

Теоретические сведения

Определение. *Временным рядом* называется последовательность измеренных через некоторые промежутки времени данные.

Определение. *Скользящим средним* будем называть среднее арифметическое значений исходной функции за установленный период

Определение. Временной ряд x_t называется *строго стационарным*, если совместное распределение вероятностей n наблюдений x_1, \dots, x_n такое же, как и n наблюдений x_{1+k}, \dots, x_{n+k} для любых n, k

Определение. *Автокорреляционной функцией* называют зависимость вида

$$\rho(k) = \frac{E[(x_t - a)(x_{t+k} - a)]}{\sigma(t)\sigma(t+k)}$$

Определение. Временной ряд называется *слабо стационарным* если его математическое ожидание и дисперсия не зависят от времени, и если ковариация между его значениями в моменты времени t и $t+s$ зависят только от s , но не от t .

Определение. *Аддитивная модель* – представление ряда в виде $Y = T + S + E$, где

T - трендовая составляющая

S - сезонная составляющая

E - остаток

Определение. *Мультипликативная модель* – представление ряда в виде $Y = TSE$

Описание задачи

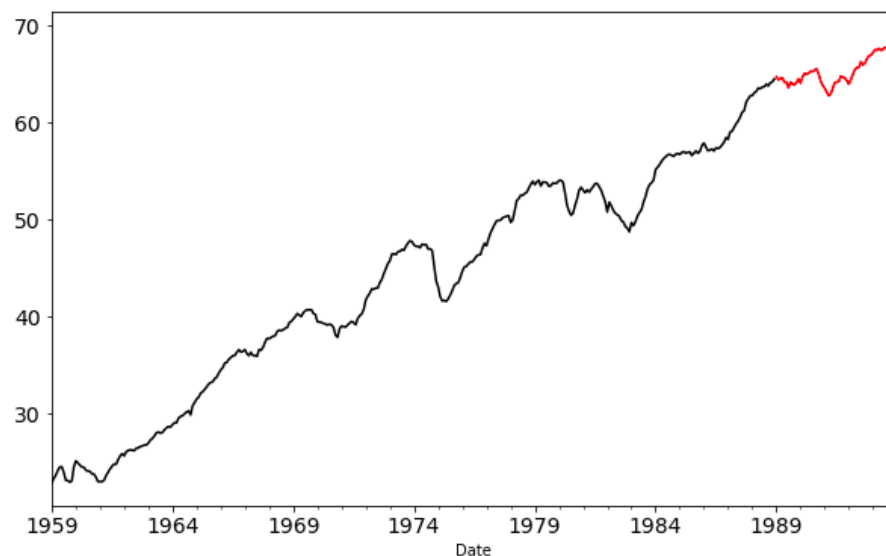
Задание состоит в:

