

Семинар 8-9. Тесты единичного корня. Моделирование с помощью ARIMA.

План

1. Тесты единичного корня: DF, KPSS, PP.
2. Запись ARIMA-моделей.
3. Методология Бокса-Дженкинса.
4. Моделирование на основе 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.

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	∞
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

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

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

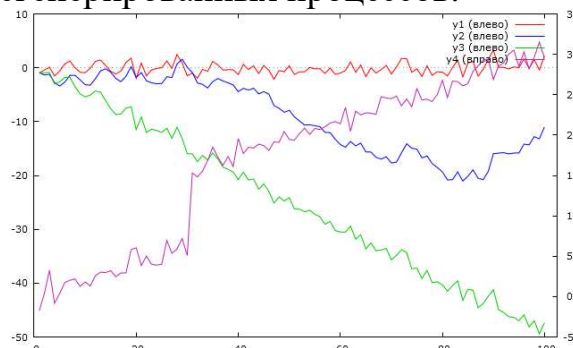
На уровне значимости 1% проверить гипотезу единичного корня.

Работа в Gretl.

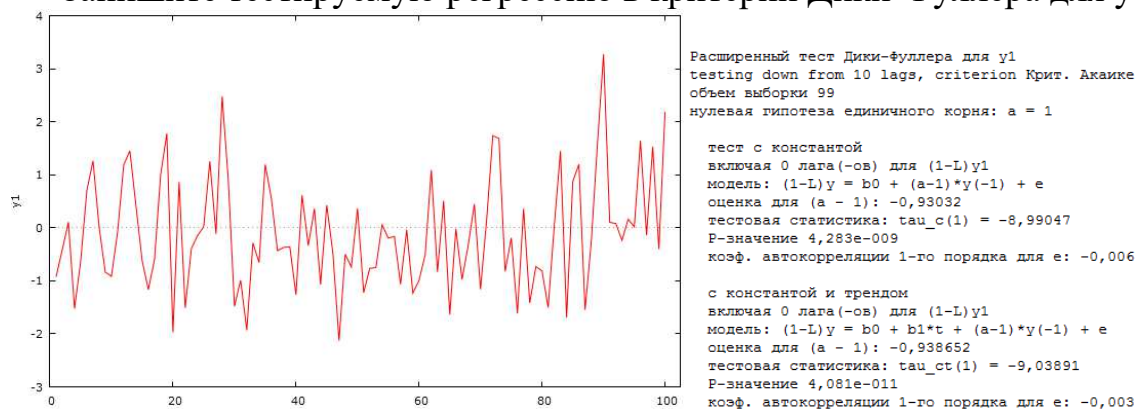
Задача 3. Даны случайные процессы $y_1, y_2, y_3, y_4, y_{21}$.

Файл: **DF.dta** откройте в Gretl

Исследуйте поведение сгенерированных процессов.



- 1.1. сформулируйте и проверьте гипотезу о наличии единичного корня. Запишите тестируемую регрессию в критерии Дики-Фуллера для y_1 .



- 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)
y_1						

Замечание. Не забывайте включать достаточное количество лагов в ADF-тест (в Gretl рассчитывается автоматически). Правило Шверта:

$$p_{\max} = \left\lceil 12 \left(\frac{T}{100} \right)^{1/4} \right\rceil$$

Schwert G.W. Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data // Journal of Monetary Economics. 1987. Vol. 20. P. 73–105.

Schwert. Test for Unit Roots: A Monte Carlo Investigation, JBES, 1989.

3.2. Разностно-стационарные ряды. Исследуйте y_2 и y_{21} . Используйте тест Дики-Фуллера для первой разности изучаемых процессов (в случае необходимости). Сделайте вывод о порядке **интегрируемости** процессов (после какой разности процесс стал стационарным).

