

Семинар 10. ARIMA-2

План

1. Вычисление прогнозов для ARIMA-процессов.
2. Моделирование на основе ARIMA: случай параболического тренда.
3. Устранение автокорреляции в остатках в ARIMA-моделях.
4. Структурные сдвиги в ARIMA-моделях.

Задание 1. Рассчитайте прогноз на 1-2 шаг вперед для ARIMA и 95% доверительный интервал:

1. $(1 - 0.1L)y_t = 1 + \varepsilon_t$, $T = 50$, $y_T = 10$, $y_{T-1} = 9$.
2. $(1 - 0.7L)\Delta y_t = 2 + \varepsilon_t + 0.1t$, $T = 100$, $y_T = 10$, $y_{T-1} = 8$.
3. $(1 + 0.3L)\Delta^2 y_t = 3 + (1 - 0.7L)\varepsilon_t$, $T = 150$, $y_T = 12$, $y_{T-1} = 6$.

Задание 2. Методология Бокса-Дженкинса

Исходные данные:

Файл: TFR_LE (Gretl).

Обозначение переменной	Описание
TFR	Коэффициент суммарной рождаемости в России (1990-2014 г.г.)
LE_M_m, LE_M_f	Ожидаемая продолжительность жизни при рождении в Москве (мужчины, женщины).
LE_r_m, LE_R_f	Ожидаемая продолжительность жизни при рождении в России (мужчины, женщины).

Данные Росстата <http://www.gks.ru>

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.

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.

Вопросы:

1. Как определить порядок интегрируемости ряда?
2. Как проверить адекватность модели?

Задание:

- Подберите подходящую модель ARIMA,
- обоснуйте выбор модели,
- запишите полученную модель (математическую модель, через лаговый оператор).
- Проверьте адекватность модели (стационарность, обратимость, анализ остатков),
- Оцените качество прогноза
- постройте прогноз на 6 лет (точечная и интервальная оценка).

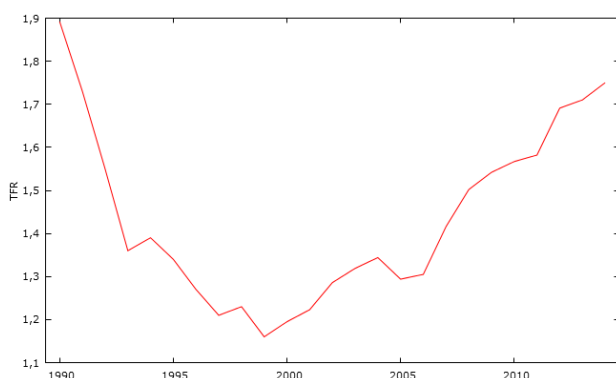
Задание 3. Параболический тренд. TFR - Коэффициент суммарной рождаемости в России (1990-2014 г.г.)

3.1. Проанализируйте стационарность ряда.

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

модели ARIMA	Ошибка модели, инф.критерии	Стационарность, обратимость процесса	Анализ остатков (автокорреляция, нормальность)	Качество прогноза
1. ARIMA(p,1,q)				
2. ARIMA(p,2,q)				
3. ARIMA(p,1,q)+лин.тренд				
4. ARIMA(p,0,q)+параб тренд				

Подсказка.

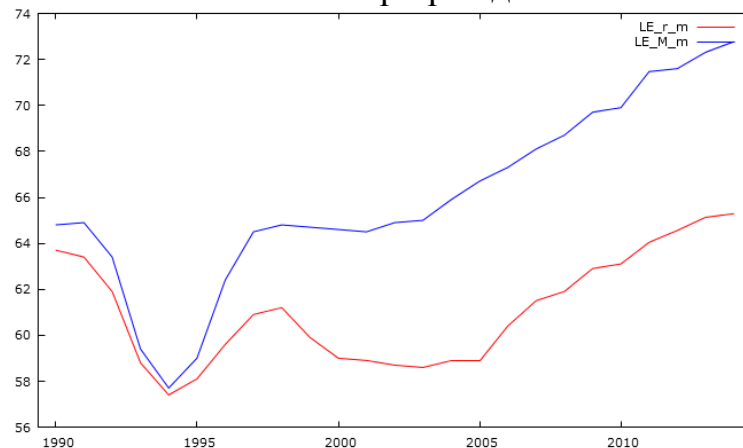


Модель 8: ARIMA, использованы наблюдения 1991-2014 (T = 24)
Оценено при помощи фильтра Калмана (Kalman) (точный метод МП)
Зависимая переменная: (1-L) TFR
Стандартные ошибки рассчитаны на основе Гесса

