Занятие 14. Временные ряды

Колмагоров Евгений
ml.hse.dpo@yandex.ru

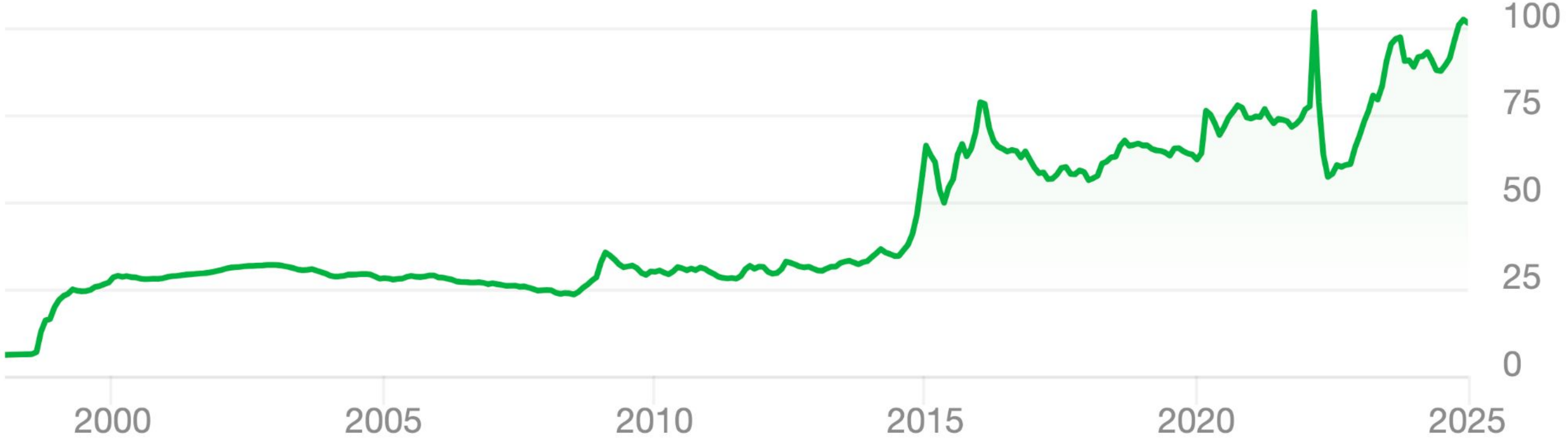
27 января 2025

План лекции

1. Виды временных рядов
2. Стационарность
3. Эконометрические модели
4. Адаптивные методы
5. Модели с трендом и сезонностью
6. Особенности обработки временных рядов

Понятие временного ряда

Временной ряд – это последовательность значений описывающих протекающий во времени процесс, измеренных в последовательные моменты времени, обычно через равные промежутки.



В каких задачах встречаются временные ряды

Задачи, связанные с обработкой и предсказаниями временных рядов широко встречаются в различных областях, но наиболее часто в экономических задачах:

- Предсказание стоимости ценных бумаг
- Курсов валют
- Цен различных активов: недвижимости, нефти, металлов и тд.
- Уровня ВВП
- Спроса на товары и услуги
- Прогнозирование погоды

Задача прогнозирования временного ряда

$y_0, y_1, \ldots, y_t, \ldots$ – временной ряд, $y_i \in R$

Задача: Построить функцию

$$\hat{y}_{t+d}(w) = a_{t,d}(y_1, \ldots, y_t; w)$$

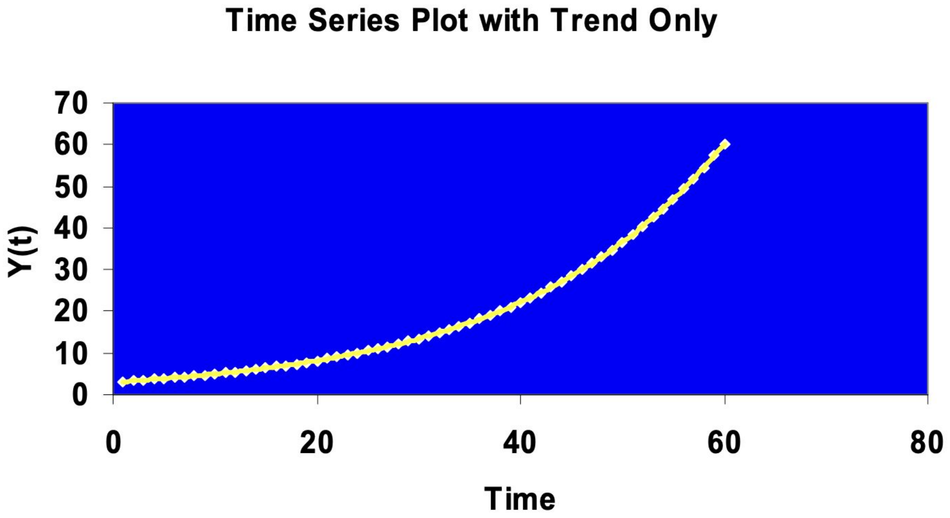
- d = 1, …, D, где D – горизонт прогнозирования
- w – вектор параметров модели

Особенности временных рядов

- в стандартных задачах машинного обучения предполагается, что наблюдения независимы и одинаково распределены
- в задаче анализа временных рядов предполагаем, что <u>ряд в прошлом содержит информацию о поведении ряда в будущем</u>

Компоненты временного ряда. Тренд

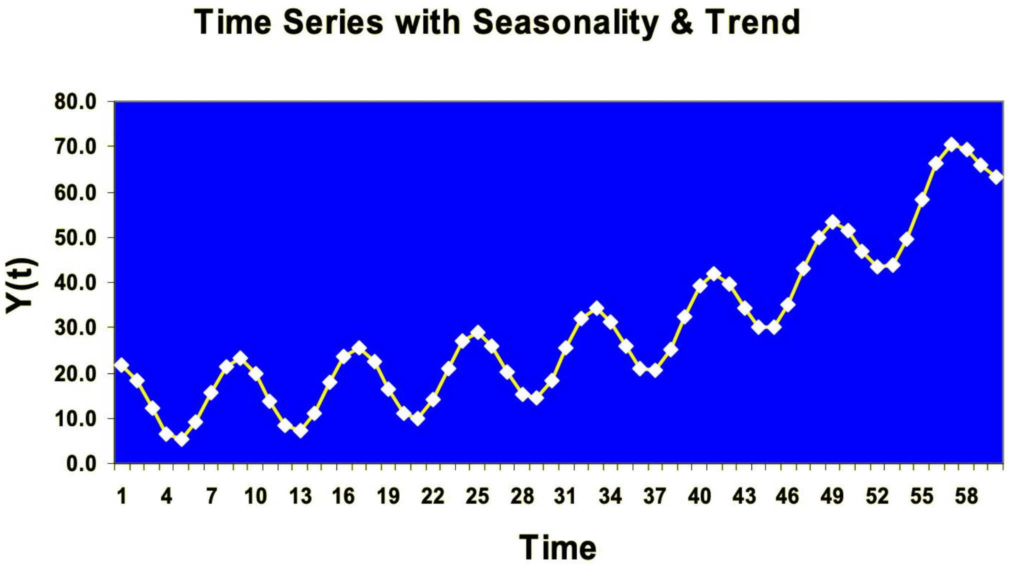
Тренд – плавное долгосрочное увеличение уровня ряда.



