

---

---

# Анализ и предсказание Временных рядов

---

---

Подготовили:

Захаров Владимир

Щёголев Михаил

Харазян Давид

Московский Государственный Университет  
имени Ломоносова

2019

Кафедра Исследования операций

# 1 Определения

- Временным рядом называется последовательность значений признака  $y$ , измеряемого через постоянные временные интервалы

$$y_1, ..., y_T, ..., y_t \in \mathbb{R}$$

- Анализ временных рядов совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования.
- Тренд — плавное долгосрочное изменение уровня ряда. Эту характеристику можно получить, наблюдая ряд в течение достаточно долгого времени.
- Сезонность — циклические изменения уровня ряда с постоянным периодом.
- Цикл — изменение уровня ряда с переменным периодом.
- Ошибка — непрогнозируемая случайная компонента ряда.
- Количественной характеристикой сходства между значениями ряда в соседних точках является автокорреляционная функция (или просто автокорреляция), которая задаётся следующим соотношением:

$$r_\tau = \frac{E((y_t - Ey)(y_{t+\tau} - Ey))}{D}$$

- Значимость автокорреляции вычисляется с помощью критерия Стьюдента:

временной ряд:	$y^T = y_1, ..., y_T$
нулевая гипотеза:	$H_0: r_\tau = 0$
альтернатива:	$H_1: r_\tau \neq 0$
статистика:	$T(y^T) = \frac{r_\tau \sqrt{T - \tau - 2}}{\sqrt{1 - r_\tau^2}}$
нулевое распределение:	$T(y^T) \sim St(T - \tau - 2)$

- Анализировать величину автокорреляции при разных значениях лагов удобно с помощью графика, который называется коррелограммой. По оси ординат на нём откладывается автокорреляция, а по оси абсцисс — размер лага  $\tau$ .
- Временной ряд  $y_1, ..., y_T$  называется стационарным, если  $\forall s$  (ширина окна) распределение  $y_t, ..., y_{t+s}$  не зависит от  $t$ , т.е. его свойства не зависят от времени.

- Формально гипотезу о стационарности можно проверить с помощью критерия Дики-Фуллера:

временной ряд:	$y^T = y_1, \dots, y_T$
нулевая гипотеза:	$H_0$ : ряд нестационарен
альтернатива:	$H_1$ : ряд стационарен
статистика:	DF-статистика
нулевое распределение:	табличное

- Дифференцирование - это переход к попарным разностям соседних значений:  $y' = y_t - y_{t-1}$
- Временной ряд имеет единичный корень, или порядок интеграции один, если его первые разности образуют стационарный ряд.
- Скользящая статистика - общее название для семейства функций, значения которых в каждой точке определения равны среднему значению исходной функции за предыдущий период. Скользящая статистика обычно используется с данными временных рядов для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов.
- Аддитивная модель имеет вид:  $Y = T + S + E$   
где  $T$  - компонента тренда,  $S$  - компонента сезонности,  $E$  - случайная компонента
- Мультипликативная модель имеет вид:  $Y = T * S * E$   
где  $T$  - компонента тренда,  $S$  - компонента сезонности,  $E$  - случайная компонента
- Временной ряд является интегрированным порядка  $k$ , если его разности порядка  $k$  образуют стационарный ряд.

## 2 Описание

Целью задачи является провести анализ временного ряда и попробовать предсказать значения для последующих месяцев.

1-й этап:

Для начала требуется проверить ряд на стационарность. Один из способов - это визуальная оценка путем рисования ряда и скользящей статистики (Рис. 1).

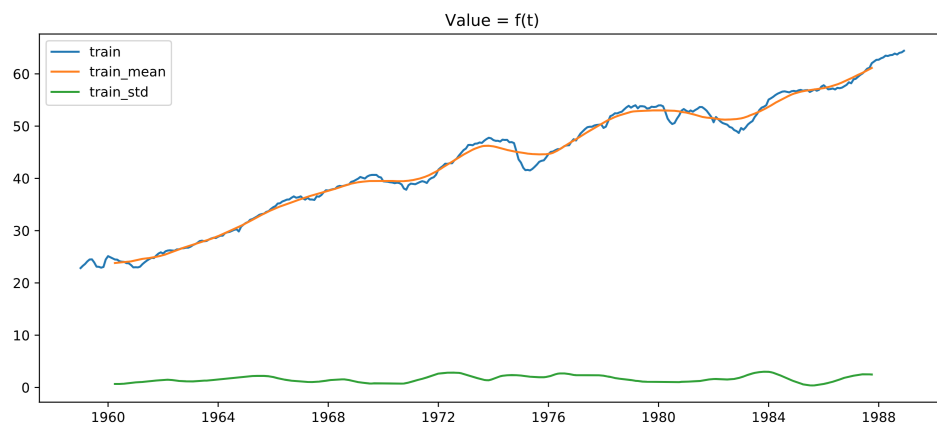


Рис. 1:

Как видим, посчитанные скользящие статистики показывают НЕстационарность ряда.