Родионова Л.А. Майнор «Прикладной статистический анализ»
 Временные ряды и их практическое применение

Процесс	Уровни	Первая разность	Вторая разность	Порядок интегрируемости
y2				
y21				

3.3. Тренд-стационарный ряд. Исследуйте y_3 , предположив наличие в процессах детерминированного тренда, проведите тест Дики-Фуллера. Сделайте вывод.

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: $\tau_{a-1}(1) = -1,18331$
Р-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: $\tau_{a-1}(1) = -0,571397$
Р-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: $\tau_{a-1}(1) = -0,751946$
асимпт. р-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: $\tau_{a-1}(1) = -9,03891$
Р-значение 4,08e-011
коэф. автокорреляции 1-го порядка для e : -0,003

3.4. Альтернативные тесты единичного корня. Проведите PP- и KPSS-тесты для y_1 , сравните результаты.

- Какие тесты единичного корня Вы знаете? Проведите PP- и KPSS-тесты для y_1 , 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: a = 1</p> <p>тест с константой включая 0 лага(-ов) для (1-L)y1 модель: (1-L)y = b0 + (a-1)*y(-1) + e оценка для (a - 1): -0,93032 тестовая статистика: тау_ct(1) = -8,99047 Р-значение 4,283e-009 коэф. автокорреляции 1-го порядка для e: -0,006</p> <p>с константой и трендом включая 0 лага(-ов) для (1-L)y1 модель: (1-L)y = b0 + b1*t + (a-1)*y(-1) + e оценка для (a - 1): -0,938652 тестовая статистика: тау_ct(1) = -9,03891 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: Z_t = -8,97733 (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 > .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	R	Stata	Gretl
Тест Дики-Фуллера для y	adf.test(y) adf.test(y, k=5)	dfuller y dfuller y, trend regress lags(2)	Menu path: /Variable/Unit root tests/Augmented Dickey-Fuller test
Тест KPSS для y	kpss.test(y)	kpss y kpss y, notrend auto	Расширенный тест Дики-Фуллера (ADF-тест) ADF-GLS тест KPSS тест
Тест Филиппса-Перрона для y	pp.test(y)	pperron y pperron y, regress	Levin-Lin-Chu test
Тест Эндрюса-Зивота		zandrews y zandrews y, lagmethod(BIC) graph	Частичное интегрирование Phillips-Perron test HEGY

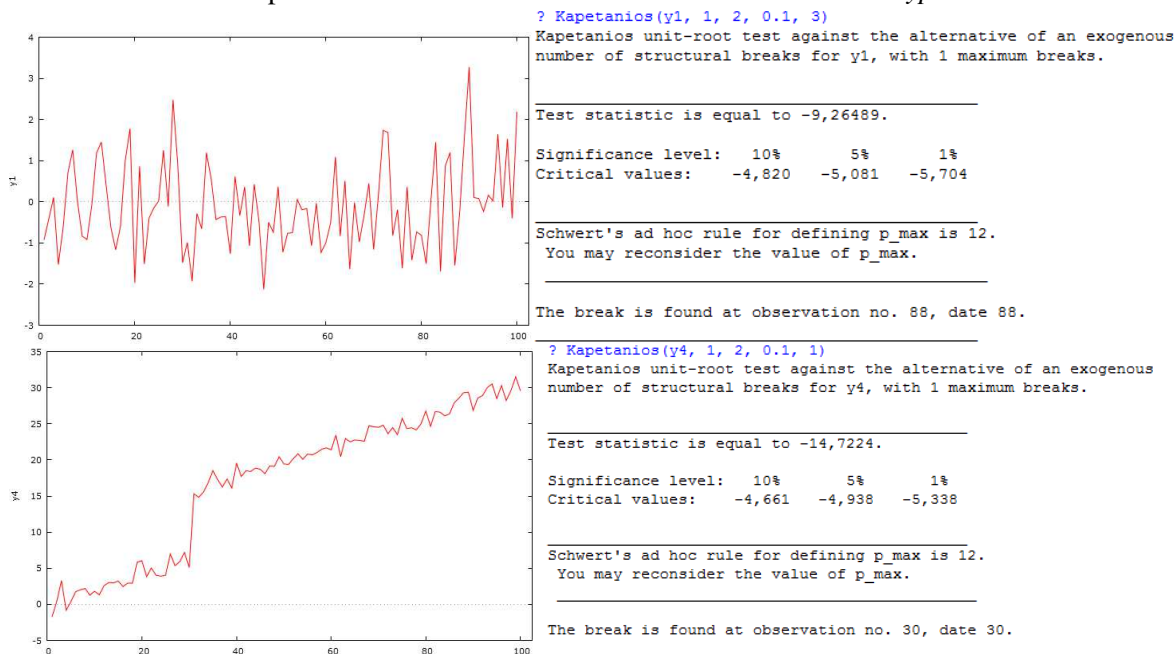
3.5. Тесты единичного корня с учетом структурного сдвига. Проведите тесты единичного корня для y_1, y_4 , в предположении наличия структурного сдвига (Gretl: Kapetanios' (2005) unit root test). Сделайте выводы.

