

## Занятие 8-9. Тесты единичного корня. Моделирование с помощью ARIMA.

### Часть 1. Тесты единичного корня. Расширенный тест Дики-Фуллера

1. В чем суть теста Дики-Фуллера? В чем суть расширенного теста Дики-Фуллера? Какое количество лагов необходимо включать? К каким последствиям это приведет?
2. Как проверить наличие второго единичного корня?
3. Какие тесты единичного корня еще Вы знаете?

**Задача 1.** Для некоторого временного ряда  $y_t$  ( $T=100$ ) получена следующая модель (в скобках стандартные ошибки коэффициентов):

$$\Delta \hat{y}_t = 2,95 - 0,18y_{t-1} \\ (1,147) \quad (0,069)$$

На уровне значимости 5% сформулировать и проверить гипотезу единичного корня (порядок интегрируемости ряда  $y_t$  равен 1, против альтернативной, что порядок интегрируемости равен 0).

**Односторонние критические значения статистики Дики-Фуллера (Магнус)**

$$y_t = b_1 y_{t-1} + \varepsilon_{1t}, \quad (11.48)$$

$$y_t = a_2 + b_2 y_{t-1} + \varepsilon_{2t}, \quad (11.49)$$

$$y_t = a_2 + b_3 y_{t-1} + c_3 t + \varepsilon_{3t} \quad (11.50)$$

Доверительный уровень	Размер выборки			
	25	50	100	$\infty$
AR модель (11.48)				
0.010	-2.66	-2.62	-2.60	-2.58
0.025	-2.26	-2.25	-2.24	-2.23
0.050	-1.95	-1.95	-1.95	-1.95
AR модель с константой (11.49)				
0.010	-3.75	-3.58	-3.51	-3.43
0.025	-3.33	-3.22	-3.17	-3.12
0.050	-3.00	-2.93	-2.89	-2.86
AR модель с константой и трендом (11.50)				
0.010	-4.38	-4.15	-4.04	-3.96
0.025	-3.95	-3.80	-3.69	-3.66
0.050	-3.60	-3.50	-3.45	-3.41

Источник: (Fuller, 1976).

**Задача 2.** Для некоторого временного ряда  $y_t$  ( $T=500$ ) получена следующая тестовая регрессия:

Dickey-Fuller test for unit root      Number of obs =      499

Interpolated Dickey-Fuller		
1% Critical Value	5% Critical Value	10% Critical Value
-3.440	-2.870	-2.570

D. x1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1						
L1.	-1.00511	.0449404	-22.37	0.000	-1.093407	-.9168135
_cons	-.0038383	.0459519	-0.08	0.933	-.0941223	.0864457

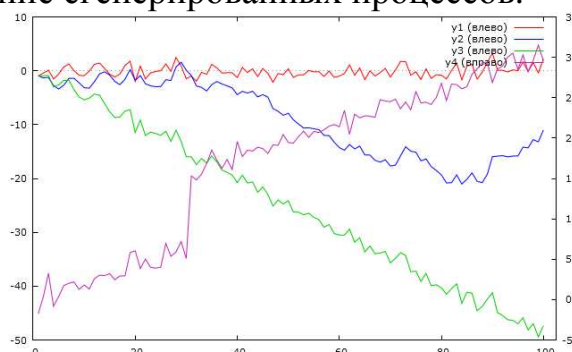
На уровне значимости 1% проверить гипотезу единичного корня.

### Работа в Gretl.

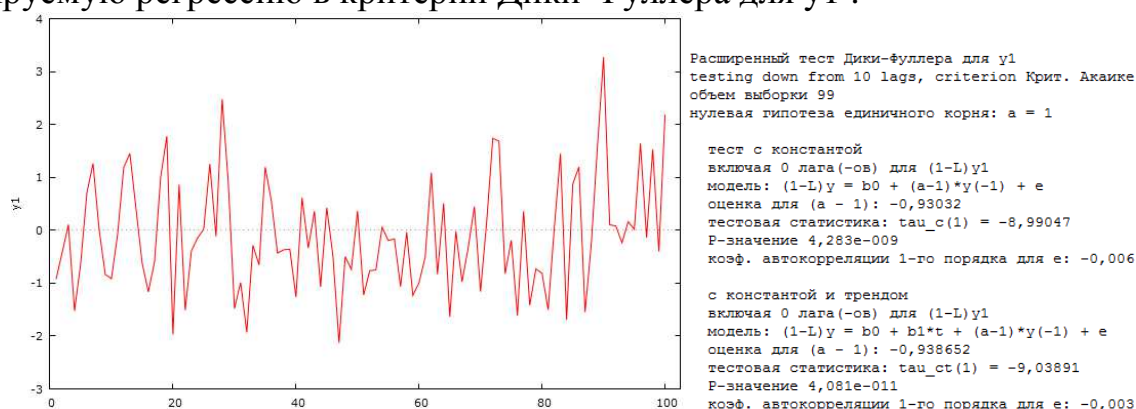
**Задача 3.** Даны три случайных процесса  $y_1$ ,  $y_2$ ,  $y_3$ ,  $y_4$ .