В данном случае ряд имеет тренд, так как значение Y монотонно возрастает

Компоненты временного ряда. Сезонность

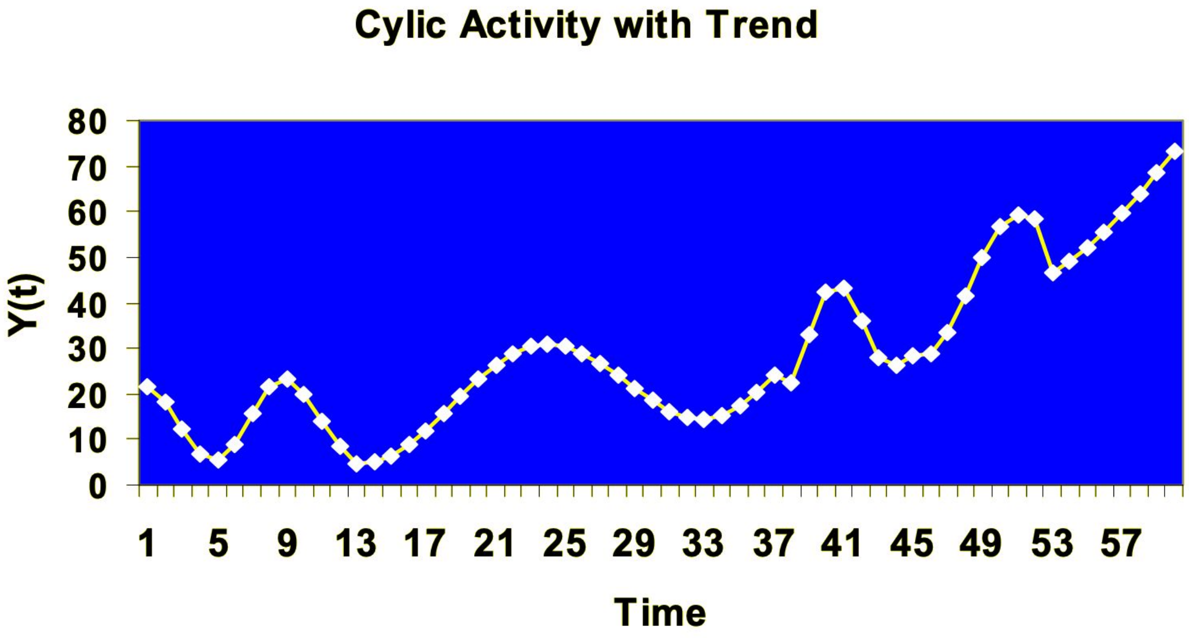
Сезонность – циклическое изменение уровня ряда с постоянным периодом



Ряд имеет как тренд так и сезонную компоненту

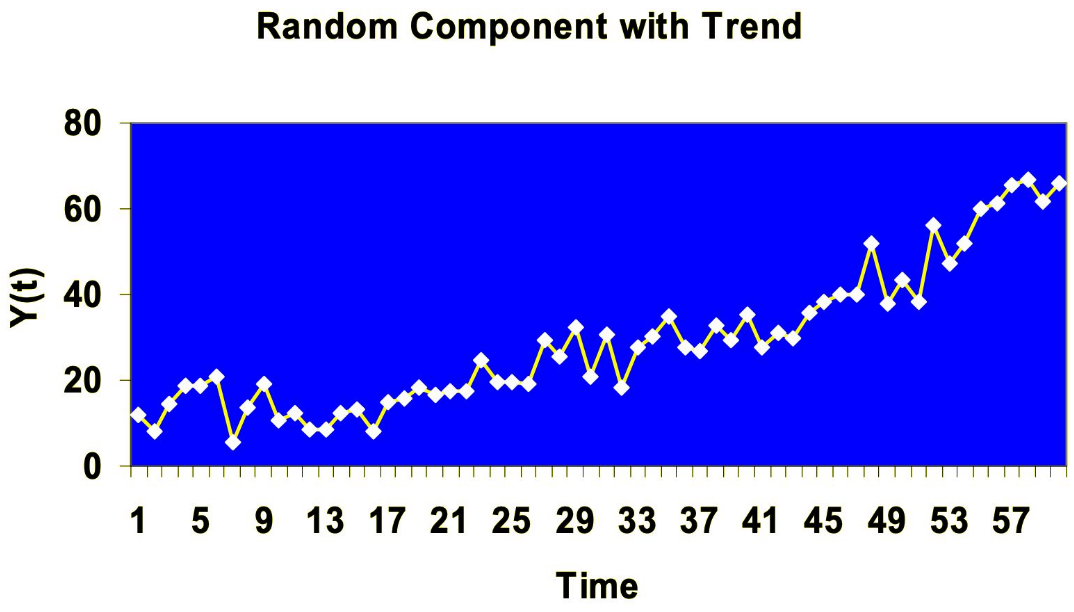
Компоненты временного ряда. Цикличность

Циклы – изменение уровня ряда с переменным периодом



Компоненты временного ряда. Шум

Шум (ошибка) – непрогнозируемая случайная компонента ряда



Стационарность в узком смысле

Ряд $y_1, ..., y_T$ стационарен в узком смысле, если для любого s распределение $y_1, \ldots, y_m$ совпадает с распределением $y_s, \ldots, y_{m+s}$. То есть свойства ряда не зависят от времени.

$$F(y_1, \ldots, y_m) = F(y_s, \ldots, y_{m+s})$$

Стационарность в широком смысле

Ряд Y_t называется стационарным в широком смысле, если

- $E(Y_t) = \mu$ при всех t
- $Var(Y_t) = \sigma$ при всех t
- $cov(Y_t, Y_{t+s}) = f(s)$ – зависит только от s, но не от t

Почему важна стационарность

- По стационарному ряду просто построить прогноз, так как мы полагаем, что его будущие статистические характеристики не будут отличаться от наблюдаемых текущих
- Если ряд имеет тренд или сезонность, то тогда ряд не стационарен, поэтому при прогнозировании стараются избавиться от этих двух компонент

Пример. Проверка стационарности

Попробуем доказать, что ряд ниже стационарен

$$Y_t = \beta_1 Y_{t-1} + \varepsilon, \text{ где } -1 < \beta < 1 \text{ и } \varepsilon \sim N(0, \sigma)$$

