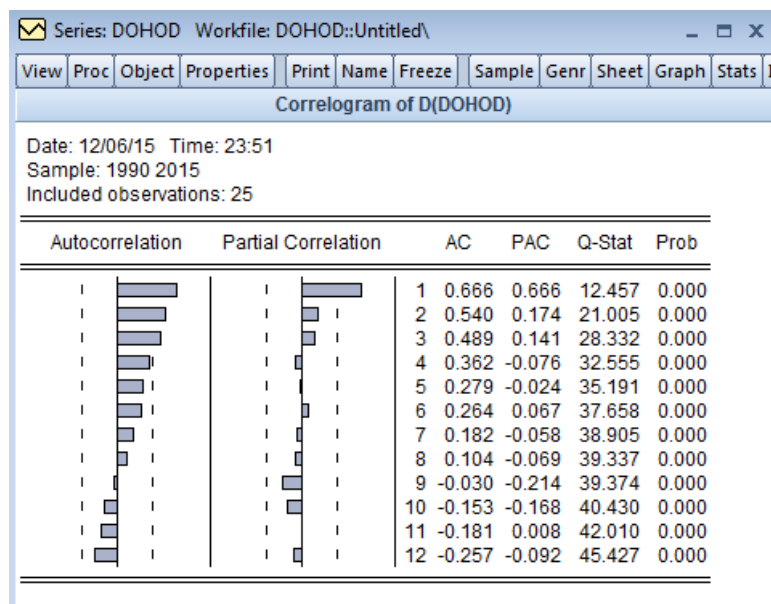


Рисунок 11 – Коррелограмма 1-ой разности

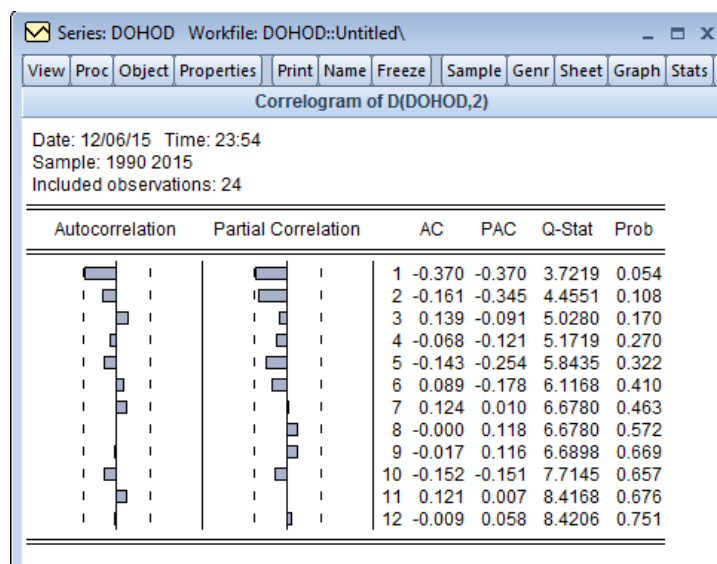


Рисунок 12 – Коррелограмма 2-ой разности

Анализ АКФ и ЧКФ

- Если коррелограмма исходного ряда имеет медленно убывающую либо не убывающую по модулю АКФ при это ЧКФ имеет первый значимый лаг почти единицу, то процесс является нестационарным и возможно содержит детерминированный тренд.
- Если процесс TS или $DS(I_1)$, то уже на первых разностях коррелограммы будут резко убывающими и иметь значимыми только первые несколько лагов ЧКФ.
- Если процесс $TS + DS$ или $DS(I_2)$, то коррелограммы для первых разностей не убывают, а убывают лишь на вторых разностях.

- Если АКФ и ЧКФ сразу убывают на исходном ряде, то этот процесс является $DS(I_0)$.

Сложности выводов по коррелограмме связаны со случаями коротких временных рядов, где наблюдается быстрое убывание АКФ, а сделать уверенный вывод о стационарности исходного временного ряда невозможно. Также неоднозначна ситуация, даже для ряда с медленно скользящей средней, поскольку это может свидетельствовать как о не стационарности, так и о стационарном авторегрессионном процессе с близкими к 1 значениями коэффициента авторегрессии.

Стационарность временного ряда можно проверить, исходя из ее определения. Для этого ряд нужно разделить на несколько частей и проверить гипотезы о равенстве дисперсий и мат. ожиданий этих частей. При положительном результате проверки соответствующих гипотез, возможно дополнительная проверка равенства АКФ.

3. Применение расширенного теста Дики – Фуллера для определения типа процесса.

В EViews стационарность временного ряда можно проверить критерием Дикки-Фуллера. Этот критерий также называется тестом на наличие единичного корня. Суть заключается в следующем: делается предположение о типе процесса, породившего данный ряд, строится вспомогательная модель и проверяются гипотезы о коэффициентах этой модели, после чего можно сделать вывод о стационарности ряда.

Для проведения теста Дикки-Фуллера в EViews воспользуемся командой *View-Unit Root Test* (рис.13).

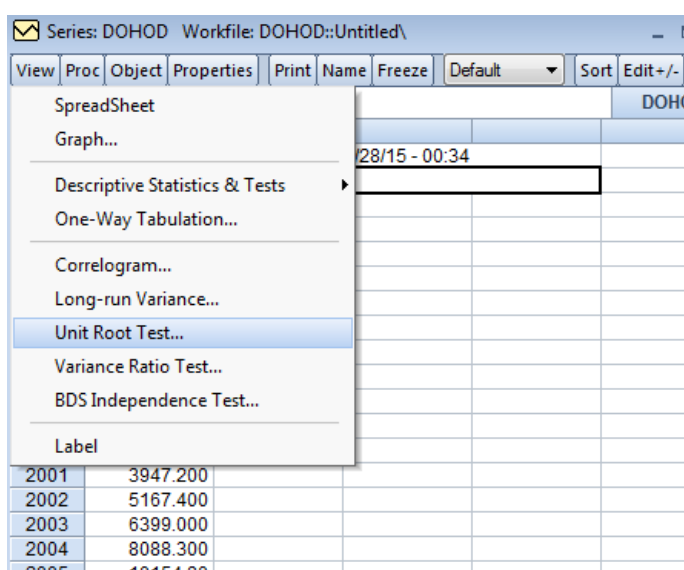


Рисунок 13 – Вызов теста Дикки-Фуллера

Окно теста выглядит следующим образом:

