

Практическое занятие 14. «Сезонность3. SARIMA-модели»

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) (0,1,1)₄

SARIMA(1,1,1) (1,1,1)₆

SARIMA(2,1,1) (1,1,2)₁₂

SARIMA(2,1,2) (2,1,2)₄

Пример. SARIMA(1,1,1) (1,1,1)₁₂

SARIMA(1,1,1) **(1,1,1)₁₂** - период сезонности
 Несезонная часть Сезонная часть
 (p,d,q) (P,D,Q)
 p – несезонный порядок AR-части P – сезонный порядок AR-части
 d – несезонный порядок интегрируемости D – сезонный порядок интегрируемости
 q – несезонный порядок MA-части 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

$$\Delta_4 y_t =$$

$$\Delta \Delta_4 y_t =$$

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

2.2. Рассчитайте прогноз на 1 лаг для моделей

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

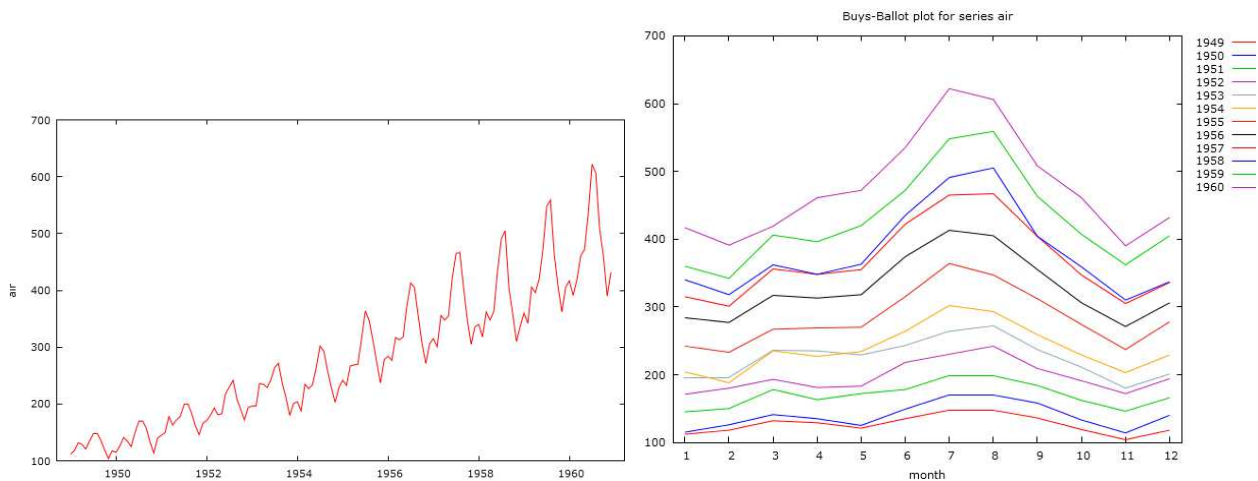
$$(2) \quad (1 - 0.7L)(1 - 0.1L^{12}) \Delta \Delta_{12} y_t = 1 + (1 - 0.9L) \varepsilon_t$$

Задание 3. GRETЛ. Исходные данные: air

Исходные данные: «Авиаперевозки пассажиров».

