

Анализ временных рядов • Time Series Analysis

- области применения методов анализа временных рядов
- базовые методы
- учет различных компонентов во временных рядах
- библиотеки для анализа временных рядов



- финансовая область
- мониторинг состояния оборудования
- климатические исследования
- естественно-научные области

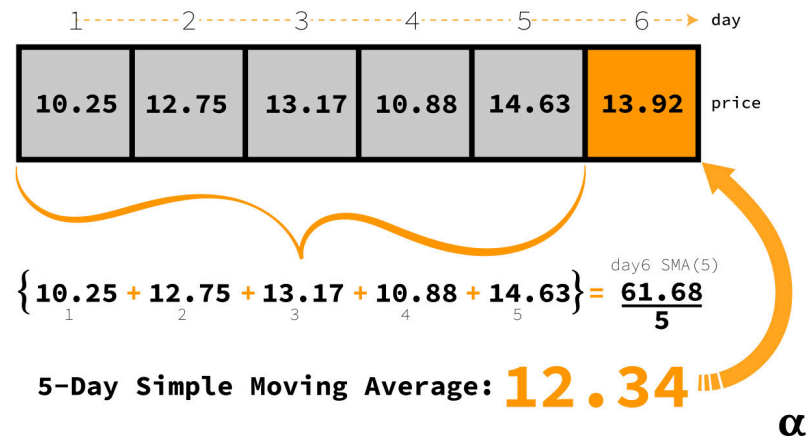
Естественным образом возникают в областях, где измерения носят интервальный характер.

Две самые типичные задачи: определение природы временного ряда и прогнозирование значения временного ряда

Скользящее среднее

$$\hat{y}_t = \frac{1}{k} \sum_{n=1}^k y_{t-n}$$

В качестве следующего значения используем среднее значение последних k наблюдений.



$$\hat{y}_t = \sum_{n=1}^k \omega_n y_{t-n}$$

В качестве следующего значения используем средне-взвешенное значение последних k наблюдений с ограничением $\sum_{n=1}^k w_n = 1$