Файл: **DF.dta** откройте в Gretl

3.1. Исследуйте поведение сгенерированных процессов.



3.2. сформулируйте и проверьте гипотезу о наличие единичного корня. Запишите тестируемую регрессию в критерии Дики-Фуллера для  $y_1$ .



Расширенный тест Дики-Фуллера для  $y_2$   
 testing down from 10 lags, criterion Крит. Акаике  
 объем выборки 99  
 нулевая гипотеза единичного корня:  $a = 1$

тест с константой  
 включая 0 лага(-ов) для  $(1-L)y_2$   
 модель:  $(1-L)y = b_0 + (a-1)y(-1) + e$   
 оценка для  $(a - 1)$ : -0,0180801  
 тестовая статистика:  $\tau_{a\_c}(1) = -1,18331$   
 P-значение 0,6793  
 коэф. автокорреляции 1-го порядка для e: 0,072

с константой и трендом  
 включая 0 лага(-ов) для  $(1-L)y_2$   
 модель:  $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + e$   
 оценка для  $(a - 1)$ : -0,0210462  
 тестовая статистика:  $\tau_{a\_ct}(1) = -0,571397$   
 P-значение 0,9783  
 коэф. автокорреляции 1-го порядка для e: 0,074

Расширенный тест Дики-Фуллера для  $y_3$   
 testing down from 10 lags, criterion Крит. Акаике  
 объем выборки 90  
 нулевая гипотеза единичного корня:  $a = 1$

тест с константой  
 включая 9 лага(-ов) для  $(1-L)y_3$   
 модель:  $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$   
 оценка для  $(a - 1)$ : -0,00680597  
 тестовая статистика:  $\tau_{a\_c}(1) = -0,751946$   
 асимпт. p-значение 0,8317  
 коэф. автокорреляции 1-го порядка для e: -0,019  
 лаг для разностей:  $F(9, 79) = 8,806 [0,0000]$

с константой и трендом  
 включая 0 лага(-ов) для  $(1-L)y_3$   
 модель:  $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + e$   
 оценка для  $(a - 1)$ : -0,938652  
 тестовая статистика:  $\tau_{a\_ct}(1) = -9,03891$   
 P-значение 4,08e-011  
 коэф. автокорреляции 1-го порядка для e: -0,003

3.3. предположив наличие в процессах детерминированного тренда, проведите тест Дики-Фуллера. В чем суть процедуры Доладо? Сделайте вывод.

3.4. используйте тест Дики-Фуллера для первой разности изучаемых процессов (в случае необходимости). Сделайте вывод о порядке интегрируемости процессов.

3.5. используйте расширенный тест Дики-Фуллера. Какое количество лагов необходимо добавить? Сделайте вывод: как меняется DF-статистика при

добавлении дополнительных лагов и выводы относительно наличия единичного корня?

	DF	Тренд	DF(1)	DF(2)	DF(3)	DF(4)
y1						
y2						
y3						

3.6. Какие тесты единичного корня Вы знаете? Проведите PP- и KPSS-тесты для y1, сравните результаты.