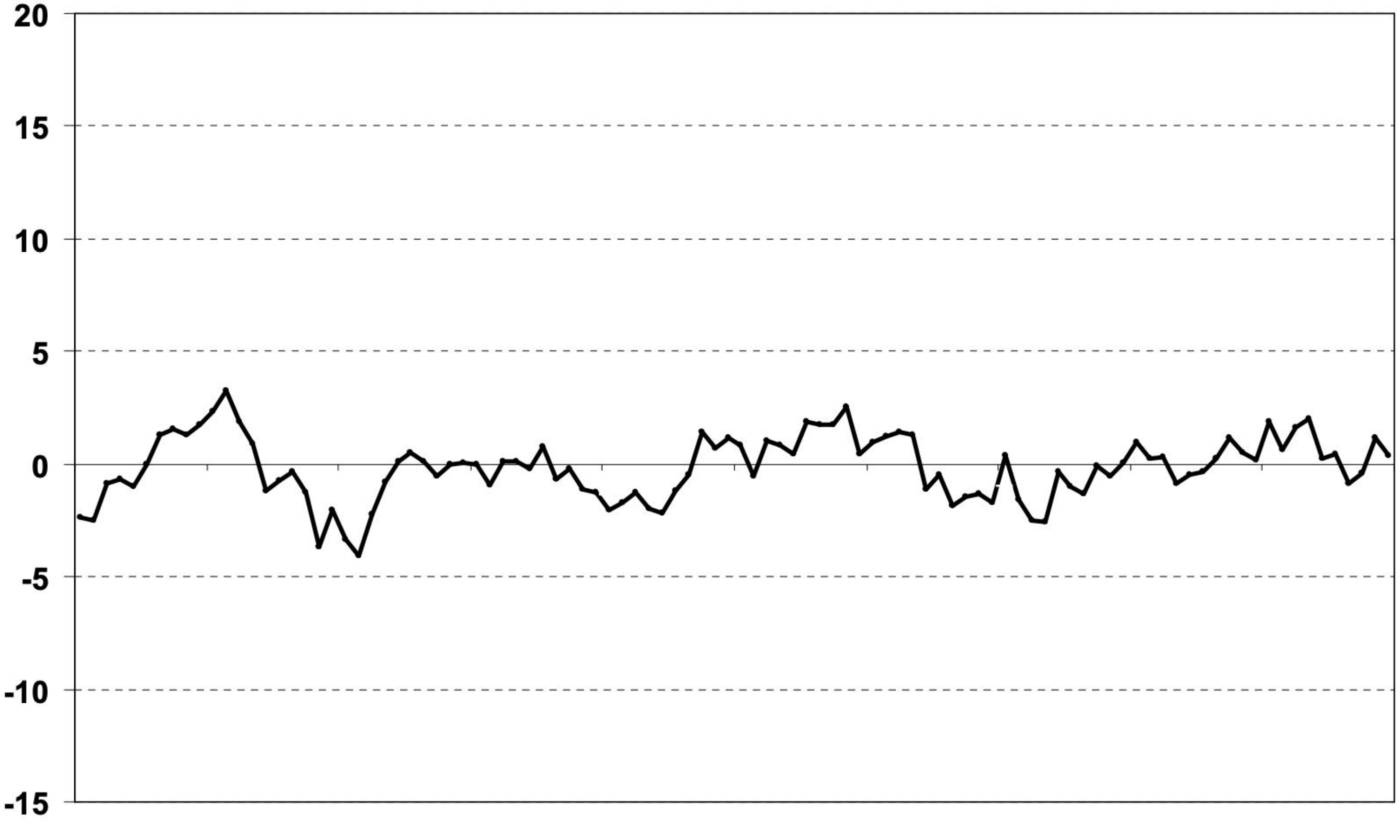
$$Y_t = \beta_1(\beta_1 Y_{t-1} + \varepsilon) + \varepsilon = \ldots = \beta_1^t Y_0 + \beta_1^{t-1}\varepsilon_1 + \ldots + \beta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$\mathbb{E}(Y_t) = \beta_1^t Y_0 + \beta_1^{t-1}\mathbb{E}\varepsilon_1 + \ldots + \beta_1 \mathbb{E}\varepsilon_{\text{t-1}} + \mathbb{E}\varepsilon_{\text{t}} = \beta_1^t Y_0 \to 0$$

$$\sigma^2(Y_t) = \mathbb{D}[\beta_1^t Y_0 + \beta_1^{t-1}\varepsilon_1 + \ldots + \beta_1 \varepsilon_{t-1} + \varepsilon_t] =$$

$$= \beta_1^{2t-2}\sigma_\varepsilon^2 + \ldots + \beta_1^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 = \frac{1-\beta^{2t}}{1-\beta^2}\sigma_\varepsilon^2 \to \frac{1}{1-\beta^2}\sigma_\varepsilon^2$$

Пример стационарного ряда



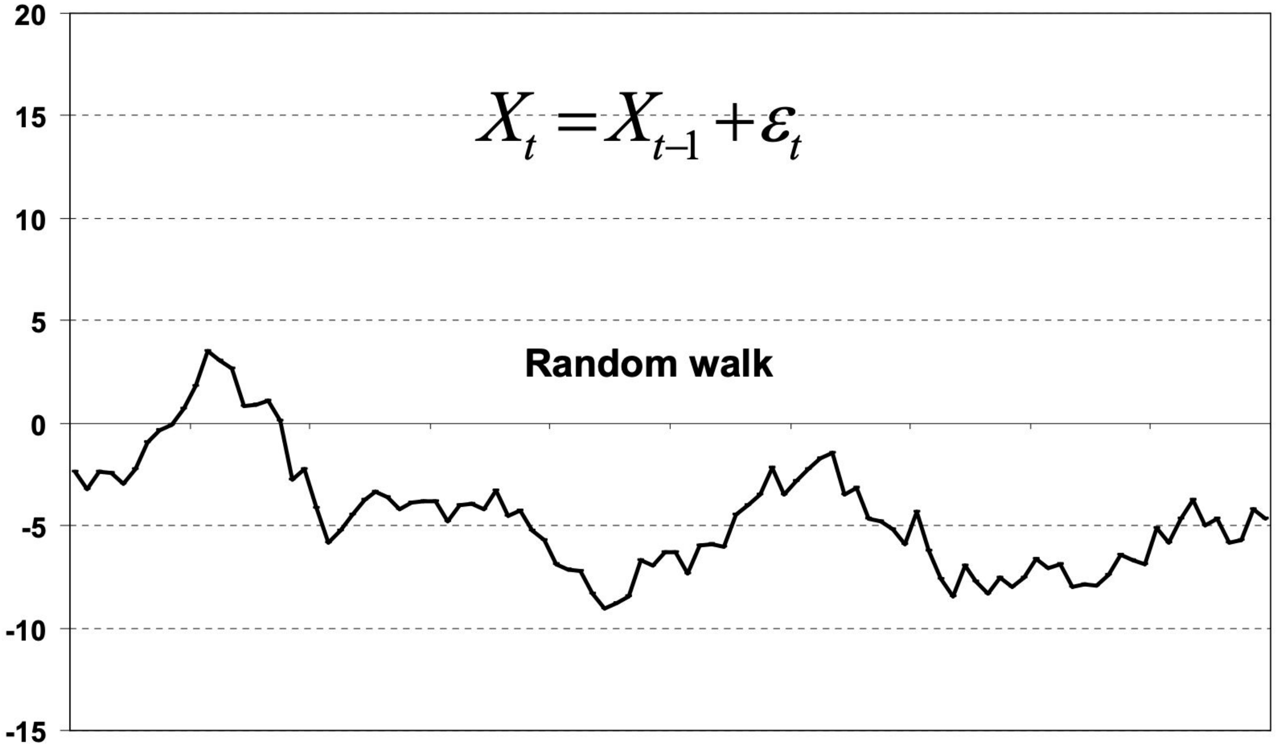
Единичный корень

Определение: Если ряд $Y_t = \beta_1 Y_{t-1} + \varepsilon$ имеет коэффициент β равный 1, то говорят, что ряд имеет единичный корень