Проведем дифференцирование первого порядка. Получим следующий ряд (Рис. 2):

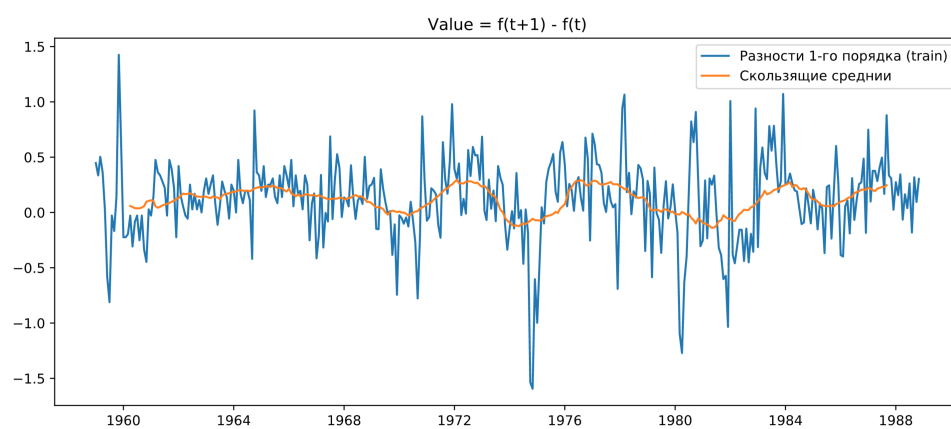


Рис. 2:

Наблюдаем стационарность разностей первого порядка данного ряда.

2-й этап:

Проведём разложение временного ряда на тренд и сезонность:

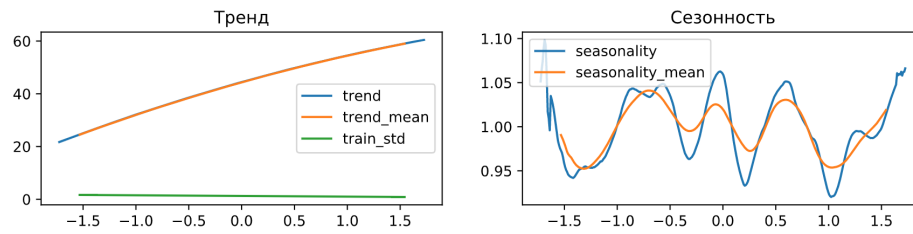


Рис. 3: Аддитивная модель

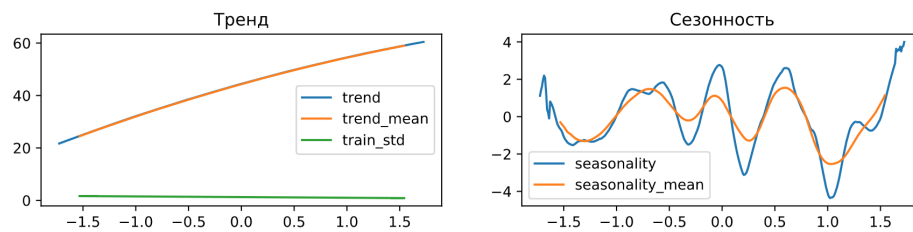


Рис. 4: Мультипликативная модель

Вывод: наблюдается тренд, что означает, что ряд не является стационарным

3-й этап:

Ряд является интегрированным порядка 1. Следовательно мы можем применить к нему модель ARIMA. Проведем отбор параметров. Для этого нарисует графики автокорреляции и функции частичной автокорреляции:

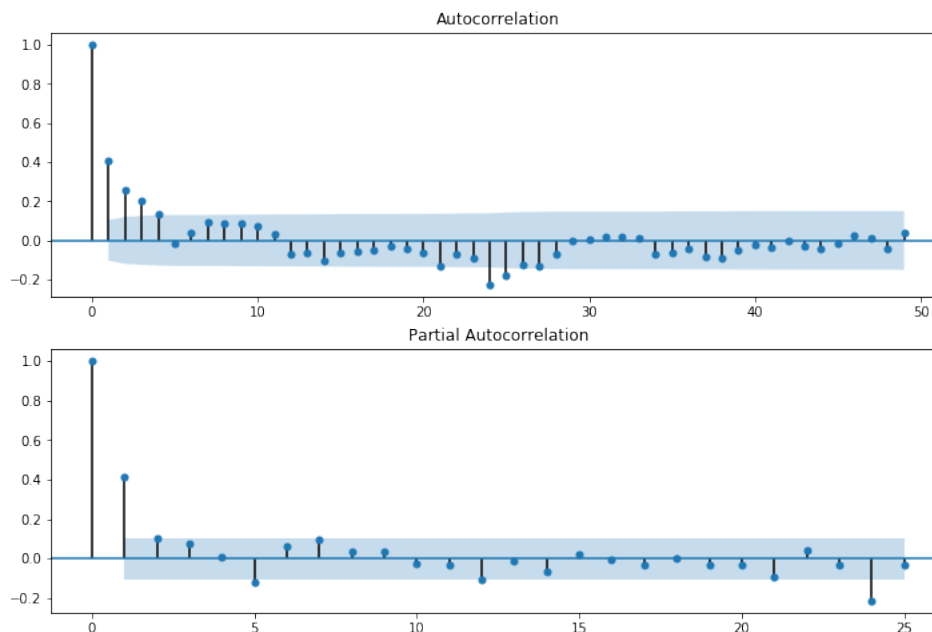


Рис. 5:

Начальное значение для параметра  $Q * S$  даёт номер последнего сезонного лага, при котором автокорреляция значима. В рассматриваемом примере сезонных лагов со значимой корреляцией нет, значит, начальное приближение  $Q = 0$ . Параметр  $q$  задаётся номером последнего несезонного лага, при котором автокорреляция значима. В данном случае можно взять начальное значение  $q = 3$ .

Значения параметров  $p, P$  подбираются с использованием не автокорреляционной функции, а частичной автокорреляционной функции. Частичная автокорреляция — это автокорреляция после снятия авторегрессии предыдущего порядка. Например, чтобы подсчитать частичную автокорреляцию с лагом  $\tau = 2$ , требуется построить авторегрессию порядка 1, вычесть эту авторегрессию из ряда и подсчитать автокорреляцию на полученных остатках.

Начальное приближение для параметра  $P * S$  задаёт номер последнего сезонного лага, при котором частичная

автокорреляция значима. В данных мы видим, что  $P = 0$   
 Аналогично,  $p$  задаётся как номер последнего несезонного лага, при котором частичная автокорреляция значима. В данном случае можно взять начальное приближение  $p = 1$ .  
 Теперь, отобрав параметры, проведём отбор лучшей модели по критерию Акаике:

$$AIC = -2\ln L + 2k$$

parameters	aic
(1, 1, 1)	255.641355
(1, 1, 2)	256.958699
(1, 1, 3)	258.926232
(1, 1, 0)	263.241050
(0, 1, 3)	266.970398

Как видно, критерий Акаике говорит, что лучшая модель, это модель с параметрами (1, 1, 1)

Построим модель с этими параметрами:

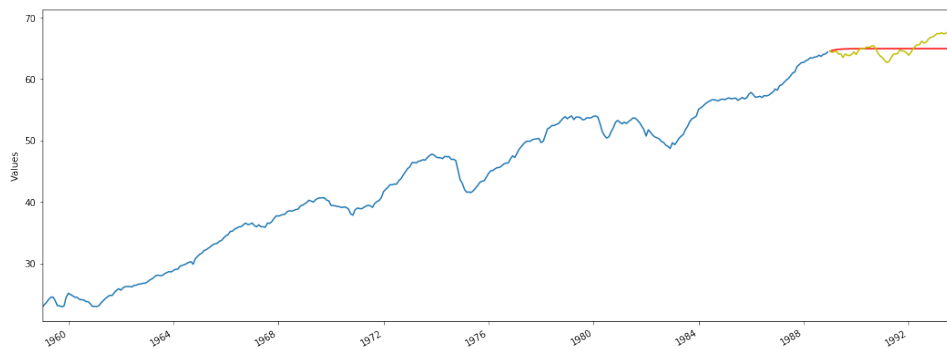


Рис. 6: ARIMA(1, 1, 1), тренировочная и тестовая выборка

Вывод: Модель ARIMA достаточно плохо моделирует наш временной ряд.

Для оценки качества модели используется метрика  $r^2 score$  библиотеки *sklearn*, получающая на вход предсказанные моделью значения и истинные значения ряда



### 3 Необходимое ПО

Библиотеки:

- warnings
- matplotlib.pyplot
- pandas
- statsmodels.api
- itertools
- product
- sklearn

## 4 Вклад участников

- Захаров Владимир - Проверка ряда на стационарность посредством визуализации статистик
- Щёголев Михаил - Разложение временного ряда на тренд, сезональность и шум , полная сборка программы и оформление пакета
- Харазян Давид - Построение прогнозирующей модели, составление ReadMe