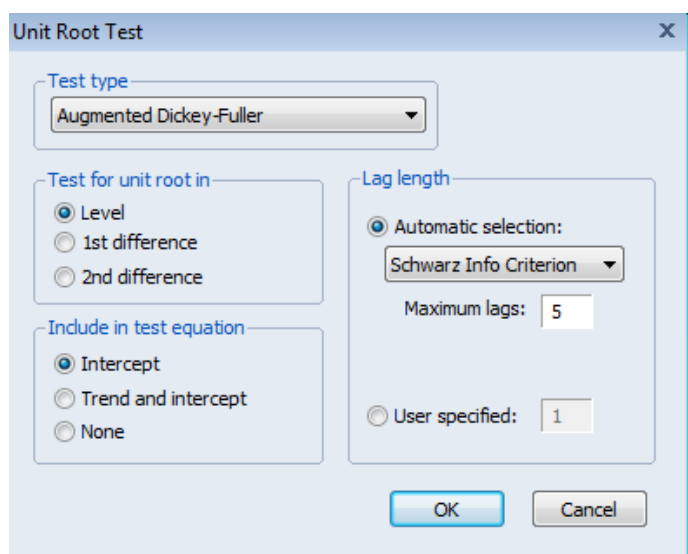


Рисунок 14 – Параметры теста Дикки-Фуллера

Рассмотрим подробнее окно с параметрами теста Дикки-Фуллера.

Во вкладке *Test for unit root in* отвечает за выбор ряда, который необходимо проанализировать в тесте (*level* – исходный ряд, *1st difference* – ряд первой разности, *2nd* – ряд второй разности).

Вкладка *Include in test equation* включает три составляющие *None*, *Intercept* и *Trend and intercept*.

Intercept означает, что проверяется авторегрессионная модель со свободным членом: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$, которую преобразовали в $\Delta y_t = \alpha_0 + (\alpha_1 - 1)y_{t-1} + \varepsilon_t$, где $\Delta y_t = y_t - y_{t-1}$. Если гипотеза верна и $\alpha_1 = 1$, т.е. коэффициент при y_{t-1} незначим, а α_0 значим, то исходный ряд является случайным блужданием с дрейфом $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ и его можно свести к стационарному взятием одной разности $\Delta y_t = \alpha_0 + \varepsilon_t$, что можно проверить, повторив тест для первых разностей (*1st difference*).

Trend and intercept предполагает тестирование модели $y_t = \alpha_0 + \alpha_1 y_{t-1} + \gamma t + \varepsilon_t$ в которую добавлен линейный тренд. Эта модель аналогично преобразуется в $\Delta y_t = \alpha_0 + (\alpha_1 - 1)y_{t-1} + \gamma t + \varepsilon_t$ и может дать дополнительную информацию о значимости переменной t , т.е. о наличии линейного тренда. Соответственно могут быть получены выводы о стационарности ряда (в разностях и трендово), нестационарности как за счет тренда, так и одновременно за счет детерминированного и стохастического трендов.

None тестирует модель случайного блуждания без дрейфа, т.е. $y_t = \alpha_1 y_{t-1} + \varepsilon_t$.

Для определения типа процесса необходимо провести 5 различных тестов Дикки-Фуллера.

1 тест

Во вкладке *Test for unit root in* выбираем *1st difference*, а в *Include in test equation* - *Trend and intercept*. Результаты теста 1 представлены на рисунке 15.

Null Hypothesis: D(DOHOD) has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 0 (Automatic - based on SIC, maxlag=5)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-4.014201	0.0223
Test critical values:	1% level		-4.394309	
	5% level		-3.612199	
	10% level		-3.243079	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DOHOD,2)				
Method: Least Squares				
Date: 12/10/15 Time: 12:05				
Sample (adjusted): 1992 2015				
Included observations: 24 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(DOHOD(-1))	-0.929796	0.231627	-4.014201	0.0006
C	-132.0337	262.6838	-0.502633	0.6205
@TREND("1990")	94.29386	30.44750	3.096932	0.0055
R-squared	0.437975	Mean dependent var		65.67500
Adjusted R-squared	0.384449	S.D. dependent var		715.3922
S.E. of regression	561.2758	Akaike info criterion		15.61477
Sum squared resid	6615642.	Schwarz criterion		15.76203
Log likelihood	-184.3772	Hannan-Quinn criter.		15.65384
F-statistic	8.182440	Durbin-Watson stat		1.884163
Prob(F-statistic)	0.002357			

Рисунок 15 – Результат 1 теста Дикки-Фуллера

Анализ результатов теста Дикки-Фуллера на примере теста 1.

Тест считается выполненным, если

1. Выполняется проверка на единичный корень.

Выдвигаем нулевую гипотезу: имеется единичный корень; а односторонняя альтернативная гипотеза: корень меньше единицы.

Существуют таблицы критических значений ADF статистики, которые зависят от того, какого вида модель оценивается. Мак-Киннон (James G. MacKinnon канадский эконометрист, род. в 1951 г.) нашел приближенную формулу расчета критических значений, которая и используется в Eviews. ADF статистика по смыслу (и формуле) представляет собой статистику Стьюдента, но имеет несколько другое распределение, поэтому используются другие критические значения. Если $t - statistica$ (отвечающему за процесс единичного корня) лежит левее критического значения MacKinnon, то есть $t_{расч} \leq t_{табл.}$, то нулевая гипотеза отвергается (рисунок 16).

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.014201	0.0223
Test critical values:		
1% level	-4.394309	
5% level	-3.612199	
10% level	-3.243079	

*MacKinnon (1996) one-sided p-values.