	Коэффициент	Ст. ошибка	z	P-значение	
const	-0,0272319	0,0582723	-0,4673	0,6403	
phi_1	0,909270	0,131630	6,908	4,92e-012	***
theta_1	-0,484943	0,259550	-1,868	0,0617	*
Среднее зав. перемен	-0,005917	Ст. откл. зав. перемен	0,082732		
Среднее инноваций	0,011308	Ст. откл. инноваций	0,063833		
Лог. правдоподобие	31,55210	Крит. Акаике	-55,10421		
Крит. Шварца	-50,39199	Крит. Хеннана-Куинна	-53,85406		
Действ. часть Миним. часть Модуль Частота					
AR					
Корень 1	1,0998	0,0000	1,0998	0,0000	
MA					
Корень 1	2,0621	0,0000	2,0621	0,0000	

Задание 4. Устранение автокорреляции в остатках. LE_M_m

- Ожидаемая продолжительность жизни при рождении в Москве (мужчины)



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

модели ARIMA	Ошибка модели, инф.критерии	Стационарность, обратимость процесса	Анализ остатков (автокорреляция, нормальность)	Качество прогноза
1. ARIMA(1,1,0)				
2. ARIMA(3,1,1) с 1-м и 3-м лагом				
3. на свой выбор				

4.2. Ожидаемая продолжительность жизни при рождении в России (мужчины) LE_r_m

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

модели ARIMA	Ошибка модели, инф.критерии	Стационарность, обратимость процесса	Анализ остатков (автокорреляция, нормальность)	Качество прогноза
1. ARIMA(0,2,0)				
2. ARIMA(1,0,0) с квадратичным трендом				
3. на свой выбор				

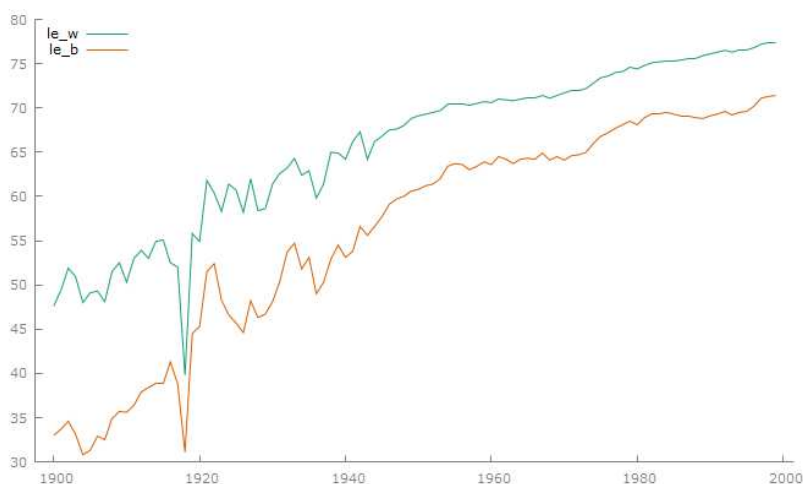
Вопрос: Согласно майским указам Президента (2018 г.) одной из национальных целей и стратегических задач развития РФ на период до 2024 г. является «повышение ожидаемой продолжительности жизни до 78 лет». Проверим реалистичность данной цели, используя статистические подходы.

Задача 5. Оценивание ARIMA со структурным сдвигом

Исходные данные: ОПЖ в США в период с 1900 по 1999

Файл с данными: LE_USA.gdt

Замечание. Необходимо учитывать структурный сдвиг 1918 года, связанный с эпидемией гриппа «Испанки».



Сравните ОПЖ белых и чернокожих. Какие изменения происходили на протяжении 20 века?

1. ОПЖ белых. Проведите тесты единичного корня с учетом структурных сдвигов, сравните с ADF/KPSS-тестом.