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.

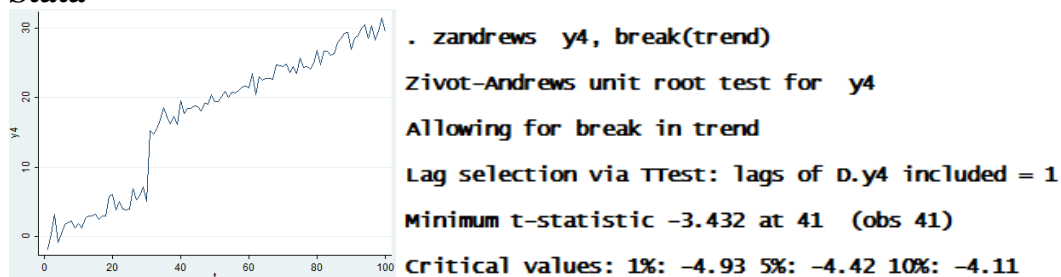
H0: DS

H1: TS

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



Stata



Рекомендуется по результатам теста делать сводную таблицу:

ВР	Тест	Нулевая гипотеза	Статистика критерия	р-значение	Вывод
y	DF(с трендом/без)				
	PP(с трендом/без)				
	KPSS(с трендом/без)				
Δy	DF(с трендом/без)				
	PP(с трендом/без)				
	KPSS(с трендом/без)				

Домашнее задание (ТДЗ) 8. Тесты единичного корня

По данным Всемирного банка выберите 3 показателя за 20-60 лет (опишите какие показатели были взяты для анализа, за какой период).

Файл: **WB_Russia.xls** (закладка **Data**)

Источник: <https://data.worldbank.org/country/russian-federation?locale=ru>

! можно взять свои данные по другой стране.

1. Опишите выбранные показатели.

показатель	
страна	
период	

Постройте графики выбранных показателей. Сделайте вывод о стационарности рядов, исходя из построенных графиков.

2. 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/альтернативная 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?

ARIMA-модели. Основные виды моделей:

Модель авторегрессии	ARMA(1,0)=AR(1) ARIMA(1,0,0)	$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$ или $(1 - \beta_1 L)y_t = \beta_0 + \varepsilon_t$
Модель скользящего среднего	ARMA(0,1)=MA(1) ARIMA(0,0,1)	$y_t = \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ или $y_t = \theta_0 + (1 + \theta_1 L)\varepsilon_t$
Модель авторегрессии – скользящего среднего	ARMA(1,1) ARIMA(1,0,1)	$y_t = \theta_0 + \beta_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ или $(1 - \beta_1 L)y_t = \theta_0 + (1 + \theta_1 L)\varepsilon_t$
модель авторегрессии – интегрированного скользящего среднего	ARIMA(1,1,1)	$\Delta y_t = \theta_0 + \beta_1 \Delta y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ или $(1 - \beta_1 L) \Delta y_t = \theta_0 + (1 + \theta_1 L)\varepsilon_t$

Методология Бокса-Дженкинса.