Рисунок 16 – Проверка на единичный корень

Отвержение нулевой гипотезы говорит об отсутствии необходимости брать разности для приведения ряда к стационарному виду.

2. Проверка на статистическую значимость F-statistic (рис.15). F-statistic является значимой, если соответствующая вероятность p-value ($Prob(F - statistic) < 0,05$).

3. Все остальные коэффициенты статистически значимы, то есть соотносящиеся им уровни доверительной вероятности p-value $< 0,05$.

Если не все лаговые переменные статистически значимы, то следует изменить длину *max* – ого лага запаздывания в специальном тесте на единичный корень на меньшую длину.

4. Статистика *Durbin-Watson Stat.* (рис.15) должна лежать в границах 1,6 и 2,4.

5. Коэффициент детерминации (*R-squared*) должен стремиться к 1.

2 тест

Во вкладке *Test for unit root in* выбираем *level*, а в *Include in test equation - Trend and intercept*.

Null Hypothesis: DOHOD has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic - based on SIC, maxlag=5)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.392596	0.8381
Test critical values:				
1% level			-4.374307	
5% level			-3.603202	
10% level			-3.238054	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation Dependent Variable: D(DOHOD) Method: Least Squares Date: 12/10/15 Time: 14:15 Sample (adjusted): 1991 2015 Included observations: 25 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DOHOD(-1)	-0.052349	0.037591	-1.392596	0.1777
C	-485.3778	326.3462	-1.487309	0.1511
@TREND("1990")	163.0425	46.53854	3.503387	0.0020

R-squared	0.695219	Mean dependent var	1173.387
Adjusted R-squared	0.667511	S.D. dependent var	913.8611
S.E. of regression	526.9490	Akaike info criterion	15.48425
Sum squared resid	6108856.	Schwarz criterion	15.63052
Log likelihood	-190.5531	Hannan-Quinn criter.	15.52482
F-statistic	25.09146	Durbin-Watson stat	1.829525
Prob(F-statistic)	0.000002		

Рисунок 17 – Результат 2 теста Дикки-Фуллера

3 тест

Во вкладке *Test for unit root in* выбираем *2nd difference*, а в *Include in test equation* - *none*.

Null Hypothesis: D(DOHOD,2) has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=5)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.807998	0.0000
Test critical values:		
1% level	-2.669359	
5% level	-1.956406	
10% level	-1.608495	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DOHOD,3)
Method: Least Squares
Date: 12/10/15 Time: 15:10
Sample (adjusted): 1993 2015
Included observations: 23 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(DOHOD(-1),2)	-1.358755	0.199582	-6.807998	0.0000
R-squared	0.678087	Mean dependent var	-12.35448	
Adjusted R-squared	0.678087	S.D. dependent var	1209.055	
S.E. of regression	685.9864	Akaike info criterion	15.94210	
Sum squared resid	10352701	Schwarz criterion	15.99147	
Log likelihood	-182.3341	Hannan-Quinn criter.	15.95451	
Durbin-Watson stat	2.198437			

Рисунок 18 – Результат 3 теста Дикки-Фуллера

Расширенный 3-ий тест

Во вкладке *Test for unit root in* выбираем *2nd difference*, а в *Include in test equation* – *Intercept*.

Null Hypothesis: D(DOHOD,2) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=5)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.212139	0.0004
Test critical values:		
1% level	-3.769597	
5% level	-3.004861	
10% level	-2.642242	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DOHOD,3)
Method: Least Squares
Date: 12/10/15 Time: 15:11
Sample (adjusted): 1994 2015
Included observations: 22 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(DOHOD(-1),2)	-1.950232	0.374171	-5.212139	0.0000
D(DOHOD(-1),3)	0.411748	0.226917	1.814532	0.0854
C	158.7811	148.9121	1.066274	0.2997
R-squared	0.731201	Mean dependent var	-16.67168	
Adjusted R-squared	0.702906	S.D. dependent var	1237.326	
S.E. of regression	674.4203	Akaike info criterion	15.99171	
Sum squared resid	8642011	Schwarz criterion	16.14049	
Log likelihood	-172.9088	Hannan-Quinn criter.	16.02676	
F-statistic	25.84241	Durbin-Watson stat	2.047435	
Prob(F-statistic)	0.000004			

Рисунок 19 – Результат расширенного 3 теста Дикки-Фуллера

4 тест

Во вкладке *Test for unit root in* выбираем *1st difference*, а в *Include in test equation* – *none*.

Null Hypothesis: D(DOHOD) has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic - based on SIC, maxlag=5)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-0.921839	0.3070
Test critical values:	1% level		-2.664853	
	5% level		-1.955681	
	10% level		-1.608793	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DOHOD,2)				
Method: Least Squares				
Date: 12/10/15 Time: 15:12				
Sample (adjusted): 1992 2015				
Included observations: 24 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(DOHOD(-1))	-0.090231	0.097881	-0.921839	0.3662
R-squared	0.027150	Mean dependent var		65.67500
Adjusted R-squared	0.027150	S.D. dependent var		715.3922
S.E. of regression	705.6140	Akaike info criterion		15.99679
Sum squared resid	11451495	Schwarz criterion		16.04587
Log likelihood	-190.9614	Hannan-Quinn criter.		16.00981
Durbin-Watson stat	2.564105			

Рисунок 20 – Результат 4 теста Дикки-Фуллера

Расширенный 4-ый тест

Во вкладке *Test for unit root in* выбираем *1st difference*, а в *Include in test equation* – *Intercept*.

