

План занятия

1. Запись SARIMA-моделей
2. SARIMA: оценивание и представление результатов.
3. Прогнозирование на основе моделей SARIMA.

Задание 1. Запись SARIMA-моделей. Запишите модели в аддитивной и мультипликативной форме

SARIMA(0,1,2) (1,1,3)₄

SARIMA(2,2,1) (1,1,2)₁₂

SARIMA(1,0,1) (3,2,2)₄

Пример. SARIMA(1,1,1) (1,1,1)₁₂

SARIMA(1,1,1)

Несезонная часть
(p,d,q)

p – несезонный порядок AR-части

d – несезонный порядок интегрируемости

q – несезонный порядок MA-части

(1,1,1)₁₂ - период сезонности

Сезонная часть

(P,D,Q)

P – сезонный порядок AR-части

D – сезонный порядок интегрируемости

Q – сезонный порядок MA-части

Мультипликативная модель	$(1 - \alpha_1 L)(1 - \alpha_{12} L^{12}) \Delta \Delta_{12} y_t = (1 + \theta_1 L)(1 + \theta_{12} L^{12}) \varepsilon_t$
Аддитивная модель	$(1 - \alpha_1 L - \alpha_{12} L^{12}) \Delta \Delta_{12} y_t = (1 + \theta_1 L + \theta_{12} L^{12}) \varepsilon_t$

Задание 2. Прогноз по SARIMA-моделям.

2.1. Запишите

$$\Delta_4 y_t =$$

$$\Delta \Delta_4 y_t = \Delta^2 \Delta_4 y_t = \Delta^2 \Delta_4^2 y_t =$$

$$\Delta \Delta_{12} y_t =$$

2.2. Рассчитайте прогноз на 1 шаг для моделей

$$(1) \quad (1 - 0.5L^4) \Delta \Delta_4 y_t = 2 + (1 - 0.3L^4) \varepsilon_t$$

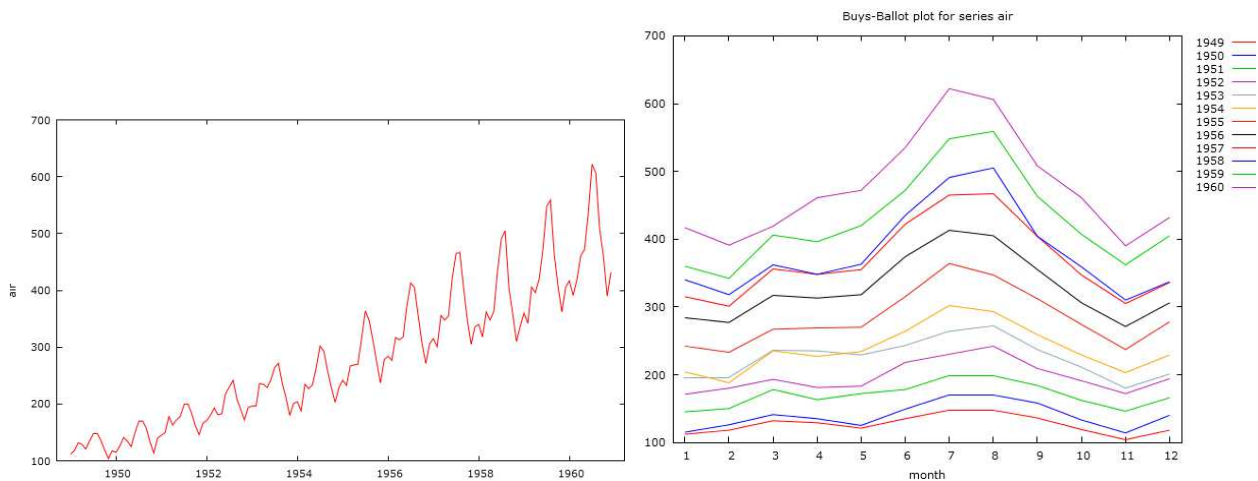
$$(2) \quad (1 - 0.7L)(1 - 0.1L^{12}) \Delta \Delta_{12} y_t = 1 + (1 - 0.9L) \varepsilon_t$$

Задание 3. GRETЛ. Исходные данные: air

Исходные данные: «Авиаперевозки пассажиров».

Выборка содержит ежемесячные данные об авиаперевозках пассажиров в период с 1949 по 1960.