1. Идентификация модели.
 - Получение стационарного ряда.
 - Подбор процесса ARMA к ВР (выбор параметров ARIMA(p=?, d=?, q=?)).
2. Оценивание модели и проверка адекватности модели.
3. Прогнозирование.

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.

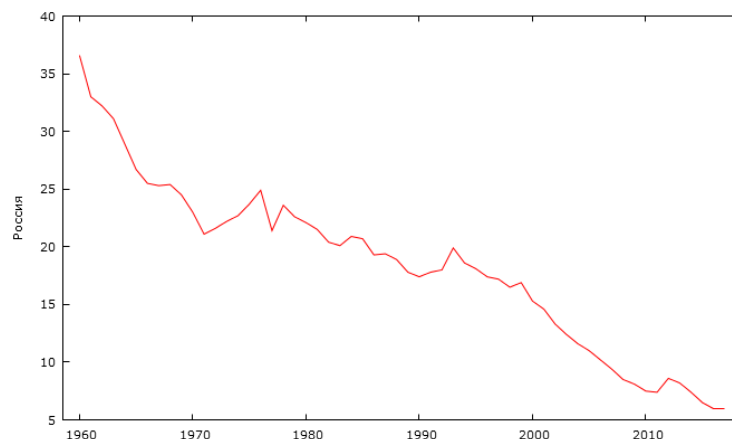
Важным показателем смертности и одновременно качества жизни является коэффициент младенческой смертности - число умерших в возрасте до 1 года в расчете на 1000 родившихся живыми. В отличие от общего коэффициента смертности, то поднимавшегося, то снижавшегося на протяжении двух последних десятилетий, 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%)^[12] и в 2012 году (на 17%)^[13]. 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. **Анализ графика.** Постройте график временного ряда младенческой смертности в России. Опишите динамику, сделайте вывод о стационарности ряда.



Анализ графика. Что необходимо заметить на графике?

- наличие тренда, его характер (убывающий/возрастающий), вид (линейный/квадратичный);
- наличие сезонности;
- наличие структурных сдвигов;
- провести предварительный анализ стационарности ряда.

3. **Стационарность.** Проверим ряд на стационарность (анализ ACF и тест Дики-Фуллера). Сформулируйте и проверьте нулевую гипотезу в критерии Дики-Фуллера.

Тесты единичного корня	
H_0 «процесс нестационарный/характеристическое уравнение процесса содержит единичный корень»	
Тест	Статистика критерия
Тест Дики-Фуллера (ADF-test)	
Тест Филипса-Перрона (PP-test)	

- Постройте коррелограмму (график автокорреляционной и частной автокорреляционной функции). Сделайте вывод о стационарности ряда.
- 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?

4. **Порядок интегрируемости ряда.** Переходим к первым разностям и заново тестируем ряд на стационарность.

Вывод: после взятия первой разности процесс стал стационарным, $d=1$, $ARIMA(p=?, d=1, q=?)$).

Замечание. Можно использовать для сравнения критические значения DF-распределения
Односторонние критические значения статистики Дики-Фуллера (Магнус)