Расширенный тест Дики-Фуллера для le_w
testing down from 12 lags, criterion Крит. Акаике
объем выборки 89
нулевая гипотеза единичного корня: $a = 1$

тест с константой
включая 10 лагов(-ов) для $(1-L)le_w$
модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
оценка для $(a - 1)$: -0,0825419
тестовая статистика: $\tau_{c(1)} = -2,88849$
асимпт. р-значение 0,04668
коэф. автокорреляции 1-го порядка для e : 0,020
лаг для разностей: $F(10, 77) = 5,931 [0,0000]$

с константой и трендом
включая 10 лагов(-ов) для $(1-L)le_w$
модель: $(1-L)y = b_0 + b_1t + (a-1)y(-1) + \dots + e$
оценка для $(a - 1)$: -0,109297
тестовая статистика: $\tau_{ct(1)} = -0,903438$
асимпт. р-значение 0,9542
коэф. автокорреляции 1-го порядка для e : 0,021
лаг для разностей: $F(10, 76) = 3,153 [0,0020]$

with constant, linear and quadratic trend
включая 3 лагов(-ов) для $(1-L)le_w$
модель: $(1-L)y = b_0 + b_1t + b_2t^2 + (a-1)y(-1) + \dots + e$
оценка для $(a - 1)$: -0,917252
тестовая статистика: $\tau_{ctt(1)} = -5,897$
асимпт. р-значение 1,349e-005
коэф. автокорреляции 1-го порядка для e : 0,014
лаг для разностей: $F(3, 89) = 2,571 [0,0592]$

Расширенный тест Дики-Фуллера для d_le_w
testing down from 12 lags, criterion Крит. Акаике
объем выборки 89
нулевая гипотеза единичного корня: $a = 1$

тест с константой
включая 9 лагов(-ов) для $(1-L)d_le_w$
модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
оценка для $(a - 1)$: -4,5711
тестовая статистика: $\tau_{c(1)} = -5,50955$
асимпт. р-значение 1,651e-006
коэф. автокорреляции 1-го порядка для e : 0,041
лаг для разностей: $F(9, 78) = 2,897 [0,0052]$

Каретанос unit-root test

Замечание. The trimming parameter, this is, the number of obs. (as % of the sample size) that must be forced between breaks and between break and end-points of the sample size.

? Каретанос(le_w , 1, 1, 0.1, 1)

Каретанос unit-root test against the alternative of an exogenous number of structural breaks for le_w , with 1 maximum breaks.

Test statistic is equal to **-8,24863.**

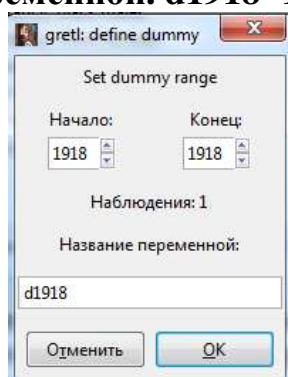
Significance level:	10%	5%	1%
Critical values:	-4,661	-4,938	-5,338

Schwert's ad hoc rule for defining p_max is 12.
You may reconsider the value of p_max .

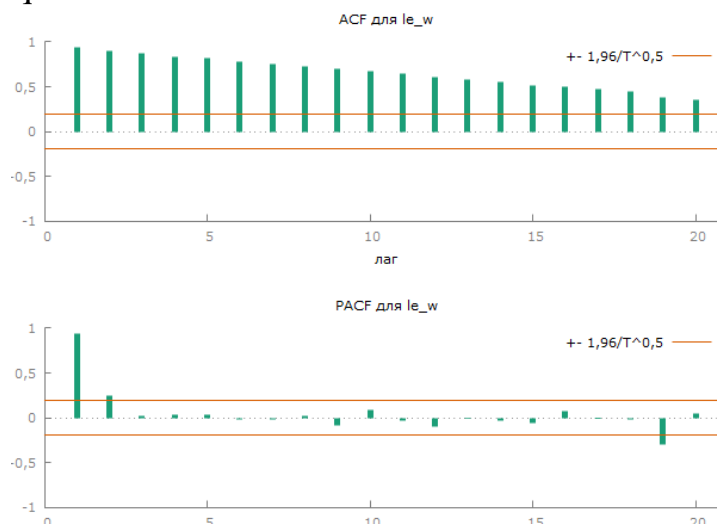
The break is found at observation no. 19, date 1918.

Можно ли предположить, что ряд является трендстационарным? Какая точка структурного сдвига?

2. Создание фиктивной переменной. $d_{1918}=1$, если (obs=1918).



3. На основе коррелограммы сделайте предположение относительно параметров p и q .