Null Hypothesis: D(DOHOD) has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=5)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-2.207159	0.2089
Test critical values:	1% level		-3.737853	
	5% level		-2.991878	
	10% level		-2.635542	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DOHOD,2)				
Method: Least Squares				
Date: 12/10/15 Time: 15:13				
Sample (adjusted): 1992 2015				
Included observations: 24 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(DOHOD(-1))	-0.327705	0.148474	-2.207159	0.0380
C	444.6482	218.4802	2.035188	0.0541
R-squared	0.181290	Mean dependent var		65.67500
Adjusted R-squared	0.144076	S.D. dependent var		715.3922
S.E. of regression	661.8534	Akaike info criterion		15.90762
Sum squared resid	9637097.	Schwarz criterion		16.00579
Log likelihood	-188.8914	Hannan-Quinn criter.		15.93367
F-statistic	4.871551	Durbin-Watson stat		2.377424
Prob(F-statistic)	0.038037			

Рисунок 21 – Результат расширенного 4 теста Дикки-Фуллера

5 тест

Во вкладке *Test for unit root in* выбираем *level*, а в *Include in test equation* - *none*.

Null Hypothesis: DOHOD has a unit root				
Exogenous: None				
Lag Length: 1 (Automatic - based on SIC, maxlag=5)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			1.635558	0.9714
Test critical values:	1% level		-2.664853	
	5% level		-1.955681	
	10% level		-1.608793	
*Mackinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DOHOD)				
Method: Least Squares				
Date: 12/10/15 Time: 15:13				
Sample (adjusted): 1992 2015				
Included observations: 24 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DOHOD(-1)	0.042834	0.026189	1.635558	0.1162
D(DOHOD(-1))	0.570882	0.227732	2.506812	0.0201
R-squared	0.451619	Mean dependent var		1222.120
Adjusted R-squared	0.426692	S.D. dependent var		899.7217
S.E. of regression	681.2430	Akaike info criterion		15.96537
Sum squared resid	10210026	Schwarz criterion		16.06354
Log likelihood	-189.5844	Hannan-Quinn criter.		15.99142
Durbin-Watson stat	2.098916			

Рисунок 22 – Результат 5 теста Дикки-Фуллера

Расширенный 5-ый тест

Во вкладке *Test for unit root in* выбираем *level*, а в *Include in test equation* – *Intercept*.

Null Hypothesis: DOHOD has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=5)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			5.043774	1.0000
Test critical values:	1% level		-3.724070	
	5% level		-2.986225	
	10% level		-2.632604	
*Mackinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DOHOD)				
Method: Least Squares				
Date: 12/10/15 Time: 15:14				
Sample (adjusted): 1991 2015				
Included observations: 25 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DOHOD(-1)	0.072684	0.014411	5.043774	0.0000
C	533.6048	180.6686	2.953500	0.0071
R-squared	0.525182	Mean dependent var		1173.387
Adjusted R-squared	0.504538	S.D. dependent var		913.8611
S.E. of regression	643.2582	Akaike info criterion		15.84759
Sum squared resid	9516965.	Schwarz criterion		15.94510
Log likelihood	-196.0948	Hannan-Quinn criter.		15.87463
F-statistic	25.43966	Durbin-Watson stat		1.348202
Prob(F-statistic)	0.000042			

Рисунок 23 – Результат расширенного 5 теста Дикки-Фуллера

Определение типа процесса по результатам проделанных тестов:

Процесс	Результаты теста
TS+DS	<ul style="list-style-type: none"> • выполняются 1, 4, 2 (статистически значимый тренд) тест • 3,5 тест не выполняются
TS	<ul style="list-style-type: none"> • 1 тест (trend значим) • 2 тест выполняется • 3,4,5 не выполняются
DS (I_2)	<ul style="list-style-type: none"> • 3 тест выполняется или его модификация • возможно, выполняется 1 тест • 2,4,5 не выполняется <p>Если сработал тест с модификацией, это значит, что в модель целесообразно включать константу (<i>Intercept</i>).</p>
DS (I_1)	<ul style="list-style-type: none"> • 3 тест выполняется • 4 тест выполняется или его модификация • 1,2,5 тесты не выполняются
DS (I_0)	<ul style="list-style-type: none"> • 3,4 тесты выполняются • 5 тест выполняется или его модификация • 1,2 тесты не выполняются

*Тест Дикки-Фуллера “может дать” ошибку, если есть структурные скачки, в этом случае необходимо использовать тест Перрона.

Проанализировав тесты, получаем следующие результаты: 1 и 3 тест считаются выполненными, а 2,4 и 5 невыполненными, из этого следует, что временной ряд является процессом DS (I_2).

Часть 2

Моделирование с помощью ARIMA – инструментов.

Цель работы: смоделировать показатель с помощью ARIMA-инструментов для построения в дальнейшем качественного прогноза.

Задачи для достижения цели:

1. Определение зависимой переменной в ARIMA-модели, подлежащей моделированию.
2. Идентификация модели – определение порядков модели ARIMA
3. Построение модели по идентифицированным порядкам.
4. Диагностика ARIMA-модели

1. Определение зависимой переменной в ARIMA-модели, подлежащей моделированию.

Правила задания зависимого процесса в EViews:

1. Если процесс DS (I_0), то в качестве зависимой переменной берется сама исходная переменная.
2. Если процесс DS (I_1), то в качестве зависимой переменной выбираем d (“название переменной”). d-означает взятие разности.
3. Если DS (I_2), то в качестве зависимой переменной берем d(d(“название переменной”)) или d (“название переменной”_2).
4. Если TS, то в качестве зависимой переменной берем исходную переменную, а в качестве независимой переменной добавляем переменную t.
5. Если процесс TS+DS, то в качестве зависимой переменной берем переменной d(“название переменной”), а в качестве независимой переменной t.

Примечание: для создания переменной t в EViews. В меню рабочего файла выбирают команду Object/New object. В появившемся диалоговом окне New Object выбирают объект типа – ряд (Series), в рамке Name of Object задают имя (рис.24).