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

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

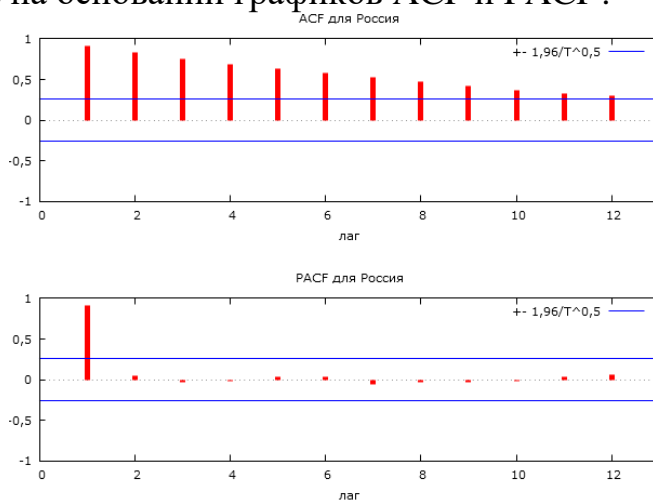
$$y_t = a_2 + b_3 y_{t-1} + c_3 t + \varepsilon_{3t} \quad (11.50)$$

Доверительный уровень	Размер выборки			
	25	50	100	∞
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).

5. Выбор параметров ARIMA(p=?, d=1, q=?). Анализ ACF/PACF

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?



Вывод: ACF – убывает, PACF – имеет пик на лаге 1 → AR(1) → p=1, q=0 → ARIMA(p=1, d=1, q=0) → ARIMA(1, 1, 0)

Поведение ACF и PACF ARMA-моделей

Вид модели	ACF	PACF
ARMA(1,0)=AR(1)	Экспоненциально убывает	Пик на лаге 1
ARMA(2,0)=AR(2)	Убывает	Пик на лаге 1,2
ARMA(0,1)=MA(1)	Пик на лаге 1	Экспоненциально убывает
ARMA(0,2)=MA(2)	Пик на лаге 1,2	Убывает
ARMA(1,1)	Убывает	Убывает

Замечание. ACF – автокорреляционная функция процесса; PACF - частная автокорреляционная функция процесса.

Этап 2. Оценивание модели.

4. Оцените и сравните несколько моделей ARIMA

- ARIMA (1, 1, 0) $(1-\alpha_1 L) \Delta y_t = \text{const} + \varepsilon_t$

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	

Как правильно записать модель?

Запись ARIMA(1,1,0) модели $(1-\alpha_1 L) \Delta y_t = \text{const} + \varepsilon_t$

Уравнение:

$\sigma=...$, AIC=..., BIC=...

Что можно сказать о качестве модели?

-запишите модель и 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 ARIMA	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)			

Информационный критерий	Формула расчета
Критерий Акаике (AIC) [Akaike (1973)]	$AIC(p, q) = \ln \hat{\sigma}^2 + 2 \frac{p+q}{T}$
Критерий Шварца (BIC) [Schwarz (1978)]	$BIC(p, q) = \ln \hat{\sigma}^2 + \ln T \frac{(p+q)}{T}$
Критерий Хеннана –Куинна (HQIC) [Hannan, Quinn (1979)]	$HQIC = \ln \hat{\sigma}^2 + \frac{2 \ln(\ln T)}{T} (p+q)$

Лучшая модель соответствует минимальному значению критериев Акаике и Шварца.

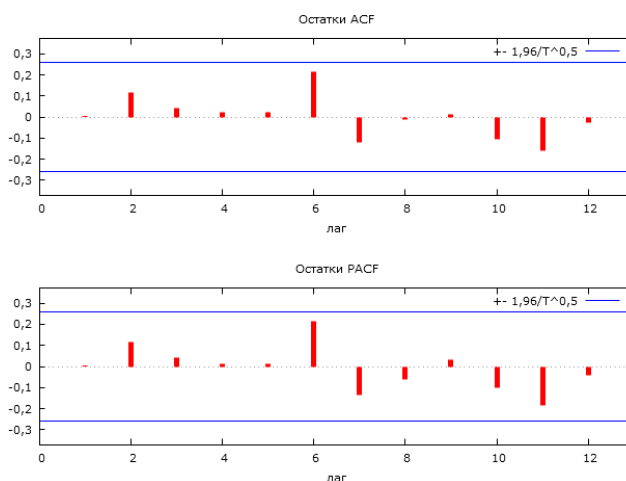
Этап 3. Диагностика моделей.