	DF	PP-тест	KPSS-тест																							
y1	<p>Расширенный тест Дики-Фуллера для y1 testing down from 10 lags, criterion Крит. Акаике объем выборки 99 нулевая гипотеза единичного корня: <math>\alpha = 1</math></p> <p>тест с константой включая 0 лагов (-ов) для (1-L)y1 модель: <math>(1-L)y = b_0 + (\alpha-1)*y(-1) + e</math> оценка для <math>(\alpha - 1)</math>: -0,93032 тестовая статистика: <math>\tau_{\alpha\_c}(1) = -8,99047</math> Р-значение 4,283e-009 коэф. автокорреляции 1-го порядка для e: -0,006</p> <p>с константой и трендом включая 0 лагов (-ов) для (1-L)y1 модель: <math>(1-L)y = b_0 + b_1*t + (\alpha-1)*y(-1) + e</math> оценка для <math>(\alpha - 1)</math>: -0,938652 тестовая статистика: <math>\tau_{\alpha\_ct}(1) = -9,03891</math> Р-значение 4,081e-011 коэф. автокорреляции 1-го порядка для e: -0,003</p>	<p>Phillips-Perron unit-root test for y1, Bartlett bandwidth 3:  <math>Z_t = -8,97738</math> (p-value = 0,0000)</p> <p>Test regression (OLS, dependent variable y1, T = 99):</p> <table><thead><tr><th></th><th>Коэффициент</th><th>Ст. ошибка</th><th>z</th><th>Р-значение</th></tr></thead><tbody><tr><td>const</td><td>-0,0925384</td><td>0,108371</td><td>-0,8539</td><td>0,3932</td></tr><tr><td>y1(-1)</td><td>0,0696797</td><td>0,103479</td><td>0,6734</td><td>0,5007</td></tr></tbody></table> <p>Sample variance of residual 1,14388 Estimated long-run error variance 1,09422</p>		Коэффициент	Ст. ошибка	z	Р-значение	const	-0,0925384	0,108371	-0,8539	0,3932	y1(-1)	0,0696797	0,103479	0,6734	0,5007	<p>KPSS тест для y1 T = 100 Параметр для усечения лагов = 4 Тестовая статистика = 0,266432</p> <table><thead><tr><th></th><th>10%</th><th>5%</th><th>1%</th></tr></thead><tbody><tr><td>Крит. значения:</td><td>0,349</td><td>0,462</td><td>0,734</td></tr></tbody></table> <p>Р-значение &gt; .10</p>		10%	5%	1%	Крит. значения:	0,349	0,462	0,734
	Коэффициент	Ст. ошибка	z	Р-значение																						
const	-0,0925384	0,108371	-0,8539	0,3932																						
y1(-1)	0,0696797	0,103479	0,6734	0,5007																						
	10%	5%	1%																							
Крит. значения:	0,349	0,462	0,734																							
y2																										

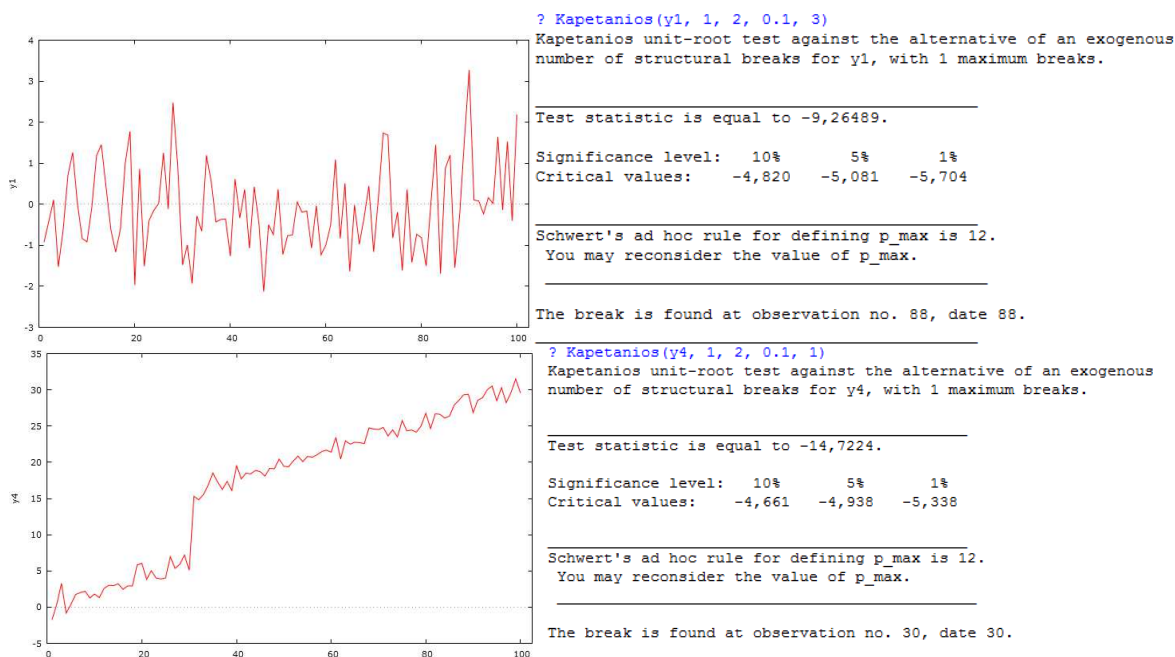
### Тесты единичного корня

Тесты единичного корня	Stata	Gretl
Тест Дики-Фуллера для y	dfuller y dfuller y, trend regress lags(2)	Menu path: /Variable/Unit root tests/Augmented Dickey-Fuller test
Тест KPSS для y	kpss y kpss y, notrend auto	Расширенный тест Дики-Фуллера (ADF-тест) ADF-GLS тест
Тест Филиппса-Перрона для y	pperron y pperron y, regress	KPSS тест Levin-Lin-Chu test
Тест Эндрюса-Зивота	zandrews y zandrews y, lagmethod(BIC) graph	Частичное интегрирование Phillips-Perron test HEGY