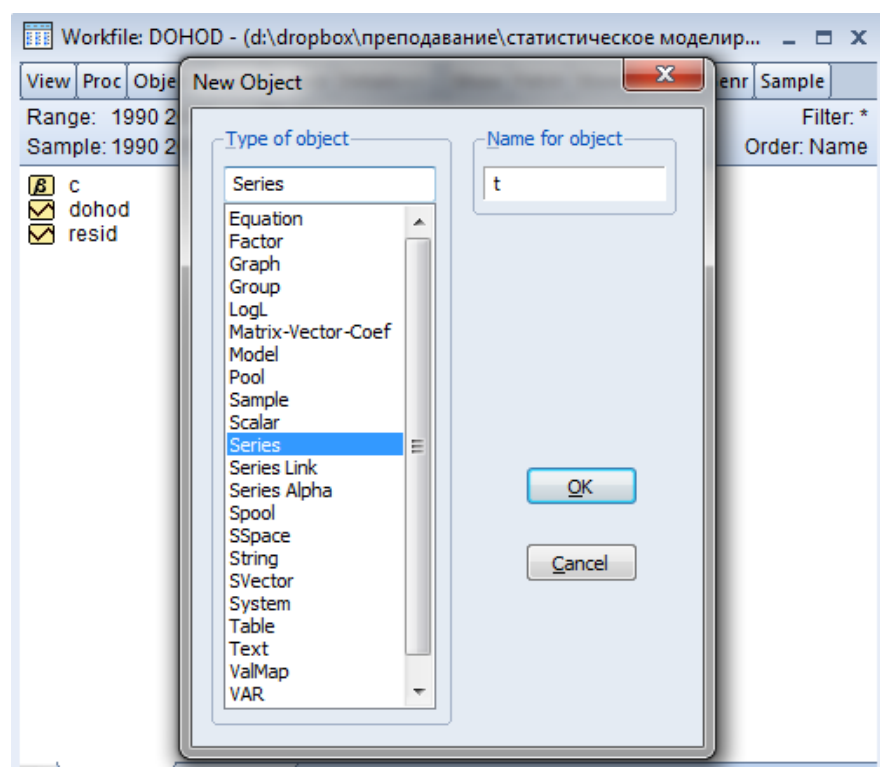


Рисунок 24 – Создание переменной t

Для непосредственного ввода данных в окне редактирования ряда *Series*: необходимо выбрать команду Edit+/- . Появится строка ввода данных. Принцип ввода информации как в любой электронной таблице (рис. 6). Переменная записывается последовательно от 1 до n, где n-количество наблюдений временного ряда.

Year	Value
1990	1.000000
1991	2.000000
1992	3.000000
1993	4.000000
1994	5.000000
1995	6.000000
1996	7.000000
1997	8.000000
1998	9.000000
1999	10.000000
2000	11.000000
2001	12.000000
2002	13.000000
2003	14.000000
2004	15.000000
2005	16.000000
2006	17.000000
2007	18.000000
2008	19.000000

Рисунок 25 – Формирование ряда данных t

2. Идентификация модели – определение порядков модели ARIMA

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \alpha_q \varepsilon_{t-q} + \varepsilon_t$$

- $\alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p}$ – процесс AR
- $\beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \alpha_q \varepsilon_{t-q} + \varepsilon_t$ – процесс MA

Для построения модели важно, чтобы была AR-часть, та как на части MA модель построить невозможно.

Примечание:

- Если рассматривается модель с запаздыванием на 1 шаг, то прогноз можно составить только на 1 шаг вперед.
- Шоком называют ε_t . Модель с бесконечными шоками можно представить конечной авторегрессионной моделью.

Для определения q и p необходимо проанализировать коррелограммы.

- Если тип процесса DS (I_0), строится коррелограмма на исходном ряде.
- Если тип процесса DS (I_1) или TS, то строится коррелограмма на 1-ых разностях.
- Если типа процесса DS (I_2) или TS + DS, то строится коррелограмма на 2-ых разностях.

Для определения порядка AR модели необходимо проанализировать ЧКФ коррелограмму, и порядок AR равен номеру лагов, соответствующих коэффициентов, выходящих за границу белого шума. Для определения порядка MA модели используется АКФ коррелограмма.

3. Построение модели по идентифицированным порядкам.

Для построения модели в EViews используется метод наименьших квадратов (МНК). В главном окне выбираем команду Quick-Estimate Equation (рис.26). В появившееся окно вводим (рис. 27) согласно правилам задания зависимого процесса (пункт 1, часть 2) уравнения для построения модели. Для AR части используется term – ar (порядок), например ar (2), ar (4), для MA части – ma (порядок), например ma(1), ma(2) и т.д.