-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?

Тесты на автокорреляцию $H_0: \rho_1 = \dots = \rho_k = 0$ («автокорреляция до лага k отсутствует»)	
Тест	Статистика критерия
Тест Бройша-Годфри (Breusch-Godfrey); (не проводится для ARIMA)	
Тест Льюинга-Бокса (Ljung-Box)	



Тест на наличие автокорреляции до порядка 10

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

Нормальность.

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)				

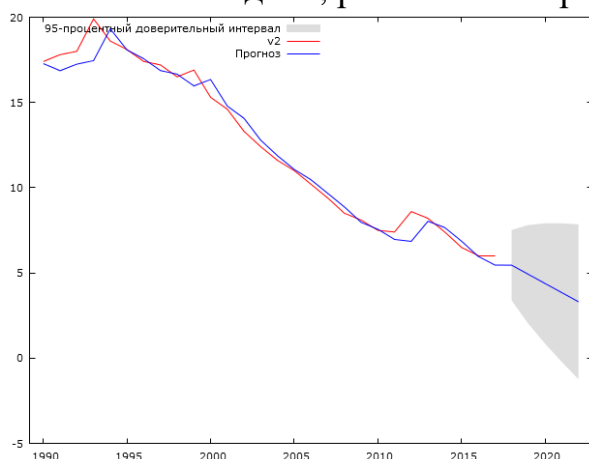
Этап 4. Прогнозирование.

По (наилучшей) модели строим прогноз, график и доверительный интервал.

$$\hat{y}_{T+1} = E\{y_{T+1} | y_1, \dots, y_T\} = E\{\alpha_0 + \alpha_1 y_T + \varepsilon_{T+1} | y_1, \dots, y_T\} = \alpha_0 + \alpha_1 y_T$$

$$\hat{y}_{T=2019} = \alpha_0 + \alpha_1 y_{T=2018} = \dots$$

- 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

Качество прогноза. Рассмотрим основные характеристики качества предсказания модели. Чем меньше значения этих характеристик, тем выше качество прогноза у полученной модели.

Характеристики точности прогноза в модели		
Характеристика	Формула вычисления	Единицы измерения
Средняя квадратичная ошибка	$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$	Зависит от единиц измерения показателя
Средняя абсолютная процентная ошибка	$MAPE = \frac{1}{n} \sum_{t=1}^n \left \frac{\hat{y}_t - y_t}{y_t} \right \cdot 100\%$	%
Средняя процентная ошибка	$MPE = \frac{1}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t} \cdot 100\%$	%

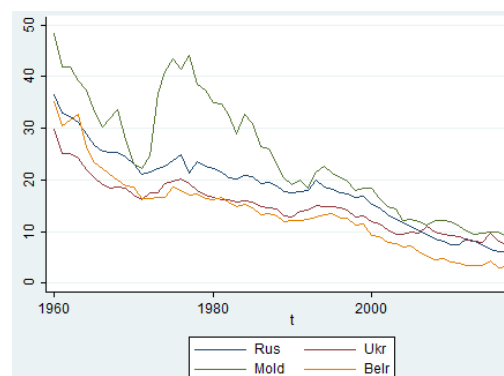
Статистика для оценки прогноза

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 (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 (самоконтроль)

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

Файл: **WB_Russia.xls** (закладка Data)

Источник: <https://data.worldbank.org/country/russian-federation?locale=ru>

!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 выбранные показатели.

показатель	
страна	
период	

Постройте графики выбранных показателей. Сделайте вывод о стационарности рядов, исходя из построенных графиков.

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	р-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?

- На основании ACF/PACF сделайте предположения о порядке ARIMA(p,d,q). Ответ обоснуйте.
- Оцените и сравните несколько ARIMA(p,d,q)-моделей: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1) и «своя модель» (обоснованная в п.3). 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).
- 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				

4. «Своя модель»				
------------------	--	--	--	--

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.