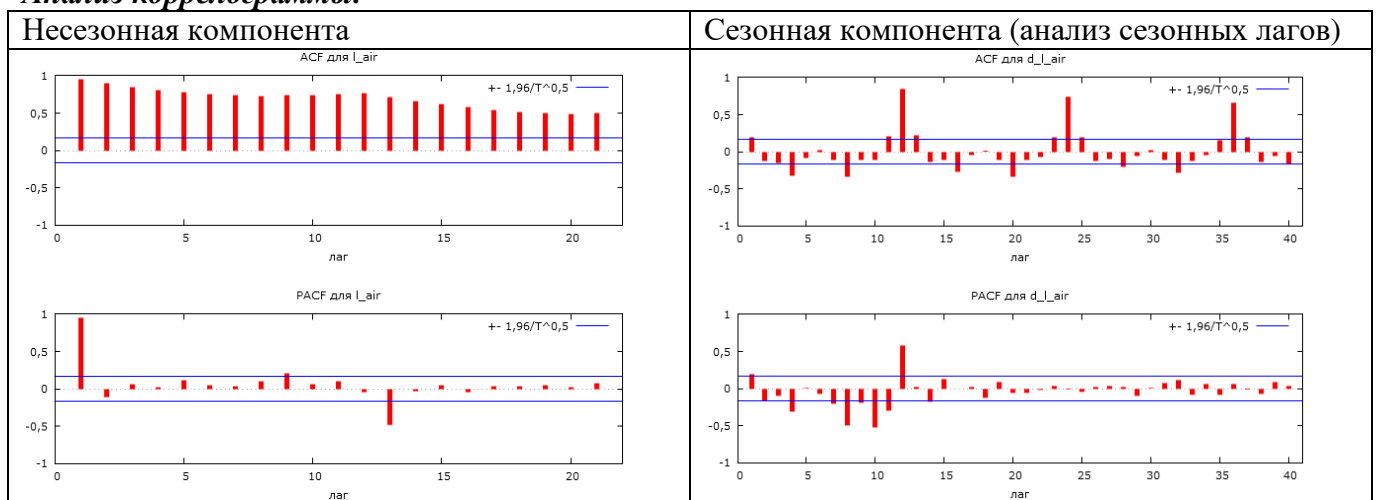
time	Time (in months)
air	Airline Passengers (1949-1960)
t	T=1,...,144



1. **Порядок несезонной/сезонной интегрируемости.** Для построения SARIMA-модели необходимо определить порядок несезонной/сезонной интегрируемости в модели $I(d)$, $I(D_s)$? Проведите серию тестов HEGY.

2. **Порядок AR/MA-частей.** На основе анализа коррелограмм сделайте предположение о порядках несезонных/сезонных AR/MA частей.

Анализ коррелограммы:



Ранее были оценены модели:

- **arima(0,0,1)** с линейным трендом (параболическим трендом)
- **arima(0,1,1)** с фиктивными сезонными переменными
- **arima(0,0,1)** с линейным трендом и с фиктивными сезонными переменными

4. **Sarima-модели.** Оцените и запишите модели в аддитивной и мультипликативной форме.

- **sarima(1,1,0)** $(1,1,0)_{12}$
- **sarima(0,1,1)** $(0,1,1)_{12}$
- **sarima(1,1,1)** $(1,1,1)_{12}$

Мультипликативная **sarima(1,1,0)** $(1,1,0)_{12}$

Аддитивная **sarima(1,1,0)** $(1,1,0)_{12}$

Модель 2: ARIMA, использованы наблюдения 1950:02–1960:12 (T = 131)
 Оценено при помощи фильтра Кальмана (Kalman) (точный метод МП)
 Зависимая переменная: (1-L) (1-Ls) 1 air
 Стандартные ошибки рассчитаны на основе Гессииана

	Коэффициент	Ст. ошибка	z	P-значение
const	6,40747e-05	0,00171181	0,03743	0,9701
phi_1	-0,374462	0,0808196	-4,633	3,60e-06 ***
Phi_1	-0,463689	0,0807987	-5,739	9,53e-09 ***
Среднее зав. перемен	0,000291	Ст. откл. зав. перемен	0,045848	
Среднее инноваций	0,000944	Ст. откл. инноваций	0,038167	
Лог. правдоподобие	240,4071	Крит. Акаике	-472,8142	
Крит. Шварца	-461,3134	Крит. Хеннана-Куинна	-468,1409	
	Действ. часть	Мним. часть	Модуль	Частота
AR				
Корень 1	-2,6705	0,0000	2,6705	0,5000
AR (сезонные)				
Корень 1	-2,1566	0,0000	2,1566	0,5000

Модель 3: ARIMA, использованы наблюдения 1950:02–1960:12 (T = 131)
 Оценено при помощи фильтра Кальмана (Kalman) (точный метод МП)
 Зависимая переменная: (1-L) (1-Ls) 1 air
 Стандартные ошибки рассчитаны на основе Гессииана