- Проверке ряда на стационарность в широком смысле

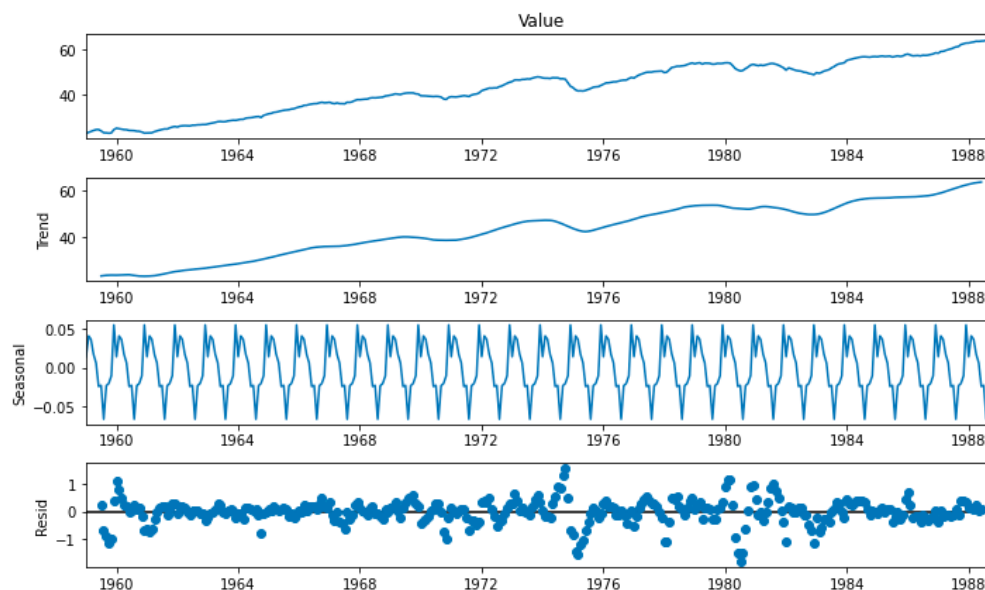
- Разложении ряда на тренд, сезонность, остаток в соответствии с мультипликативной и аддитивными моделями
- Визуализировании полученных рядов и оценке их стационарности
- Проверке ряда на интегрированность порядка k
- Применении к нему модели ARIMA
- Предсказании модели для тестовой выборки, вычислении r^2 score
- Выборе наилучшей модели с использованием информационного критерия Акаике

Подход к решению

1. С помощью matplotlib построим график временного ряда.



2. Разложим ряд на тренд, сезонность и остаток. Воспользуемся функцией `seasonal_decompose()` с параметром `model = 'additive'` (соотв. `model = 'multiply'` для мультипликативной модели). Из построенных графиков видно, что наблюдается тренд, что означает, что ряд не является стационарным.

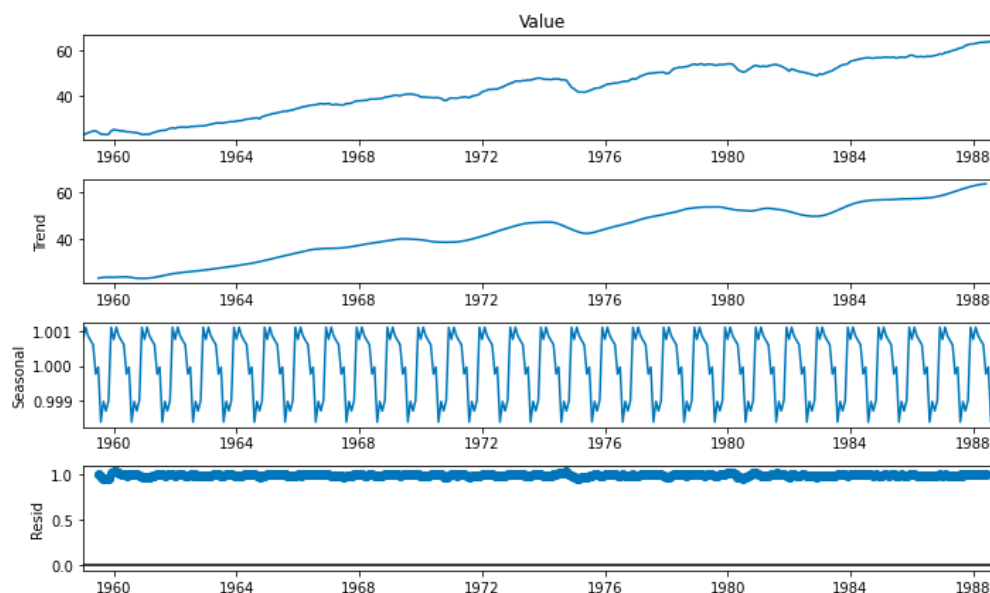


3. Проверим, является ли ряд интегрированным порядка k . Проведем обобщенный тест Дики-Фуллера на наличие единичных корней с помощью функции `adfuller()`

```
from statsmodels import adfuller
```

Имеем $k = 1$, т. е. ряд интегрируем

4. Перейдем к построению модели ARIMA. Для этого нужно подобрать параметры, определяющие порядок модели: p – порядок компоненты AR, d – порядок интегрированного ряда, q – порядок компоненты MA.