Модель случайного блуждания, канонический пример ряда с единичным корнем:

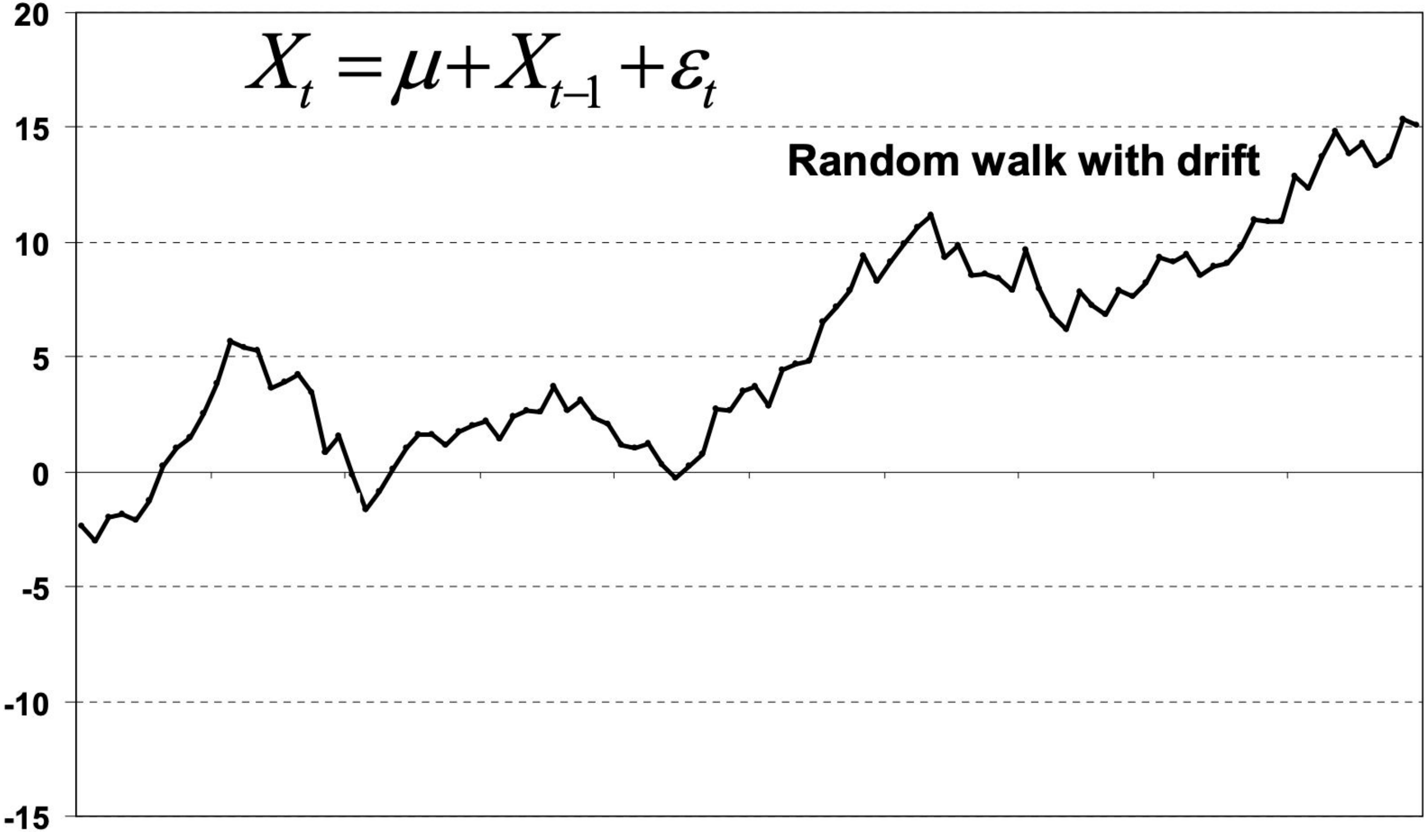
$$Y_t = Y_{t-1} + \varepsilon_t$$

Пример нестационарного ряда. Случайное блуждание



Видно, что при $\beta_1=1$ процесс не возвращается к своему среднему, а значит, не является стационарным