4. Рассмотрите и сравните следующие модели:

модели ARIMA	Ошибка модели, инф.критерии	Стационарность, обратимость процесса	Анализ остатков (автокорреляция, нормальность)	Качество прогноза
1. ARIMA(p,1,q) (свой выбор) обосновать				
2. ARIMA(p,1,q) с квадратичным трендом				
3. ARIMA(p,1,q) со структурным сдвигом				
4. ARIMA(3,1,(3)) +d1918				

Запишите модель:

ARIMA(3,1,(3)) +d1918

Модель 12: ARMAX, использованы наблюдения 1901-1999 (T = 99)

Estimated using AS 197 (точный метод МП)

Зависимая переменная: (1-L) le_w

Стандартные ошибки рассчитаны на основе Гесса

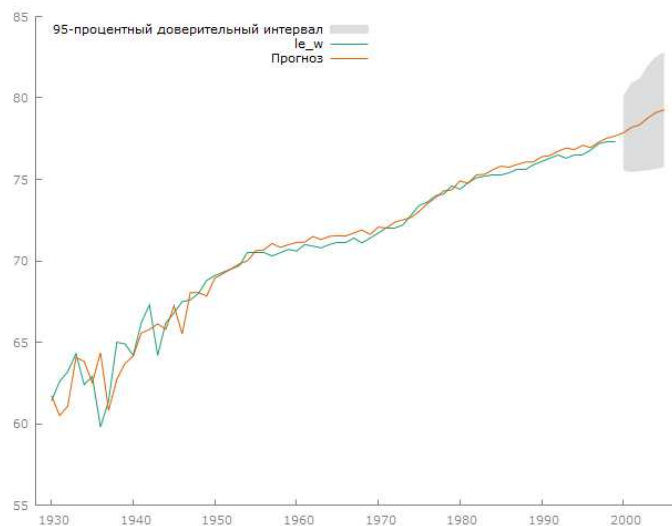
	Коэффициент	Ст. ошибка	z	P-значение	
const	0,293191	0,0568301	5,159	2,48e-07	***
phi_1	-0,354195	0,0934691	-3,789	0,0002	***
phi_2	-0,452402	0,0908780	-4,978	6,42e-07	***
phi_3	0,393878	0,102839	3,830	0,0001	***
theta_3	-0,331820	0,114533	-2,897	0,0038	***
d1918	-16,9701	0,922775	-18,39	1,57e-075	***

Среднее зав. перемен	0,300000	Ст. откл. зав. перемен	2,510549
Среднее инноваций	0,009951	Ст. откл. инноваций	1,176271
R-квадрат	0,983959	Испр. R-квадрат	0,983277
Лог. правдоподобие	-157,1139	Крит. Акаике	328,2278
Крит. Шварца	346,3936	Крит. Хенна-Куинна	335,5777

	Действ. часть	Мним. часть	Модуль	Частота
AR				
Корень 1	-0,4908	-0,9752	1,0917	-0,3242
Корень 2	-0,4908	0,9752	1,0917	0,3242
Корень 3	2,1302	0,0000	2,1302	0,0000
MA				
Корень 1	-0,7222	1,2509	1,4444	0,3333
Корень 2	-0,7222	-1,2509	1,4444	-0,3333
Корень 3	1,4444	0,0000	1,4444	0,0000

5. **auto_arima.** Используйте встроенную процедуру **auto_arima** для подбора параметров p и q .

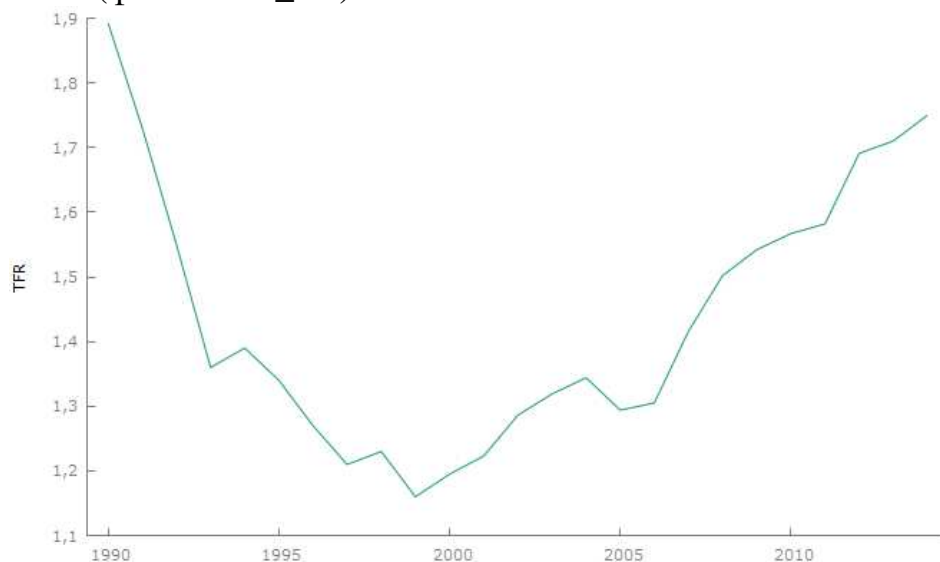
6. **Прогноз.** Постройте в одной системе координат исходный график и предсказанные по модели.