$d = 1$, остается определить p и q .

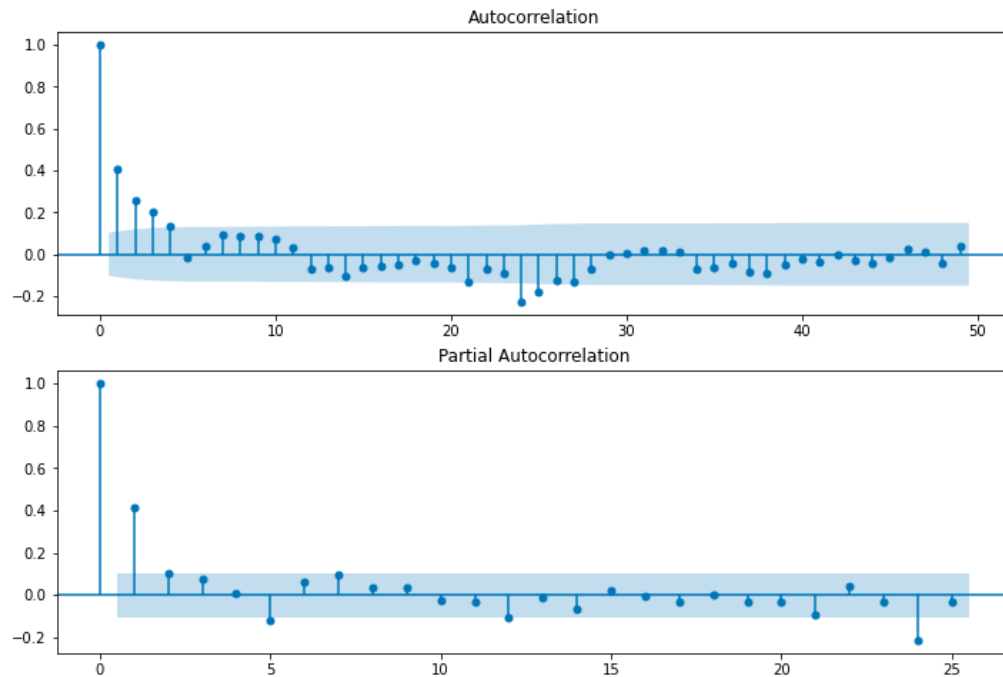
Построим автокорреляцию и частичную автокорреляцию ряда.

q определим из автокорреляции – по коррелограмме можно определить количество автокорреляционных коэффициентов, которые сильно отличаются от 0 в модели МА.

p определим из частичной автокорреляции – по ее коррелограмме можно определить максимальный номер коэффициента, сильно отличающегося от 0 в модели AR.

Для построения коррелограмм воспользуемся функциями `plotacf()` и `plotpacf()`

```
from statsmodels import plotacf, plotpacf
```



Получаем $p = 0$, $q = 3$

5. Построим теперь саму модель и осуществим прогноз. Построим график, на котором изображены данные из `testing.xlsx` и построенный прогноз. Вычислим коэффициент r_2 , чтобы определить процент наблюдений, описываемый данной моделью.

Системные требования

Необходимые библиотеки:

```
matplotlib
pandas
statsmodels
sklearn
pylab
```

Вклад участников в решение задачи

- Алексей Сомов – проверка ряда на стационарность, построение прогнозирующей модели, сборка программы
- Дмитрий Попов – написание Readme, разложение ряда на тренд, сезональность и шум
- Юлия Голубева – визуализация, тест Дики - Фуллера, сборка программы
- Алиса Боос – написание Readme, тест Дики - Фуллера, сборка программы
- Ли Юйтун – визуализация, сборка программы