Пример нестационарного ряда. Случайное блуждание



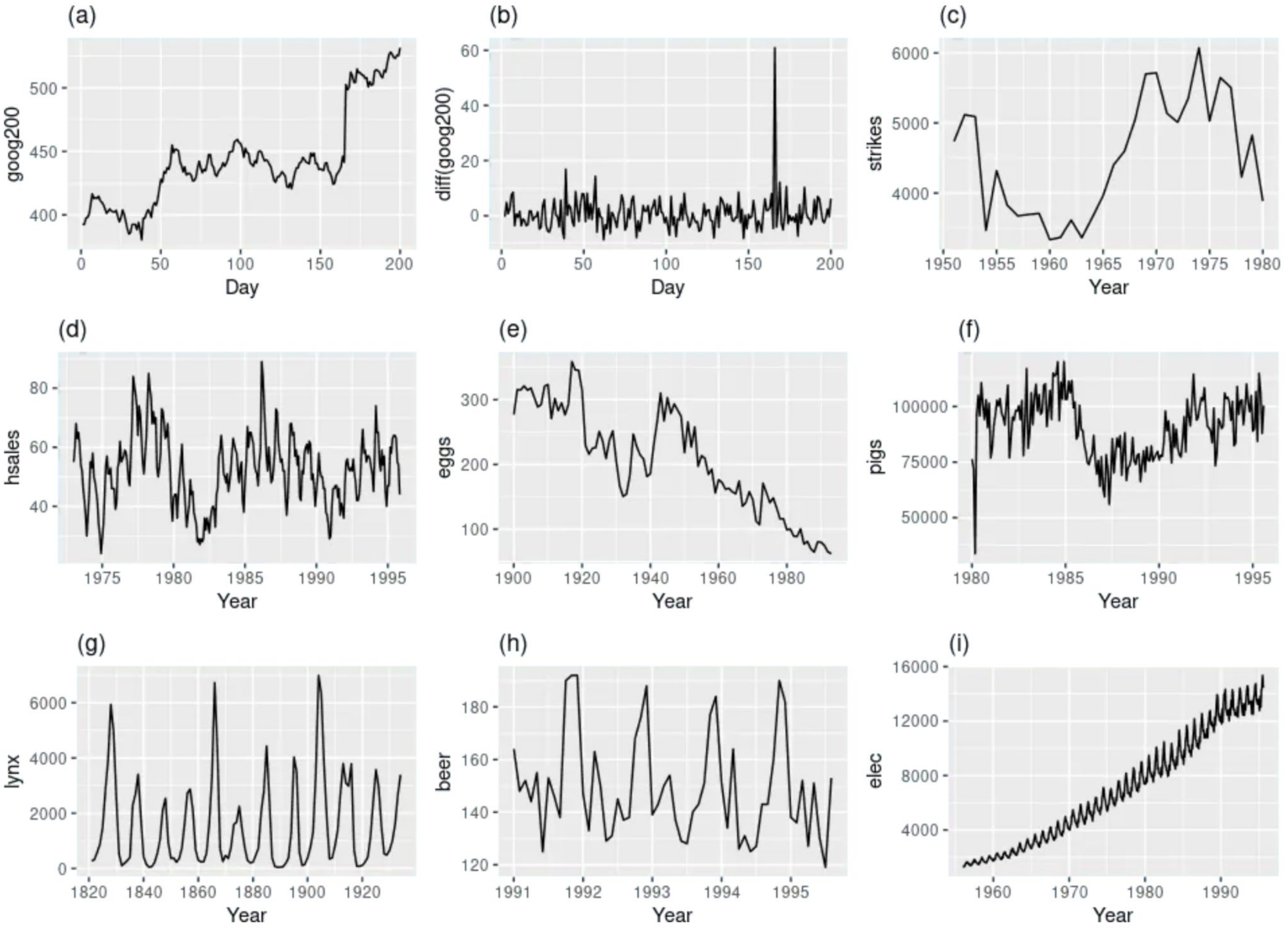
Проверка стационарности

Как можно было заметить формальное доказательство стационарности ряда представляет собой не самую простую задачу с точки зрения математики.

Для того, чтобы проводить быстрые тесты на стационарность используют методы основанные на проверки гипотез:

- Критерий Дики-Фуллера
- Критерий KPSS (Kwiatkowski–Phillips–Schmidt–Shin)

Попробуйте найти стационарный ряд

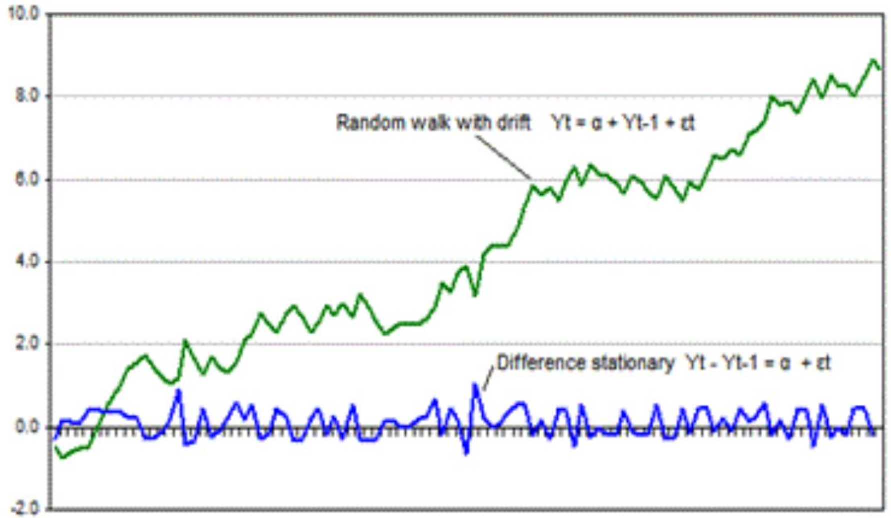


Методы избавления от нестационарности

Дифференцирование – переход к попарным разностям для соседних значений ряда

$$y'_t = y_t - y_{t-1}$$

- Стабилизирует среднее значение ряда, позволяет избавиться от тренда
- Можно применять неоднократно



Методы избавления от нестационарности

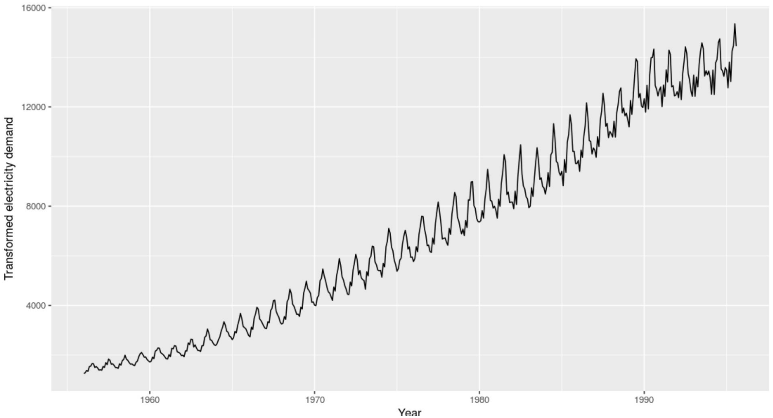
Для рядов с монотонно изменяющейся дисперсией (гетероскедастичные ряды) можно использовать стабилизирующее преобразование Бокса-Кокса:

$$y'_t = \begin{cases} ln\, y_t, & \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda}, & \text{где } \lambda \neq 0 \end{cases}$$

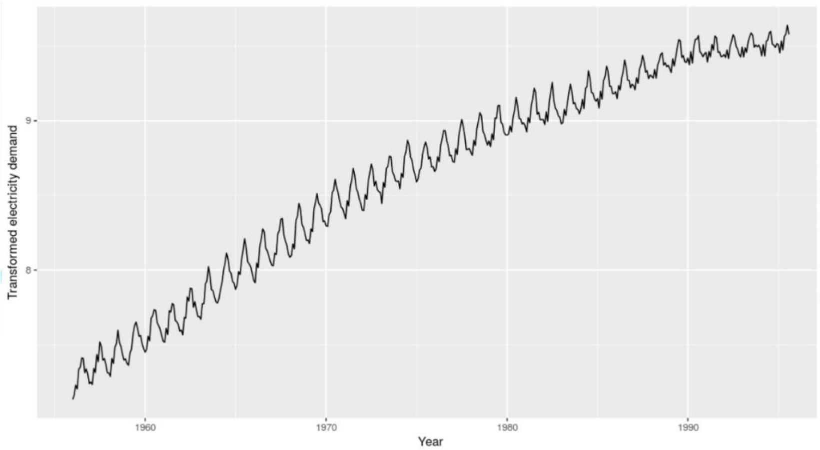
- Параметр λ подбирается так, чтобы сделать дисперсию как можно более однородной

Стабилизация дисперсии

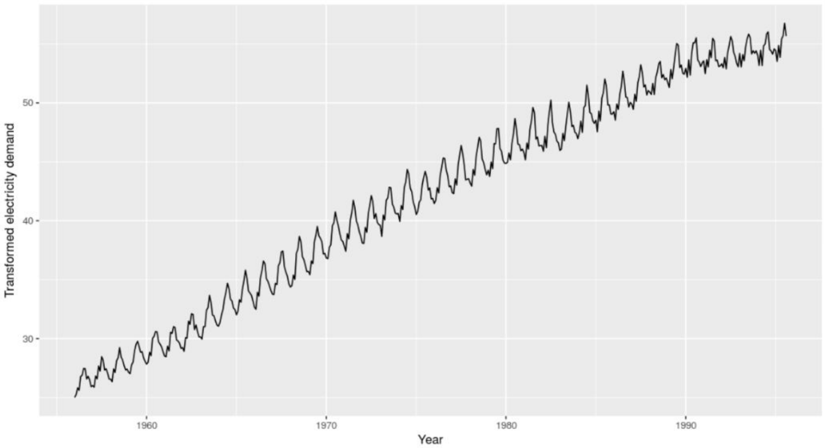
Исходный ряд



Параметр λ=0



Параметр λ=0.3



Поиск сезонности и тренда

При наличии во временном ряде тренда и сезонных колебаний значения любого последующего элемента ряда зависит от предыдущих.