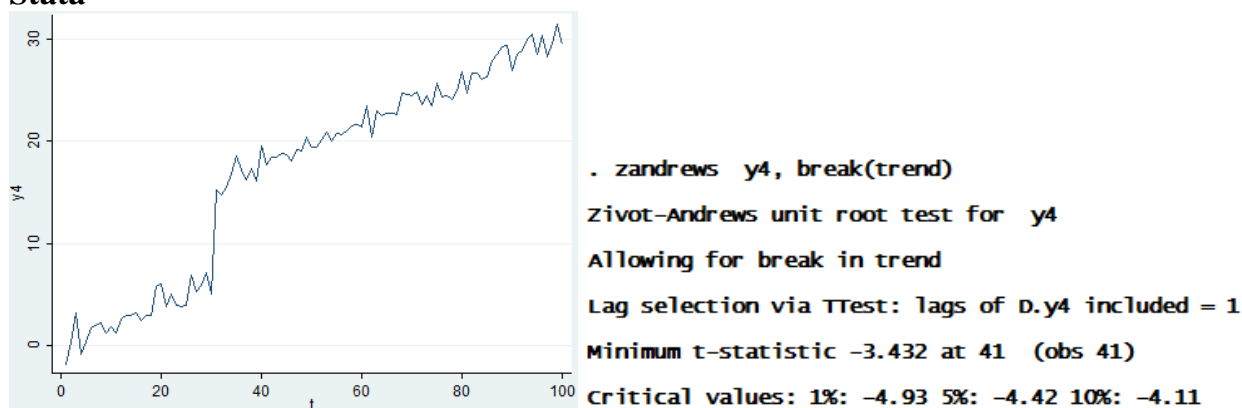
3.7. Проведите тесты единичного корня для y1,y4, в предположении наличия структурного сдвига (Gretl: Kapetanios' (2005) unit root test). Сделайте выводы.

Kapetanios, G. (2005). Unit-root testing against the alternative hypothesis of up to m structural breaks. Journal of Time Series Analysis, 26(1), 123-133.

The test controls for up to 5 level and/or trend breaks under the *alternative hypothesis* of trend stationarity.



## Stata



## Домашнее задание (ТДЗ) 8. Unit root

По данным Всемирного банка выберите 3 показателя за 40-60 лет (опишите какой показатель был взят для анализа, за какой период).

Файл: **WDI**. (закладка **Data**)

!Можно взять свои данные

- Опишите выбранные показатели. Постройте графики выбранных показателей. Сделайте вывод о стационарности рядов, исходя из построенных графиков.
- Проведите тесты единичного корня (ADF, PP, KPSS) и их модификации. Сравните результаты и сделайте вывод по результатам тестирования. Для **одного** из показателей результаты представьте в виде сводной таблицы.

ВР	Тест	Нулевая гипотеза	Статистика критерия	p-значение	Вывод
----	------	------------------	---------------------	------------	-------

у	DF(с трендом/без)				
	PP(с трендом/без)				
	KPSS(с трендом/без)				
Δу	DF(с трендом/без)				
	PP(с трендом/без)				
	KPSS(с трендом/без)				

По двум другим рядам приведите основные выводы.

3. Сделайте вывод о стационарности рядов и степени интегрируемости ( $d=?$ ).  
Являются ли ряды тренд-стационарным или разностно-стационарным?

Ряд	Стационарность
У1	
У2	
У3	

4. Для одного из показателей проведите тест единичного корня (любой) с **учетом структурного сдвига**. Какая нулевая гипотеза. Сделайте вывод по результатам тестирования.

## Часть 2. Анализ временных рядов с помощью ARIMA (Gretl).

### Методология Бокса-Дженкинса.

1. В чем суть методологии Бокса-Дженкинса?
2. Какие этапы содержит?