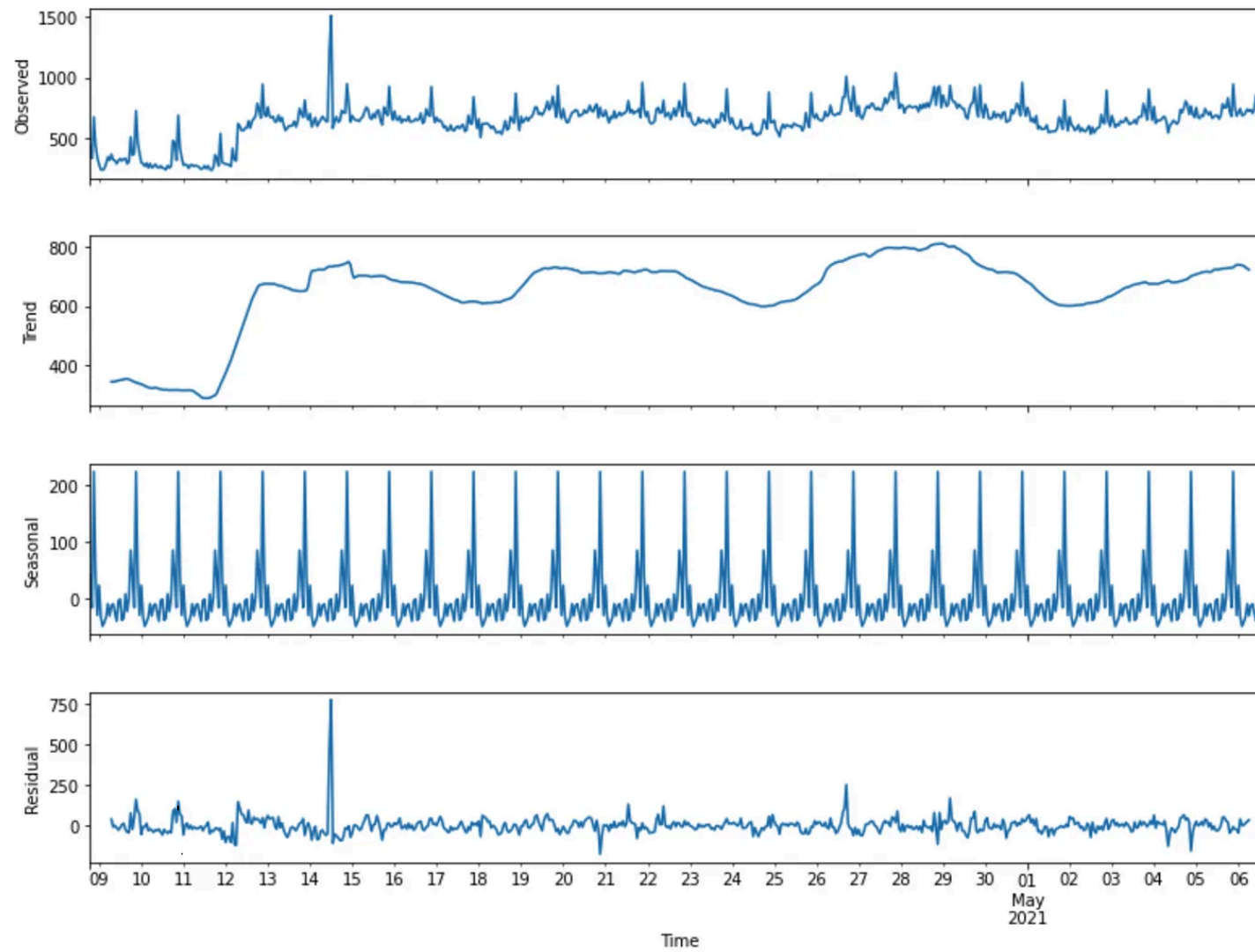
 statsmodels

- В самом простом варианте выделяет три компоненты, который образуют итоговое наблюдение

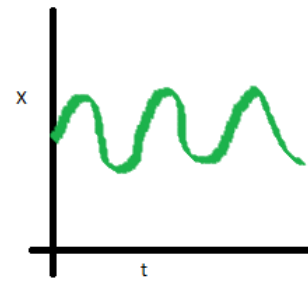
$$y(t) = h(t) + g(t) + \epsilon(t)$$

- $g(t)$ - Тренд
- $s(t)$ - Сезонность
- $e(t)$ - Остатки(ошибки)

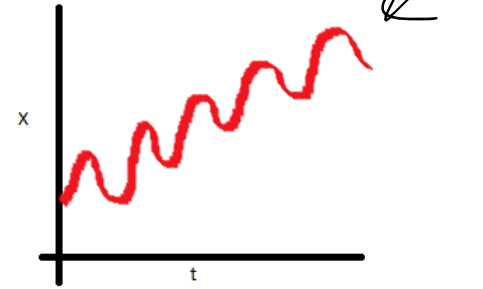
Компоненты временного ряда



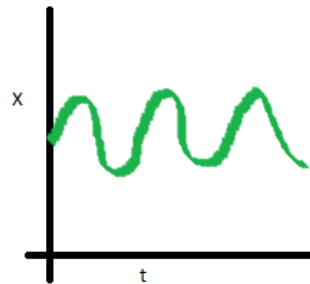
Стационарность временного ряда



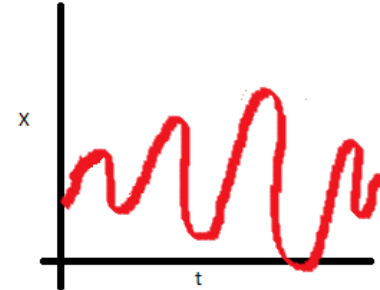
Stationary series



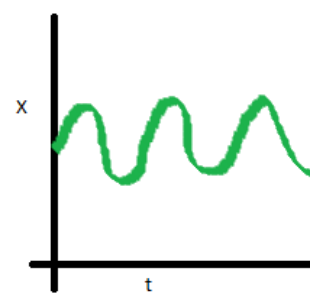
Non-Stationary series



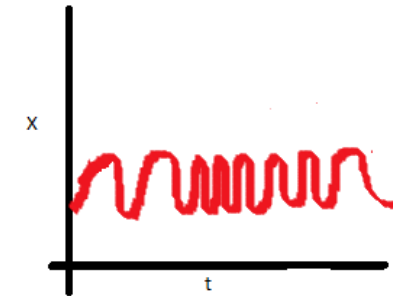
Stationary series



Non-Stationary series



Stationary series



Non-Stationary series

Для стационарного временного ряда характерно:

- отсутствие изменений матожидания во времени
- отсутствие изменений дисперсии во времени
- отсутствие изменения ковариационной функции во времени

Нестационарный временной ряд неудобен для исследования:

- большинство моделей строят прогноз по статистическим данным, если они нестабильны, то и прогноз будет нестабильным
- бороться с нестационарностью можно
- для проверки на стационарность используется тест Дики-Фуллера [1] [2]
 - Нулевая гипотеза: *временной ряд не является стационарным*

Приведение ряда к стационарному виду

1. Преобразование Бокса-Кокса

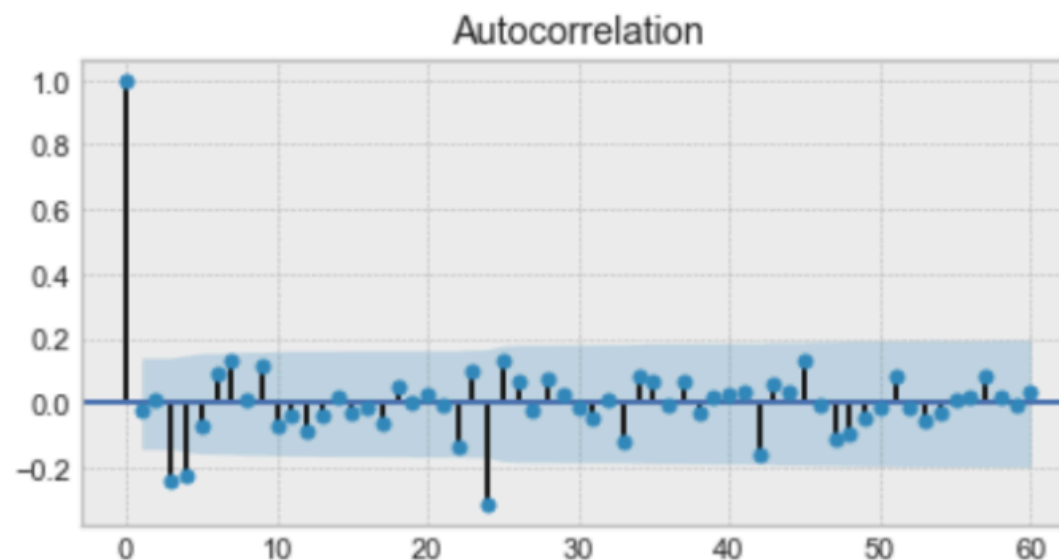
$$x'_i = \begin{cases} \frac{x_i^\lambda - 1}{\lambda}, & \lambda > 0 \\ \log(x), & \lambda = 0 \end{cases}$$

2. Дифференцирование временного ряда

y	y_{t+1}	Δ_{t+1}	y_{t-1}	Δ_{t-1}
2	NaN	NaN	4	-2
4	2	2	6	-2
6	4	2	8	-2
8	6	2	NaN	NaN

Автокорреляция / AutoCorrelationFunction / ACF

Корреляция ряда величин с самим собой: для временного ряда с самим собой со сдвигом по времени