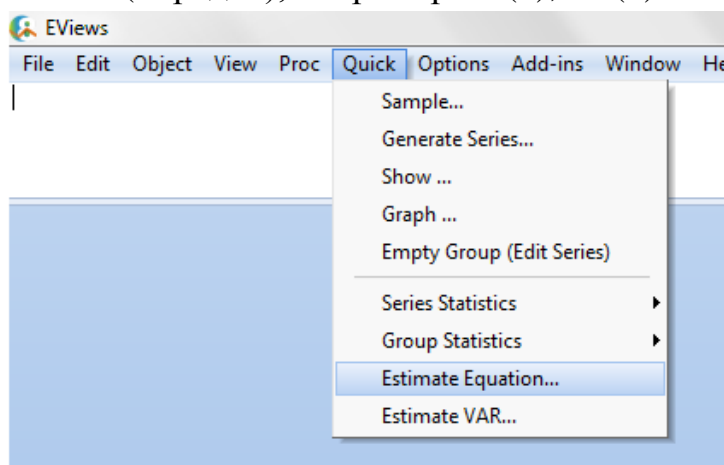


Рисунок 26 – Выбор окна ввода уравнения для построения модели ARMA

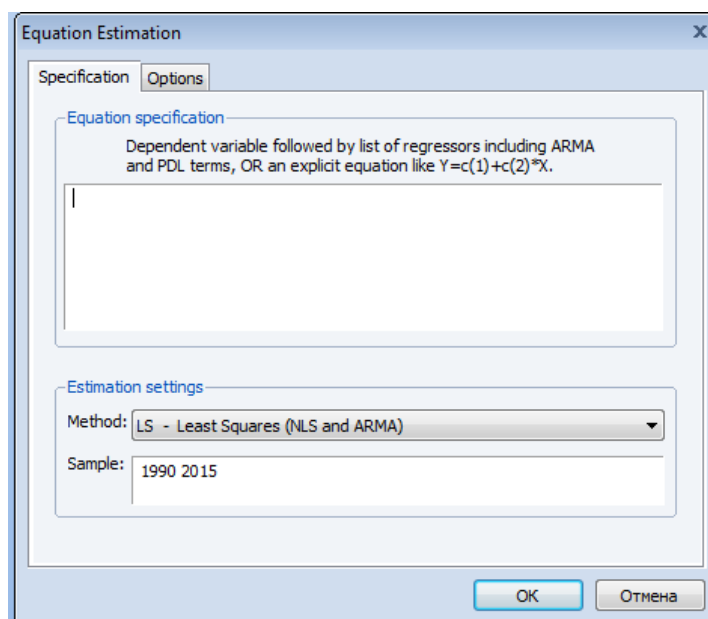


Рисунок 27 – Окно ввода модели

Примечание:

В случае сезонной модели ARMA указывается порядок сезонности. Если амплитуда сезонных колебаний постоянна, то модель аддитивная, иначе мультипликативная.

Для аддитивной модели указывается AR (период сезонности) и MA(период сезонности). Для мультипликативной модели – SAR (период сезонности) и SAM (период сезонности).

4. Диагностика ARMA-модели

Модель считается качественно построенной, если выполняются следующие условия:

1. Значимые коэффициенты. Все коэффициенты модели статистически значимы, то есть соотносящиеся им уровни доверительной вероятности $p\text{-value} < 0,05$.

2. Структура единичных корней. Обратные корни характеристического уравнения ARIMA-модели не выходят за границы единичного круга.

3. Коррелограммы АКФ и ЧКФ остатков должны указывать на то, что остатки соответствуют белому шуму.

4. Статистика Дарбина—Уотсона должна свидетельствовать об отсутствии автокорреляции в остатках, то есть находиться в границах 1,6 и 2,4.

5. Коэффициент детерминации (*R-squared*) должен стремиться к 1.

6. Построенная модель должна быть значимой. F-statistic является значимой, если соответствующая вероятность $p\text{-value}$ ($Prob(F - statistic)) < 0,05$.

7. Остатки подчиняются нормальному закону распределения. Выдвигаются H_0 : остатки подчиняются нормальному закону распределения, и H_1 : остатки не подчиняются нормальному закону распределения. Для проверки данных гипотез используется тест Бера–Жарка. Для принятия H_0 гипотезы необходимо, чтобы вероятность $p\text{-value}$ была больше 0,05.

Рассмотрим диагностику построенной модели в EViews. Так как анализируемый ряд относится к DS (I_2) необходимо перед построением ARIMA модели анализировать коррелограммы на 2-ых разностях. Для ввода уравнения воспользуемся также правилами задания зависимого процесса (пункт 1, часть 2). Получим уравнение вида, представленного на рисунке 28. Результат анализа модели представлен на рисунке 29.

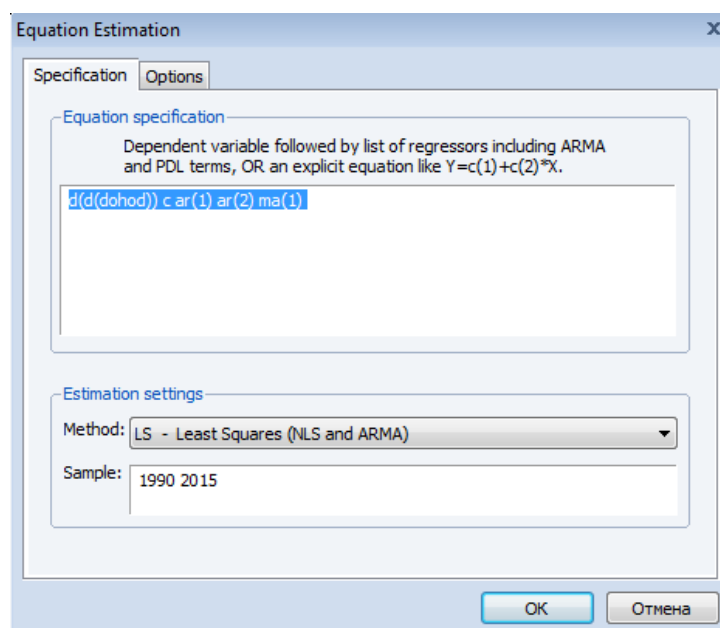


Рисунок 28 – Ввод уравнения модели

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: D(D(DOHOD))									
Method: Least Squares									
Date: 12/13/15 Time: 11:32									
Sample (adjusted): 1994 2015									
Included observations: 22 after adjustments									
Convergence achieved after 369 iterations									
MA Backcast: OFF (Roots of MA process too large)									
Variable		Coefficient	Std. Error	t-Statistic	Prob.				
C		66.04771	4.241114	15.57320	0.0000				
AR(1)		-0.186392	0.238392	-0.781874	0.4445				
AR(2)		-0.302878	0.244538	-1.238572	0.2314				
MA(1)		-1.847987	0.369938	-4.995402	0.0001				
R-squared		0.799712	Mean dependent var	64.48500					
Adjusted R-squared		0.766331	S.D. dependent var	748.5635					
S.E. of regression		361.8506	Akaike info criterion	14.78331					
Sum squared resid		2356846.	Schwarz criterion	14.98168					
Log likelihood		-158.6164	Hannan-Quinn criter.	14.83004					
F-statistic		23.95689	Durbin-Watson stat	1.886922					
Prob(F-statistic)		0.000002							
Inverted AR Roots		-.09+.54i	-.09-.54i						
Inverted MA Roots		1.85							
Estimated MA process is noninvertible									

Рисунок 29 – Окно вывода построенной модели

Продиагностируем качество построенной модели.

1. Значимость коэффициентов. В построенной модели только 2 коэффициента являются статистически значимы (коэффициент при МА (1) и свободный член).
2. Структура единичных корней. Для проверки единичных корней (рис. 30) воспользуемся командой *View-ARMA-Structure-Roots*.