**Задание 2.1. Запись ARIMA-моделей.** Запишите модели в виде через лаговый оператор, ответ поясните

ARIMA(1,1,1)

ARIMA(2,1,1)

ARIMA(1,2,2)

ARIMA(3,0,0)

### Задание 2.2. Анализ младенческой смертности

**Файл:** млад смерт РМУБ.xls

**Данные:** младенческая смертность в России за период 1960-2017 гг.

**Источник данных:** Демоскоп <http://www.demoscope.ru/weekly/pril.php>

**Требуется** подобрать соответствующую модель ARIMA для описания динамики *младенческой смертности*, оценить ее параметры и построить прогноз.

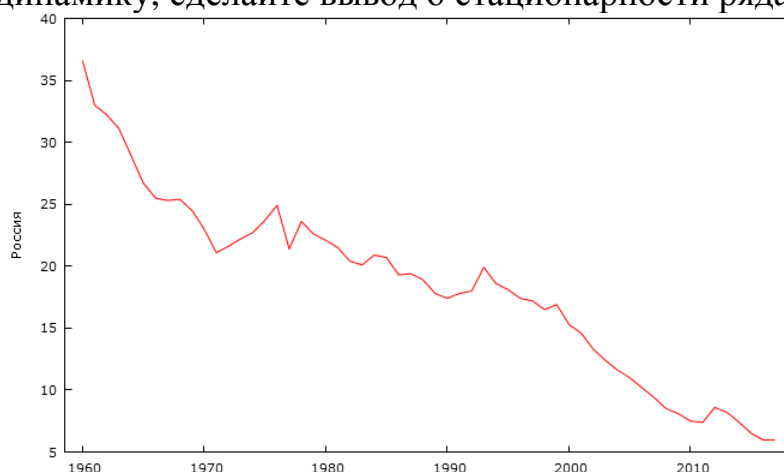
Важным показателем смертности и одновременно качества жизни является коэффициент младенческой смертности - число умерших в возрасте до 1 года в расчете на 1000 родившихся живыми. В отличие от общего коэффициента смертности, то поднимавшегося, то снижавшегося на протяжении двух последних десятилетий, коэффициент младенческой смертности довольно устойчиво снижался (рис). Наблюдавшиеся повышения значения коэффициента младенческой смертности были связаны в основном с улучшением качества учета и постепенным переходом на международный стандарт в определении живорождения - в 1993 году (на 11%)<sup>[12]</sup> и в 2012 году (на 17%)<sup>[13]</sup>. Дополнительное расширение критериев живорождения в 2013 году не привело к повышению показателей младенческой смертности. Число детей, умерших в возрасте до 1 года, быстро снижалось в 1960-е годы за счет сокращения и рождаемости, и смертности, но в 1972-1976 годах стало расти (отчасти за счет улучшения учета родившихся). С конца 1980-х годов число умерших в возрасте до 1 года неуклонно сокращалось, снизившись с 48,5 тысячи в 1987 году до 13,2 тысячи человек в 2011 году. В 2012 году число зарегистрированных смертей в возрасте до 1 года в связи с расширением критериев живорождения увеличилось до 16,3 тысячи человек, что на 24% больше, чем за 2011 год. Значение коэффициента младенческой смертности возросло до 8,6‰ против 7,4‰ за 2011 год.

<http://www.demoscope.ru/weekly/2018/0761/barom05.php>

1. Перенесите данные в в Gretl.

### Этап 1. Идентификация модели

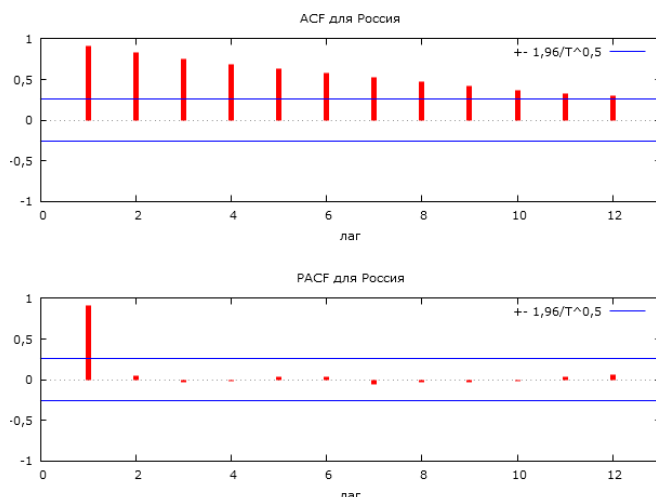
2. Постройте график временного ряда младенческой смертности в России. Опишите динамику, сделайте вывод о стационарности ряда.