Замечание. Свойства модели улучшатся, если ограничить диапазон наблюдений $t > 1945$.

Задача 6. (самостоятельно) Анализ структурных сдвигов: суммарный коэффициент рождаемости (TFR).

Gretl (файл: TFR_LE).

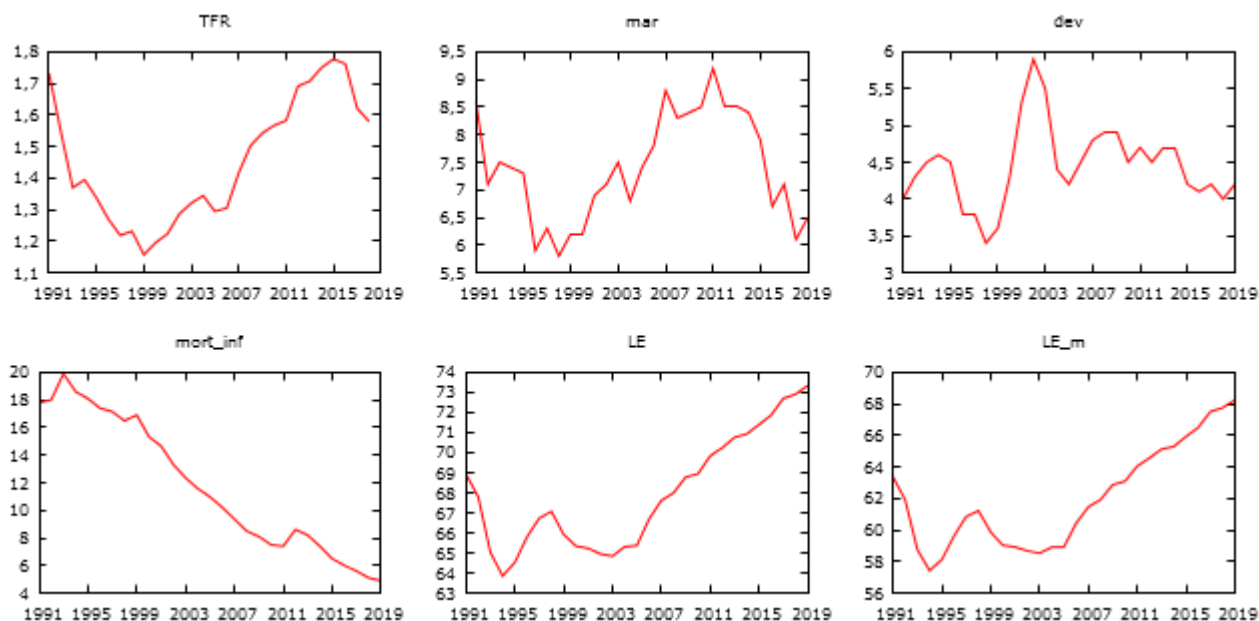


Анализ структурных сдвигов.

1. **T=2006.** С целью оказания дополнительной материальной поддержки российских семей, имеющих двух детей и более, по инициативе Президента РФ Владимира Путина в 2006 г. был принят федеральный закон от 29.12.2006 № 256-ФЗ, положивший начало новой федеральной программе материнского капитала, которая призвана создать «условия, обеспечивающие этим семьям достойную жизнь». (<http://materinskiy-kapital.molodaja-semja.ru/programma/>)
2. **T=1999.** С чем можно связать структурное изменение в данных с 1999 г.?

Задание. Проверьте влияние программы «материнский капитал» на TFR путем введения фиктивной переменной и постройте прогноз на основе ARIMA.

5. Самостоятельная работа. Файл: Russia_dem



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

модели ARIMA	Значимость коэфф-тов	Ошибка модели, инф.критерии	Стационарность, обратимость процесса (корни)	Анализ остатков (автокорреляция, нормальность)	Характеристики качество прогноза
1. ARIMA (свой выбор) обосновать					
2. ARIMA с квадратичным/лин трендом					
3. ARIMA со структурным сдвигом					

ТДЗ нет. Подготовка СР1 (см задание в ЛМС)