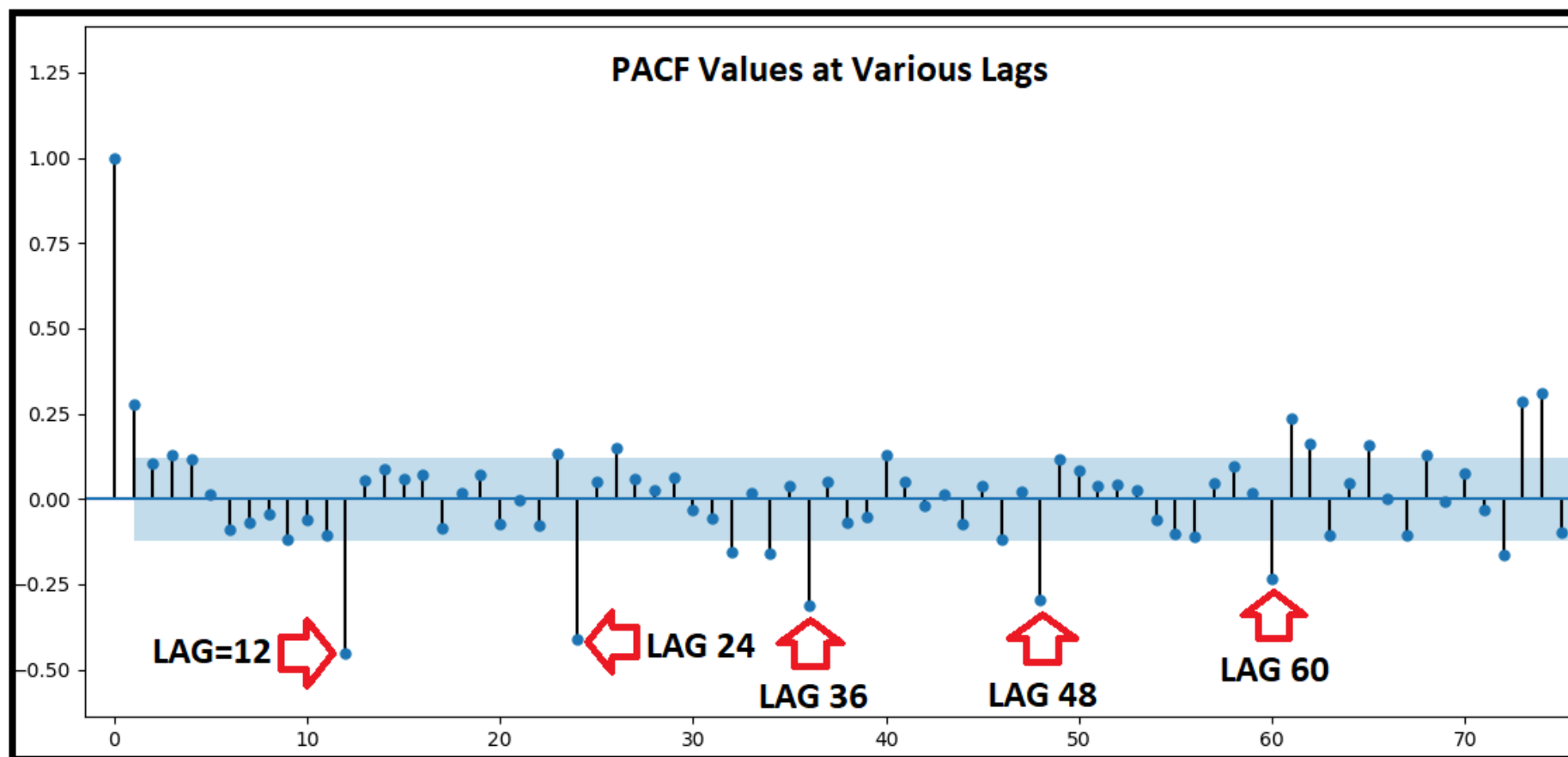


Example of an autocorrelation plot

Автокорреляционная функция показывает зависимость показателя корреляции от величины сдвига

Частичная автокорреляция / Partial ACF / PACF

Показывает корреляцию ряда с самим собой с вычетом промежуточных влияний



Модель $(S)ARIMA(p, d, q)(P, D, Q)_s$

- p – порядок авторегрессионной модели (AR), выбирается по PACF
 - номер последнего ненулевого элемента на графике PACF
- d – порядок интегрирования ряда
 - с каким лагом дифференцировали
- q – порядок модели MA (moving average), выбирается по ACF
 - номер последнего ненулевого элемента на графике ACF

Модель $(S)ARIMA(p, d, q)(P, D, Q)_s$

- P – порядок сезонной составляющей
 - смотрим на лаги, кратные размеру сезонности, например: 12, 24, 36... (PACF)
- D – порядок интегрирования сезонной составляющей
- Q – порядок сезонной составляющей
 - смотрим на лаги, кратные размеру сезонности, но на ACF
- s – размерность сезонности (месяц и т.д.)

- два самых распространенных подхода:
 - эконометрический (сезонности, тренды и тд)
 - о нем чаще спрашивают на собеседованиях
 - ML-подход: рассматриваем как задачу регрессии (при этом как приготовить признаки из одного временного ряда - отдельный вопрос)
 - В частности можно использовать рекуррентные нейронные сети, но об этом позже
- очень сильно зависит от предметной области
- доминирующие подходы сейчас - эконометрический и нейросетевой

- statsmodels
- fbprophet
- tsfresh
- autots
- Darts
- Kats
- [sktime: docs](#) / Sktime: унифицированная библиотека Python для машинного обучения и работы с временными рядами