3. Постройте коррелограмму (график автокорреляционной и частной автокорреляционной функции). Сделайте вывод о стационарности ряда.
4. Проведите тесты единичного корня (ADF, PP, KPSS) для ряда в уровнях и ряда первой разности (при необходимости). Сравните результаты и сделайте вывод о стационарности ряда и степени интегрируемости ( $d=?$ ). Является ли ряд тренд-стационарным или разностно-стационарным?

### Этап 2. Оценивание модели.

5. После определения степени интегрируемости ряда, перейдем к оцениванию параметров модели  $ARIMA(p, d, q)$ . Какие предположения относительно порядков  $p$  и  $q$  можно сделать на основании графиков ACF и PACF?



**Свойства автокорреляционных и частных автокорреляционных функций**

	AR(1)	AR(2)	MA(1)	MA(2)
ACF	Экспоненциально затухает	Экспоненциально затухает	Пик на лаге 1	Пик на лагах 1,2
PACF	Пик на лаге 1	Пик на лагах 1,2	Экспоненциально затухает	Экспоненциально затухает

## 6. Оцените и сравните несколько моделей **ARIMA**

### - **ARIMA (1, 1, 0)**

#### модель AR(1)

$$\Delta y_t = \text{const} + \alpha_1 \Delta y_{t-1}$$

Модель 5: ARIMA, использованы наблюдения 1961–2017 (T = 57)

Оценено при помощи фильтра Кальмана (Kalman) (точный метод МП)

Зависимая переменная: (1-L) v2

Стандартные ошибки рассчитаны на основе Гесса

	Коэффициент	Ст. ошибка	z	P-значение	
const	-0,535970	0,136881	-3,916	9,02e-05	***
phi_1	-0,0200474	0,142516	-0,1407	0,8881	
Среднее зав. перемен	-0,536842	Ст. откл. зав. перемен	1,062280		
Среднее инноваций	-0,001079	Ст. откл. инноваций	1,052735		
Лог. правдоподобие	-83,80900	Крит. Акаике	173,6180		
Крит. Шварца	179,7472	Крит. Хеннана-Куинна	176,0000		
Действ. часть Мним. часть Модуль Частота					
AR					
Корень 1	-49,8817	0,0000	49,8817	0,5000	

Что можно сказать о качестве модели?

- запишите модель через лаговый оператор и опишите ее статистические свойства.

- проверьте выполнение предпосылок **ARIMA** (стационарность и обратимость)

- Оцените модели ARMA(p, q), (p, q ≤ 2). Выберите лучшую модель с точки зрения информационных критериев Акаике и Шварца.

модели <b>ARIMA</b>	Ошибка модели	AIC	BIC
1. ARIMA(1,1,0)			
2. ARIMA(0,1,1)			
3. ARIMA(1,1,1)			
4. ARIMA(1,0,0)+лин.тренд			
5. ARIMA(2,1,1)			



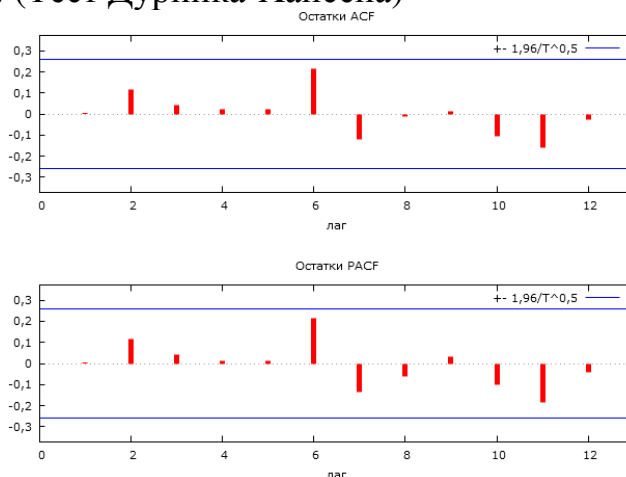
### Этап 3. Диагностика моделей.

-Выберете из полученных моделей 2 с наименьшими значениями информационных критериев.

-Оцените адекватность построенных моделей на основе **анализа остатков**.

*Анализ автокорреляций.* Постройте графики ACF/PACF остатков. Какими свойствами должен обладать ряд остатков?