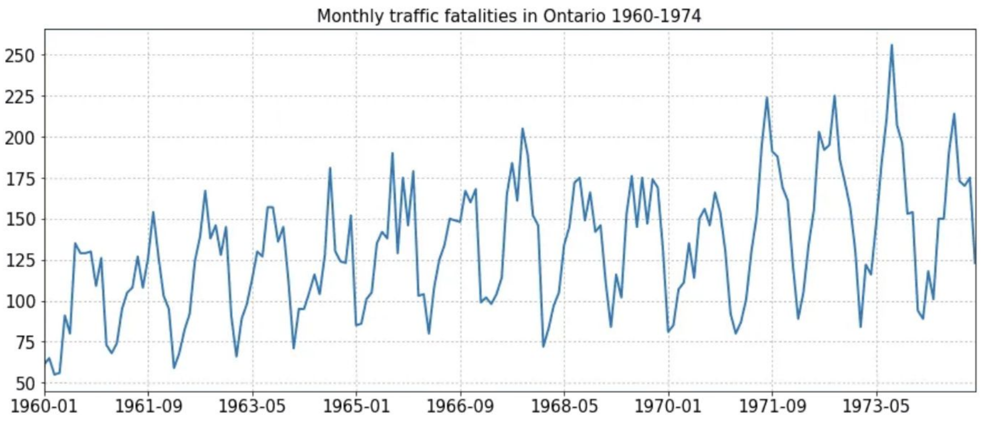
Определение: Корреляция (зависимость) между последовательными элементами временного ряда – это автокорреляция

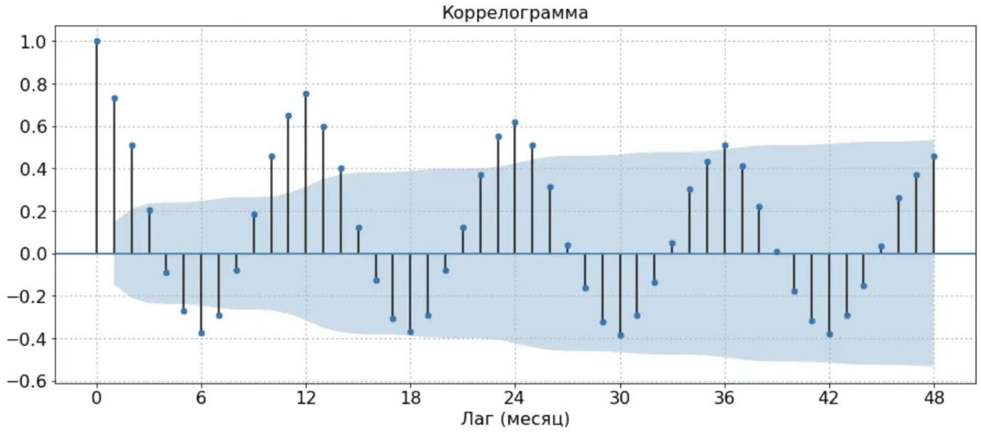
$$r_\tau = r_{y_t y_{t+\tau}} = \frac{\sum\limits_{t=1}^{T-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\sum\limits_{t=1}^{T} (y_t - \bar{y})^2}, \quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

Автокорреляция

Коррелограмма – график, визуализирующий автокорреляционную функцию

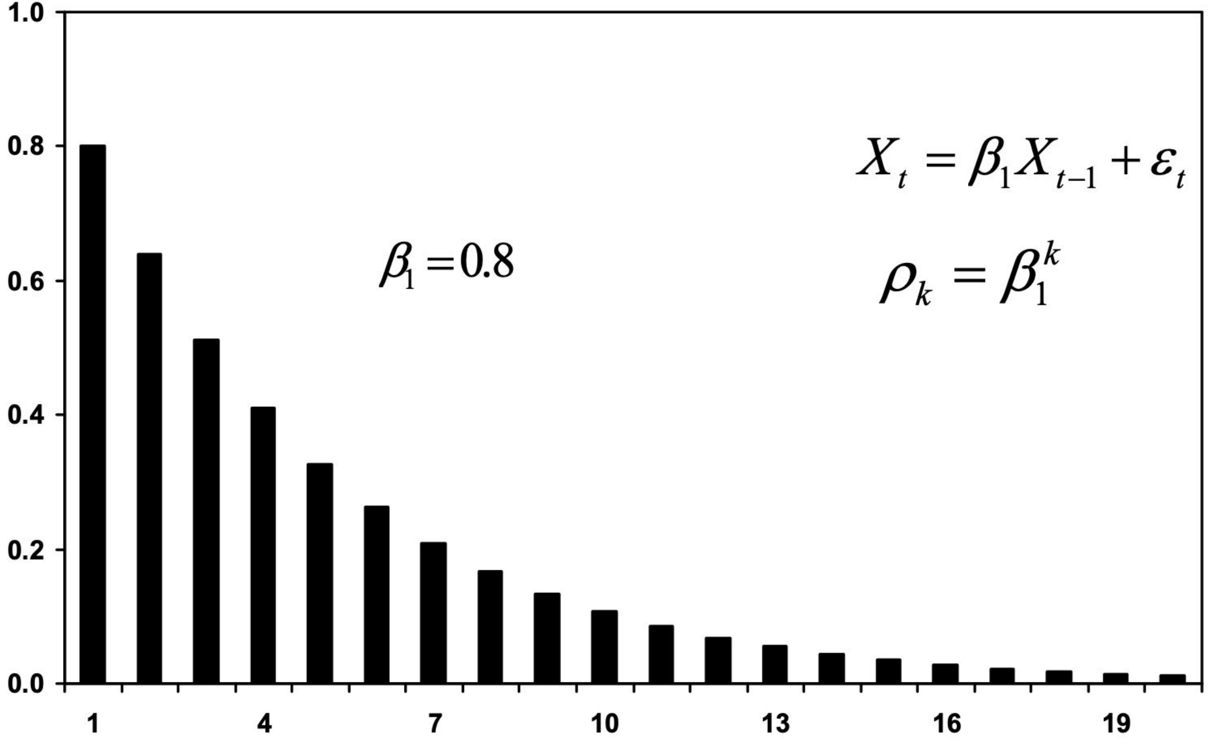
Автокорреляционная функция – множество коэффициентов корреляции для различных смещений