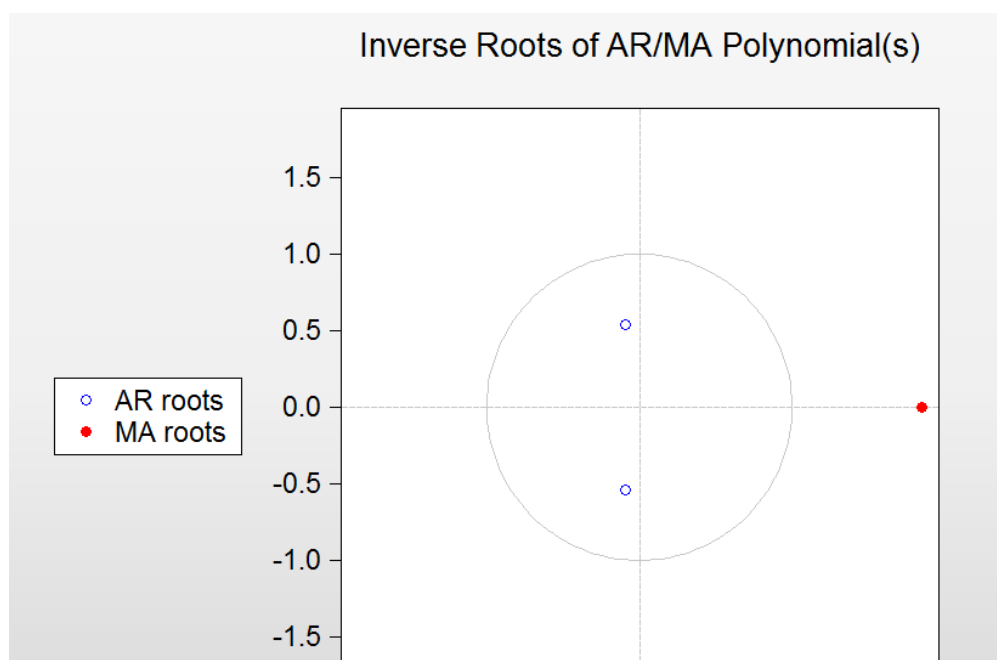


Рисунок 30 – Проверка единичных корней ARMA

Проанализировав рисунок 30 сделаем вывод, что корни характеристического уравнения AR части модели лежат в единичном круге, а MA части не лежат в единичном круге.

3. Для построения коррелограммы АКФ и ЧКФ остатков используем команду *View - Residual Diagnostics - Correlogram-Q-statistics* (рис.31).

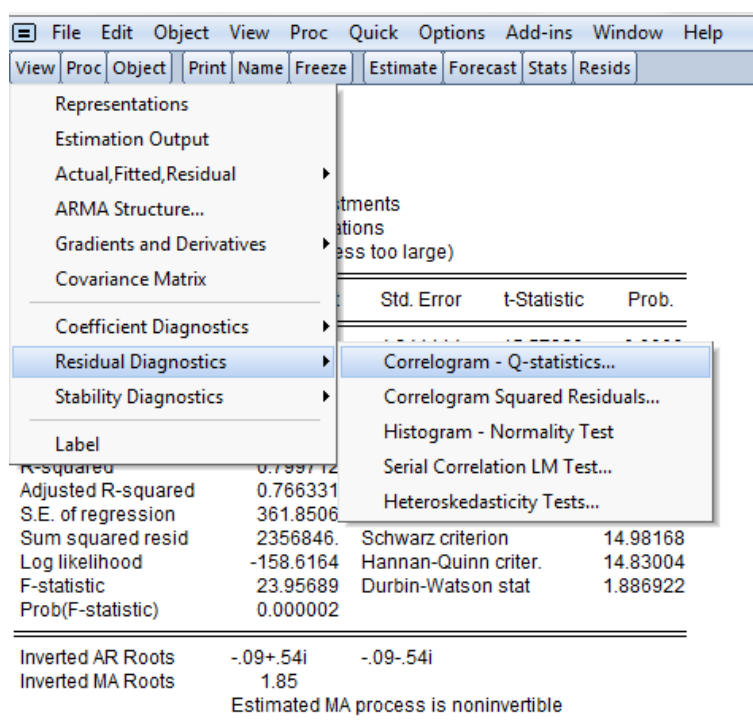


Рисунок 31 – Построение коррелограмм остатков

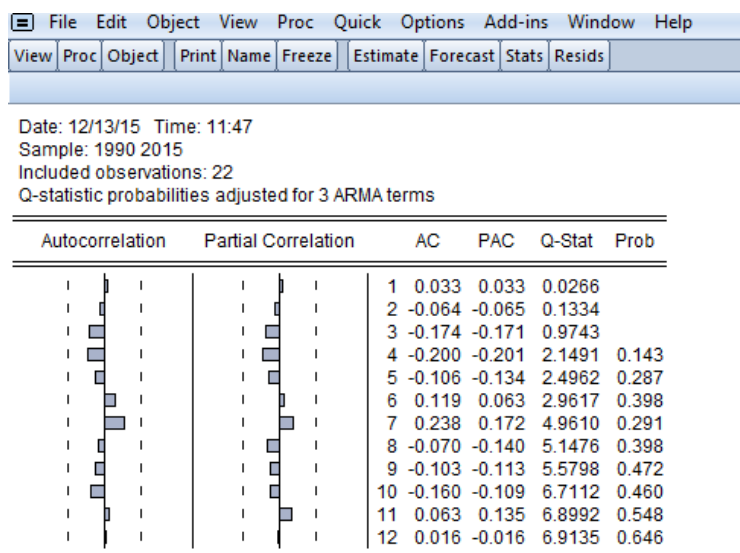


Рисунок 32 – Коррелограммы остатков

Коррелограммы (рис.32) не указывают то, что остатки соответствуют белому шуму.