*Нормальность.* (Тест Дурника-Хансена)



Тест на наличие автокорреляции до порядка 10

Ljung-Box  $Q' = 5,90074$ ,  
p-значение =  $P(\text{Хи-квадрат}(9) > 5,90074) = 0,7498$

Нулевая гипотеза - нормальное распределение:  
Хи-квадрат(2) = 8,289 p-значение 0,01585

#### Альтернативные тесты на нормальность остатков в Gretl.

1. Сохраняете остатки модели
2. Используете тесты: Переменные -Тесты на нормальное распределение

Тест на нормальное распределение uhat1:

Тест Дурника-Хансена (Doornik-Hansen) = 8,28896, p-значение 0,0158517

Тест Шапиро-Уилка (Shapiro-Wilk W) = 0,957794, p-значение 0,0448994

Тест Лиллифорса (Lilliefors) = 0,104789, p-значение ~ 0,12

Тест Жака-Бера (Jarque-Bera) = 4,95901, p-значение 0,0837846

Опишите и сравните построенные модели. Выберите наилучшую. Ответ обоснуйте.

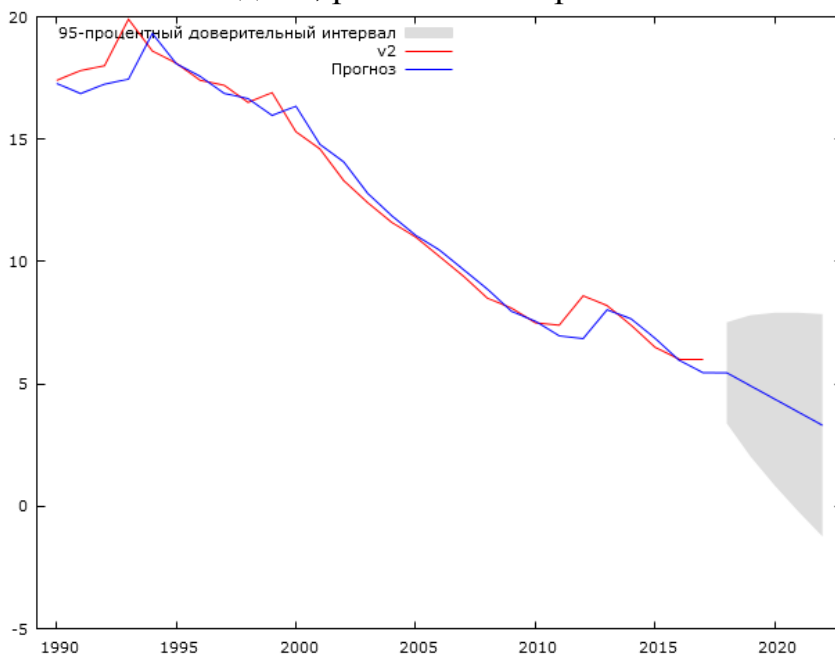
модели ARIMA	Стационарность, обратимость модели	Ошибка модели, инф.критерии	Анализ остатков (автокорреляция, нормальность)	Наилучшая модель
1. ARIMA(1,1,0)				



2. ARIMA(0,1,1)				
3. ARIMA(1,1,1)				
4. ARIMA(1,0,0)+лин. тренд				
5. ARIMA(2,1,1)				

#### Этап 4. Прогнозирование.

- Изобразите в одной системе координат исходные данные и предсказанные значения по модели, рассчитайте прогнозы.



Для 95% доверительных интервалов,  $z(0,025) = 1,96$

	v2	Предсказание	Ст. ошибка	95% доверительный интервал	
2017	6,0	5,5			
2018		5,5	1,05	3,4	7,5
2019		4,9	1,47	2,0	7,8
2020		4,4	1,80	0,9	7,9
2021		3,8	2,07	-0,2	7,9
2022		3,3	2,32	-1,2	7,9

- характеристики качества прогноза

Статистика для оценки прогноза

Средняя ошибка (ME)	-0,0010785
Корень из средней квадратичной ошибки (RMSE)	1,0528
Средняя абсолютная ошибка (MAE)	0,75233
Средняя процентная ошибка (MPE)	0,25703
Средняя абсолютная процентная ошибка (MAPE)	4,1475
U-статистика Тейла (Theil's U)	0,91019

Сделайте общий вывод о качестве полученной модели.

#### Самостоятельная работа в группе (на занятии)

**Данные:** младенческая смертность в *Украине/Молдове/Белоруссии* за период 1960-2017 гг.

Проведите аналогичный анализ в Gretl. Подберите подходящую модель ARIMA (оцените и сравните ARIMA(1,1,1), ARIMA(1,1,0), ARIMA(0,1,1)), обоснуйте выбор модели и опишите полученную модель. Проверьте адекватность модели и постройте прогноз на 6 лет (точечная и интервальная оценка).