	Коэффициент	Ст. ошибка	z	P-значение
const	9,83130e-05	0,00200764	0,04897	0,9609
phi_1	-0,311141	0,0727461	-4,277	1,89e-05 ***
phi_12	-0,411006	0,0759387	-5,412	6,22e-08 ***
Среднее зав. перемен	0,000291	Ст. откл. зав. перемен	0,045848	
Среднее инноваций	0,000775	Ст. откл. инноваций	0,038640	
Лог. правдоподобие	238,9818	Крит. Акаике	-469,9637	
Крит. Шварца	-458,4629	Крит. Хеннана-Куинна	-465,2904	
	Действ. часть	Мним. часть	Модуль	Частота
AR				
Корень 1	-1,0112	0,2603	1,0441	0,4599
Корень 2	-1,0112	-0,2603	1,0441	-0,4599
Корень 3	1,0639	-0,2915	1,1031	-0,0426
Корень 4	1,0639	0,2915	1,1031	0,0426

Опишите полученные модели:

- запишите математическую форму моделей через лаговый оператор (с учетом аддитивного и мультипликативного эффектов),
- оцените качество моделей,
- рассчитайте информационные критерии, характеристики качества прогноза.
- рассчитайте предсказанные значения по модели и постройте совмещенные графики.

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

модели SARIMA	Инф.критерии, ошибка модели	Стационарность, обратимость	Анализ остатков	Общий вывод
1. sarima(0,1,1) (0,1,1) _{s=12} Аддитивная форма Мультипликативная форма				
2. sarima(1,1,0) (1,1,0) _{s=12} Аддитивная форма Мультипликативная форма				
3. sarima(1,1,1) (1,1,1) _{s=12} Аддитивная форма Мультипликативная форма				

2. Проверка адекватности модели.

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

- Обладают ли остатки свойствами белого шума?
- автокорреляция остатков. Тесты?
- нормальность. Тесты?

Альтернативные тесты на нормальность остатков в Gretl.

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

Тест на нормальное распределение `what1`:

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 \approx 0,12

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

3. Прогнозирование.

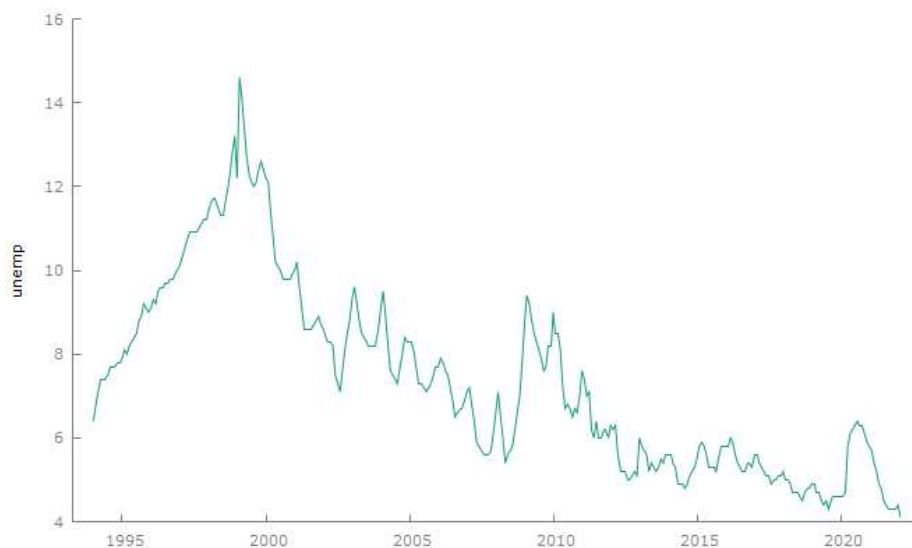
- По выбранной наилучшей модели постройте прогноз на 1 месяц вперед (аналитически)
- По выбранной наилучшей модели постройте прогноз на 12 месяцев вперед (в пакете). Постройте 95% доверительный интервал.
- Рассчитайте характеристики точности прогноза. Какие характеристики можно использовать?

4. (Самостоятельно на занятии)

Данные: Уровень безработицы в России (1994- 2022) *Файл:* UnempRus.gdt

Источник: <http://sophist.hse.ru/hse/nindex.shtml>

Уровень безработицы определяется как удельный вес численности безработных в численности экономически активного населения.



Постройте модели SARIMA с учетом сезонности. Какие модели Вы бы построили? Каким будет порядок несезонной/сезонной интегрируемости в модели $I(d)$, $I(D_s)$? Обоснуйте выбор порядков несезонных/сезонных p, d, q , детерминированных составляющих (тренда, фиктивных переменных). Проверьте адекватность полученной модели. Постройте прогноз.

Домашняя работа (ТДЗ) 14. SARIMA

Исходные данные: Число зарегистрированных браков в регионах (оперативные данные) в России по месяцам (2006- 2017; 2017-2021).