4.Статистика Дарбина—Уотсона свидетельствует об отсутствии автокорреляции в остатках, так как равна 1,89, то есть находиться в границах 1,6 и 2,4 (рис.29).

5. Коэффициент детерминации (*R-squared*) равен 0,79 (рис.29).

6. Построенная модель является значимой, так как вероятность *p-value* (*Prob (F – statistic)*) < 0,05 (рис. 29).

7. Для проверки гипотезы о нормальном законе распределения остатков воспользуемся тестом Бера–Жарка (рис.33).

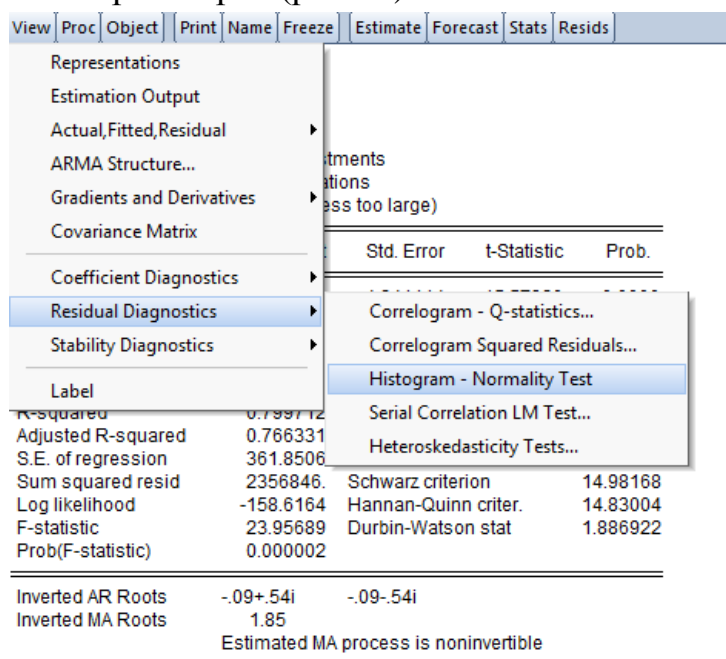


Рисунок 33 – Выбор теста Бера–Жарка

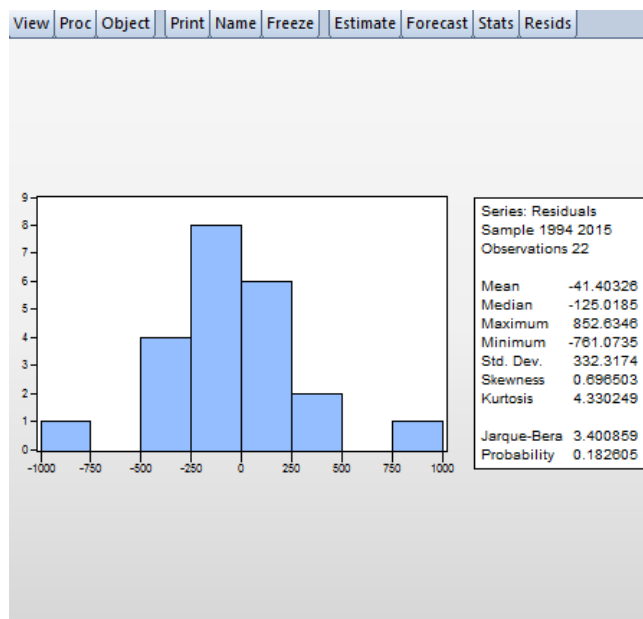


Рисунок 34 – Результат анализа теста Бера–Жарка

По рисунку 34 можно сделать вывод, что остатки распределены по нормальному закону распределения, так как вероятность p -value больше 0, 05 (0,18) для теста Бера–Жарка.

Так как не все условия выполняются, необходимо перестроить модель. Для перестройки модели необходимо воспользоваться командой *Proc-Specify/Estimate*. Для проверки качества построенной модели также используют график исходных данных и полученных с помощью модели (рис. 35).

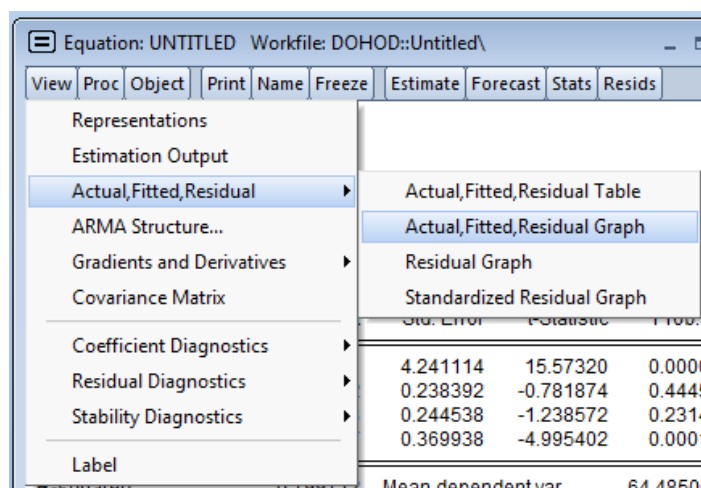


Рисунок 35 – Построение исходных данных и наложенных

Примечание.

Один и тот же процесс могут описывать несколько качественных моделей. В этом случае проводят этап селекции модели. Выбирают наилучшую модель на основе информационных критериев Акаике (Akaike info criterion), Шварца (Schwarz criterion), Ханнана-Куинна (Hannan-Quinn

criterion). Выбирается та модель, для которой значение информационных критериев наименьшее.

Задание на лабораторную работу*

1. Выбрать временной ряд для построения качественной ARMA модели. Выборка должна включать не менее 25 наблюдений.
2. Определить тип процесса, к которому относится временной ряд.
3. Построить ARMA модель для выбранного временного ряда.
4. Проверить качество построенной модели. Скорректировать модель, если это необходимо.
5. Выбрать наилучшую модель, если процесс, может быть, описан несколькими моделями.
6. Оформить результаты в виде отчета по лабораторной работе.

*Вариант у каждого студента индивидуален.