### Домашнее задание (ТДЗ) 9. ARIMA

По данным Всемирного банка выберите один показатель за 40-60 лет (опишите какой показатель был взят для анализа, за какой период).

Файл: **WDI**. (закладка **Data**)

!Можно взять свои данные

**Задание.** Требуется подобрать соответствующую модель ARIMA для описания динамики *выбранного показателя*, оценить ее параметры и построить прогноз на основании полученной модели.

1. Опишите какой показатель был взят для анализа, за какой период, постройте график и сделайте предположение о стационарности ряда.
2. Проведите тесты единичного корня (ADF, PP, KPSS) и их модификации. Сравните результаты и сделайте вывод по результатам тестирования. Результаты представьте в виде сводной таблицы.

ВР	Тест	Нулевая гипотеза	Статистика критерия	p-значение	Вывод
у	DF(с трендом/без)				
	PP(с трендом/без)				
	KPSS(с трендом/без)				
Δу	DF(с трендом/без)				
	PP(с трендом/без)				
	KPSS(с трендом/без)				

Сделайте вывод о стационарности ряда и степени интегрируемости ( $d=?$ ). Является ли ряд тренд-стационарным или разностно-стационарным?

3. На основании ACF/PACF сделайте предположения о порядке ARIMA(p,d,q). Ответ обоснуйте.
4. Оцените и сравните несколько ARIMA(p,d,q)-моделей: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1). Запишите модели в математической форме через *лаговый* оператор. Проверьте выполнение предпосылок ARIMA-моделей (стационарность, обратимость). Сравните модели между собой по статистическим свойствам (значимость коэффициентов, ошибка модели, информационные критерии).
5. Проверьте адекватность моделей на основе анализа остатков (автокорреляция, нормальность)). Оцените прогностические свойства

полученных моделей (характеристики RMSE, MPE, MAPE). Опишите и сравните построенные модели в виде сводной таблицы. Выберите наилучшую. Ответ обоснуйте.


модели ARIMA	Ошибка модели, инф.критерии	Стационарность, обратимость процесса	Анализ остатков (автокорреляция, нормальность)	Качество прогноза
1. ARIMA( ) уравнение				
2. ARIMA( ) уравнение				
3. ARIMA( ) уравнение				

6. Постройте прогноз на 6 шагов вперед (точечная и интервальная оценка) по наилучшей модели. В отчете приведите графики (наблюдаемые+предсказанные значения). Сделайте вывод как изменится показатель.

Напишите решение задач (скан рукописного варианта, где это необходимо) и краткий отчет с выводами и полученными графиками (формат pdf). Приветствуется сдача работы в группе по 2 человека (не забывайте указывать авторов).

Выполненная домашняя работа загружается в LMS. Срок выполнения – 1 неделя.

#### Команды Stata

edit	редактирование данных (открытие редактора данных)
clear	очистить память компьютера
display	Вывод на экран значения переменной или выражения
dis	калькулятор
list	Вывод на экран значений переменных из активного множества данных
<b>Описательные статистики</b>	
list [v1]	вывести значения переменных (v1) на экран, кнопка BREAK  (прервать выполнение команды)
describe [v1]	вывести описание переменных
sum [v1]	расчет дескриптивных статистик для переменной (v1)
<b>Действия над переменными</b>	
gen v2=g(v1 )	создать новую переменную v2 как функцию g от v1
drop v1	удалить переменную v1
ren v1 v2	переименовать переменную v1 в v2
egen t=seq()	Создание последовательности целых чисел
<b>Работа с временными рядами</b>	
tsset t	Объявить переменную t переменной времени
tsline y	Построить график временного ряда y
regress y t	Построить линейную регрессию

predict y1, xb	Сохранить предсказанные значения в y1
predict e1, residuals	Сохранить значения остатков в e1
ac y	построить автокорреляционную функцию для y
sum y	вычислить описательные статистики для y
pac y	построить частную автокорреляционную функцию для y
corrgram y	Вычислить значения автокорреляционной функций для y
pergram y	Построение периодограммы для y
sktest e1	Тест на нормальность для e1
wntestq e1	Статистика Льюинга-Бокса для e1
estat dwatson	Статистика Дарбина-Уотсона на наличие автокорреляции 1-го порядка
Число ПИ	_pi
dfuller y	Тест Дики-Фуллера для y
arima y, arima(1,1,1) arima y, ar(1) ma(1) arima y, ar(1 5) ma(1 3)	Оценивание arima-модели для y