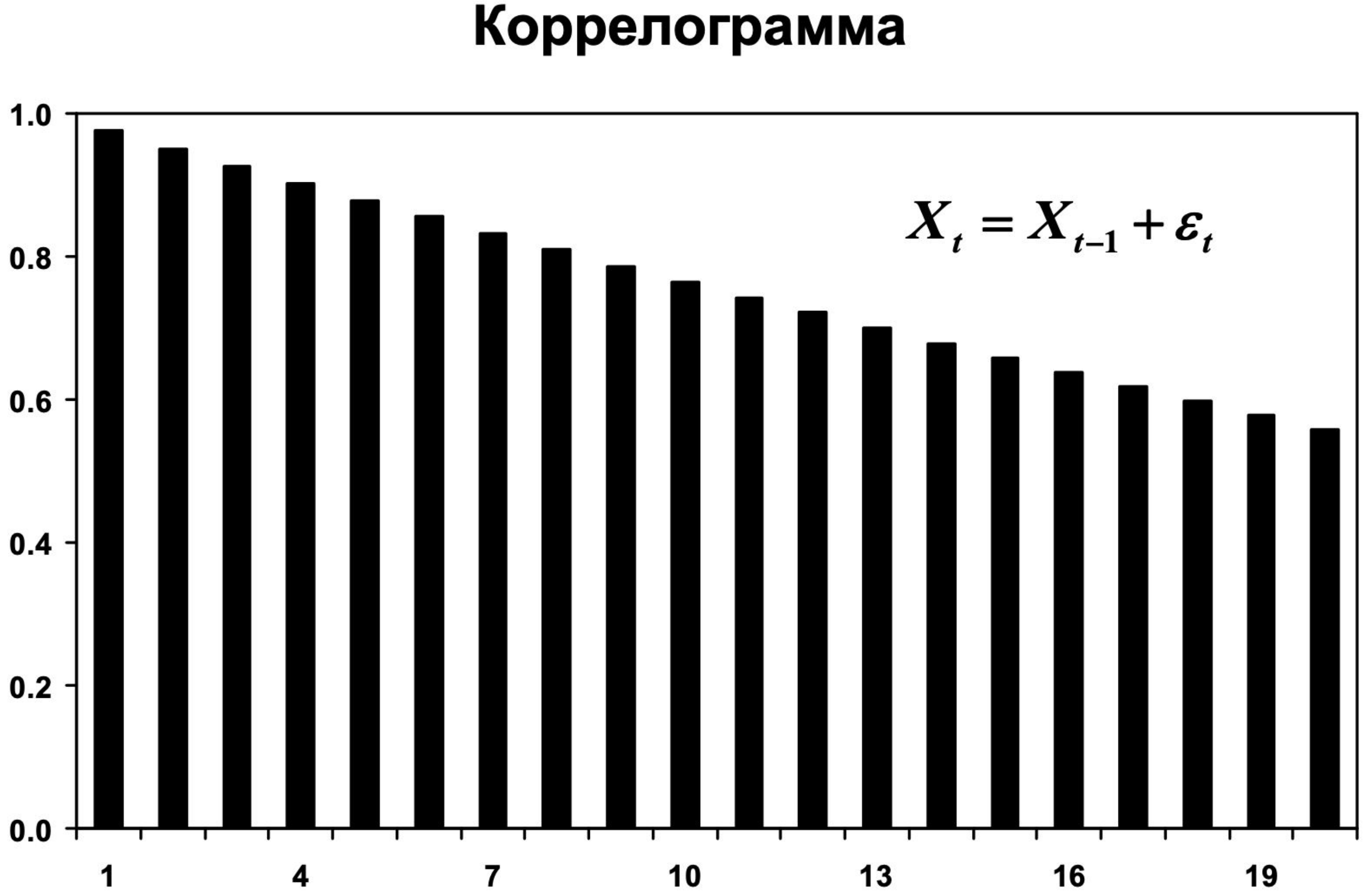


Коррелограмма стационарного процесса

Автокорреляционная функция стационарного процесса быстро убывает с ростом **τ**



Коррелограмма нестационарного процесса



Частная автокрреляция

Частная автокорреляционная функция – “чистая корреляция” между X_t и X_{t-k} при исключении влияния X_{t-1}, X_{t-2}, …, X_{t-k+1}

Вычисляется как оценка коэффициента β_k методом наименьших квадратов для в регрессии

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_k X_{t-k} + \varepsilon_t$$