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

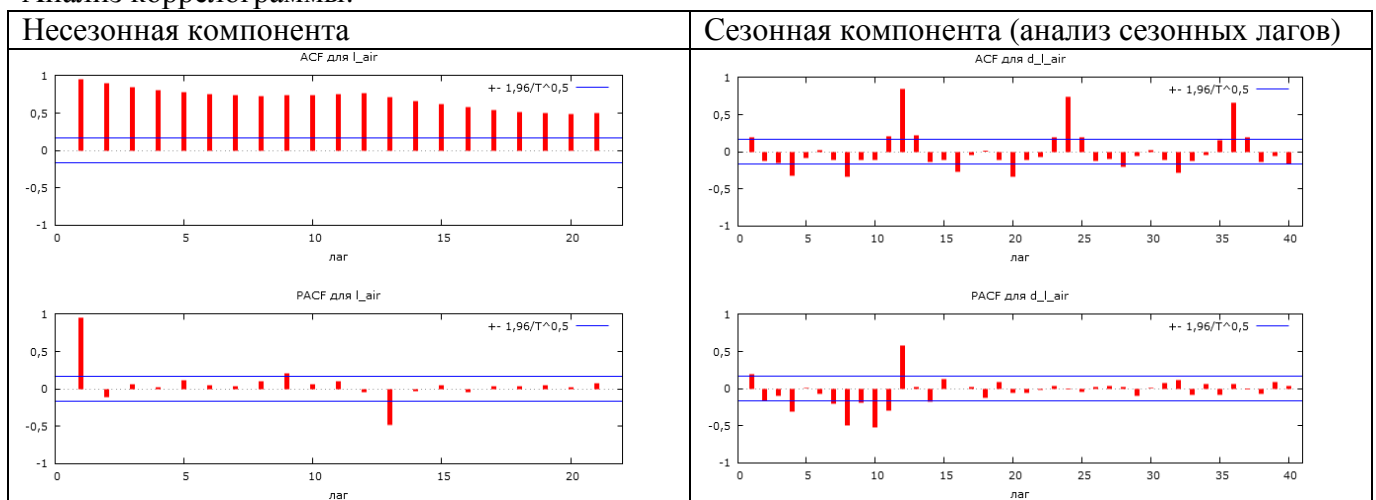
time	Time (in months)
air	Airline Passengers (1949-1960)
t	T=1,...,144



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

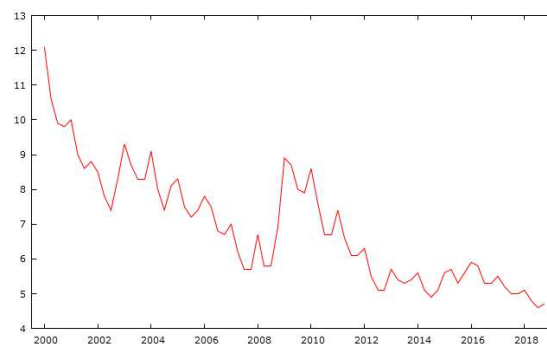
3. Прогнозирование.

- 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 (1 кв.2000- 4 кв.2018 гг.)

Файл: `unemp.gdt`



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) 13-14. SARIMA.

Исходные данные: Число зарегистрированных родившихся в регионах (оперативные данные) в России по месяцам (2006- 2018).

Файл: `рождаемость_регионы.xls` **Источник:** ЕМИСС

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

Период	2006	2007	2008	2009	2010
115 январь	109 422	125 368	141 262	137 503	132 371
114 февраль	114 069	117 186	132 405	133 343	135 015
M01_02 январь-	223 491	242 554	273 667	270 846	267 386
68 март	129 178	129 618	133 235	151 027	160 950
122 январь-март	352 669	372 172	406 902	421 873	428 336
61 апрель	113 392	116 527	140 213	142 828	143 644
117 январь-апрель	466 061	488 699	547 115	564 701	571 980
67 май	125 100	136 795	132 539	135 218	139 043
121 январь-май	591 161	625 494	679 654	699 919	711 023
66 июнь	124 238	127 509	131 845	149 348	157 813
120 январь-июнь	715 399	753 003	811 499	849 267	868 836
65 июль	125 927	142 686	163 520	163 309	158 207
119 январь-июль	841 326	895 689	975 019	1 012 576	1 027 043
60 август	140 228	149 912	148 284	151 739	159 610
116 январь-август	981 554	1 045 601	1 123 303	1 164 315	1 186 653
108 сентябрь	123 950	132 454	151 064	156 717	150 999

!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, о втором сделайте короткий вывод).

ВР	Тест	Нулевая гипотеза	Статистика критерия	р-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)

- **sarima(0,1,1)(0,1,1)**
- **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				
3. $\text{sarima}(0,1,1) (0,1,1)_{s=12}$				
4. $\text{sarima}(1,1,0) (1,1,0)_{s=12}$				
5. $\text{sarima}(1,1,1) (1,1,1)_{s=12}$				
Наилучшая модель:				
Регион2				
Наилучшая модель:				

Для второго региона приведите одну наилучшую модель.

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

5. По выбранной наилучшей модели (любой на выбор) постройте прогноз на 1 шаг вперед (аналитически, по формулам).

6. 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.

- Напишите решение задач (скан рукописного варианта) 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 LMS. 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.

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