Файл: браки.xls

Источник: ЕМИСС <https://fedstat.ru/indicator/33553>

1	Число зарегистрированных браков (оперативные данные) (единица)						
2							
3		2006					
4		январь	февраль	январь-февраль	март	январь-март	апрель
5	Российская Федерация	55 509	62 449	117 958	70 798	188 756	86 055
6	Центральный федеральный округ	14 845	16 414	31 259	15 753	47 012	21 803
7	Белгородская область	686	751	1 437	639	2 076	809
8	Брянская область	471	625	1 096	554	1 650	665
9	Владимирская область	473	623	1 096	601	1 697	779
10	Воронежская область	945	1 042	1 987	785	2 772	1 254
11	Ивановская область	347	438	785	447	1 232	529
12	Калужская область	372	431	803	463	1 266	648
13	Костромская область	306	367	673	339	1 012	426

Замечание. Обратите внимание на формат представления данных на сайте Росстат. Данные необходимо предварительно преобразовать в «длинный» ряд, удалить лишние строки в Экселе.

!Можно взять свои данные по двум странам/регионам/городам и т.д. Не забудьте приложить файл с данными.

1. Описательный анализ. Выберите два региона для исследования и сравнения (северный южный/ западный восточный). Постройте графики рядов и опишите исходные данные. Исследуйте наличие сезонности в данных (график сезонной волны (в пакете), ACF/PACF, периодограмма), сравните региональные особенности. Приведите графики и дайте интерпретацию полученным результатам.

2. Порядок интегрируемости. С помощью HEGY-тест сделайте вывод о наличии сезонных/несезонных единичных корней и оцените порядок несезонной/сезонной разности (для двух регионов). Приведите результаты тестирования различных модификаций в виде таблицы, дайте интерпретацию результатам тестирования (для одного из регионов, по второму сделайте короткий вывод).

ВР	Тест	Нулевая гипотеза	Статистика критерия	p-значение	Вывод
у	HEGY (с трендом/без)				
	HEGY (с фикт пер/без)				
	HEGY (с трендом+ фикт пер)				
Δу	HEGY (без тренда)				
	HEGY (с фикт пер)				
	HEGY (с гарм пер)				
Δ _{су}	HEGY (с трендом/без)				
	HEGY (с фикт пер/без)				
ΔΔ _{су}	HEGY (с трендом/без)				
	HEGY (с фикт пер/без)				

3. Sarima. Оцените и запишите модели в аддитивной/мультипликативной (на выбор) форме (для **одного** из двух регионов)

- $\text{sarima}(0,1,1)(0,1,1)$

- $\text{sarima}(1,0,0) (1,1,0)$
- $\text{sarima}(1,1,1) (1,1,1)$
- предложите свою модель, задав порядки p, d, q (обычные и сезонные), включив по необходимости детерминированные составляющие (тренд, сезонные фиктивные переменные, учет структурного сдвига) произвольно, ответ обоснуйте.

Опишите полученные модели в виде сводной таблицы:

- запишите математическую форму моделей (с учетом аддитивного или мультипликативного эффектов),
- оцените качество моделей.

модели SARIMA	Инф.критерии, ошибка модели	Стационарность, обратимость	Анализ остатков	Общий вывод
Регион1				
1. $\text{sarima}(0,1,1) (0,1,1)_{s=12}$				
2. $\text{sarima}(1,1,0) (1,1,0)_{s=12}$				
3. $\text{sarima}(1,1,1) (1,1,1)_{s=12}$				
Наилучшая модель:				
Регион2				
Наилучшая модель:				

Для второго региона приведите одну наилучшую модель.

4. Предпосылка стационарности. Сформулируйте условие стационарности полученной модели (любой на выбор) (аналитически, через характеристическое уравнение).

5. Прогноз. По наилучшей модели постройте прогноз (точечную и интервальную оценку) в пакете, приведите графики (наблюдаемые+ предсказанные значения). Сделайте вывод, как будет меняться анализируемый показатель в двух регионах. Насколько реалистичный прогноз получился?

- Напишите решение задач (скан рукописного варианта по необходимости) и краткий отчет с выводами и полученными графиками, где это необходимо. Допускается сдача работы в группе по 2 человека (не забывайте указывать авторов).
- Выполненная домашняя работа загружается в SmartLMS. Срок выполнения – 1 неделя.

Задания для самоконтроля (сдавать не нужно!!).

1. Проиллюстрируйте схему теста HEGY для случая *ежемесячных данных* (запишите подробно все необходимые разложения с использованием теории приближенных функций, выведите подробно тестовое уравнение). Приведите все подробные пояснения.
2. Проверить стационарность процесса, используя для вычисления корней характеристического уравнения метод Феррари. Подробно описать схему метода и привести подробные расчеты.

$$y_t = 2 + y_{t-4} - y_{t-3} + y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$$