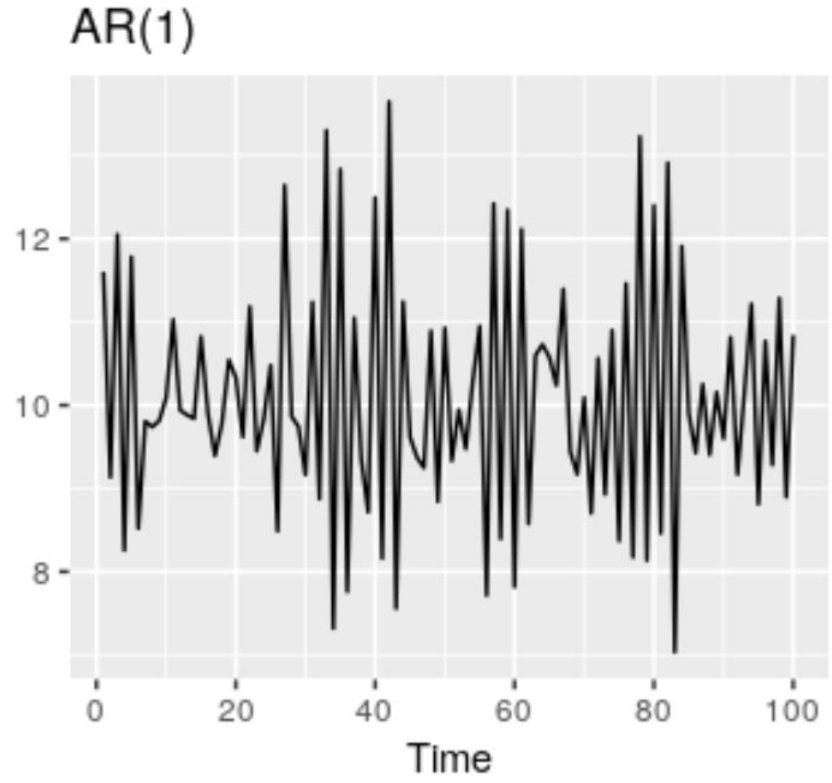
Эконометрические модели временных рядов

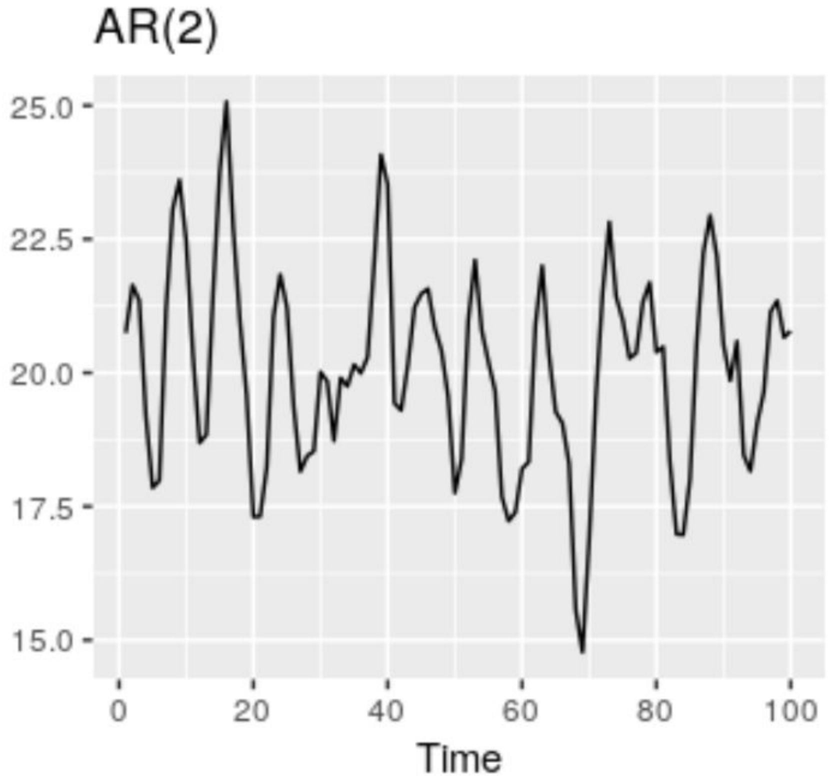
Модель авторегрессии AR(p):

$$y_t = \alpha + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t$$

- $\alpha_p \neq 0$
- ε - процесс белого шума, $\varepsilon \sim N(0, \sigma^2)$
- $cov(\varepsilon_t, \varepsilon_s) = 0$

Модель авторегрессии





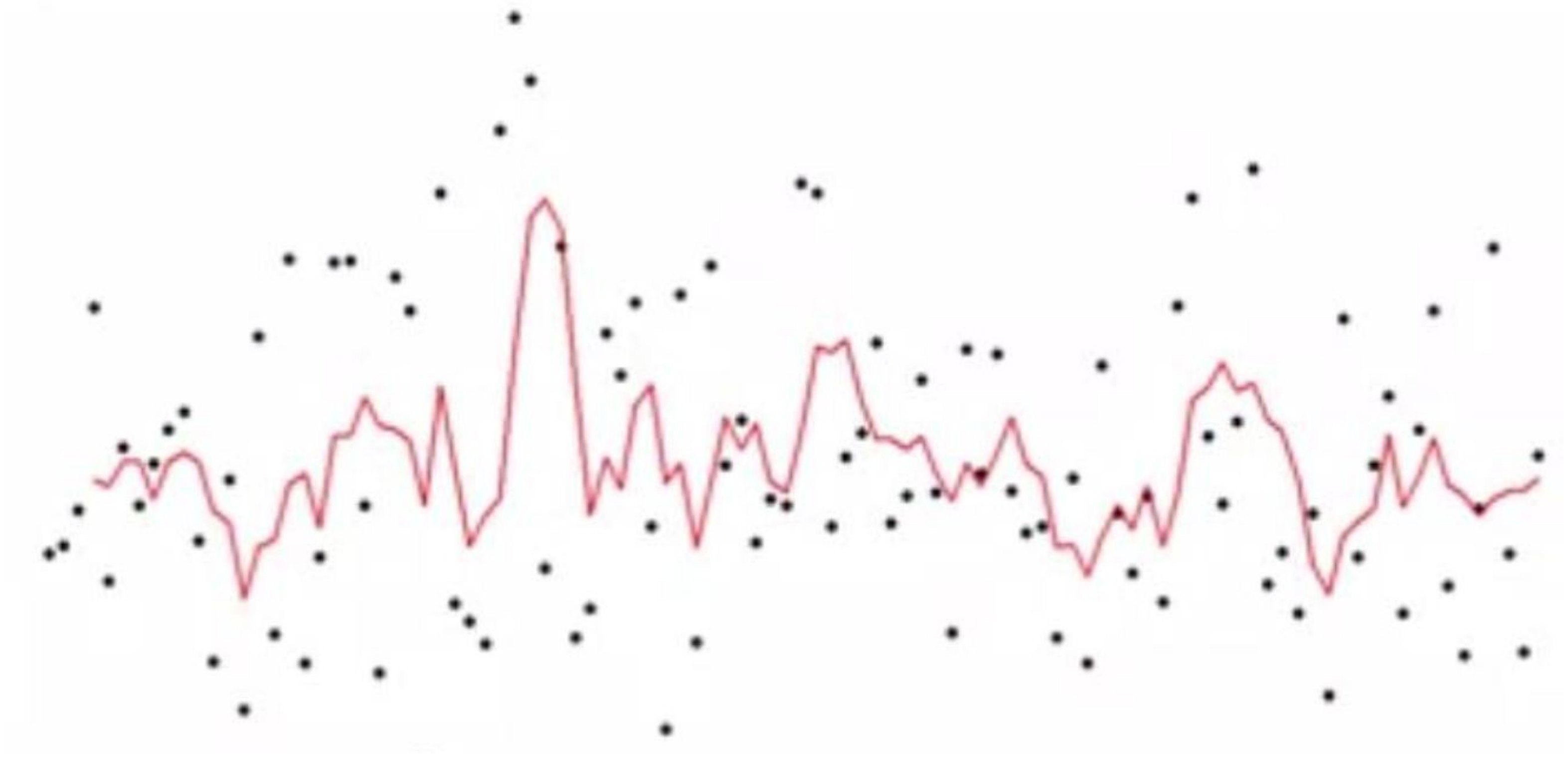
Модель скользящего среднего

Модель скользящего среднего порядка q (MA(q)):

$$y_t = \beta + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q}$$

- $\beta_q \neq 0$
- ε_t - процесс белого шума, $\varepsilon_t \sim N(0, \sigma^2)$
- $\mathrm{cov}(\varepsilon_t, \varepsilon_s) = 0$
- Модель скользящего среднего всегда стационарна

Модель скользящего среднего

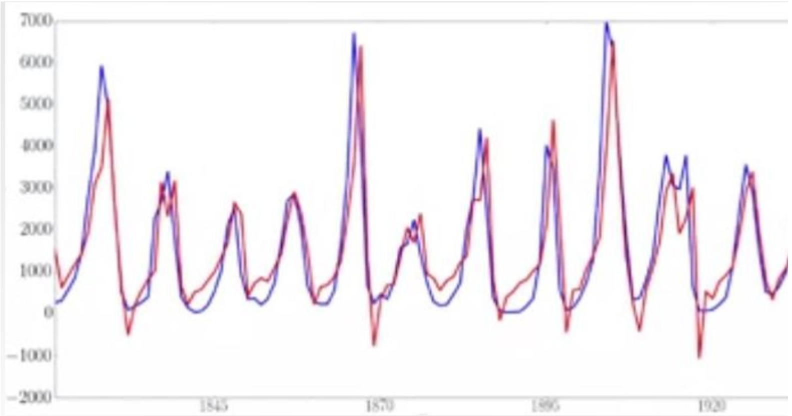


Смешанный процесс авторегрессии

Смешанный процесс авторегрессии ARMA(p, q):

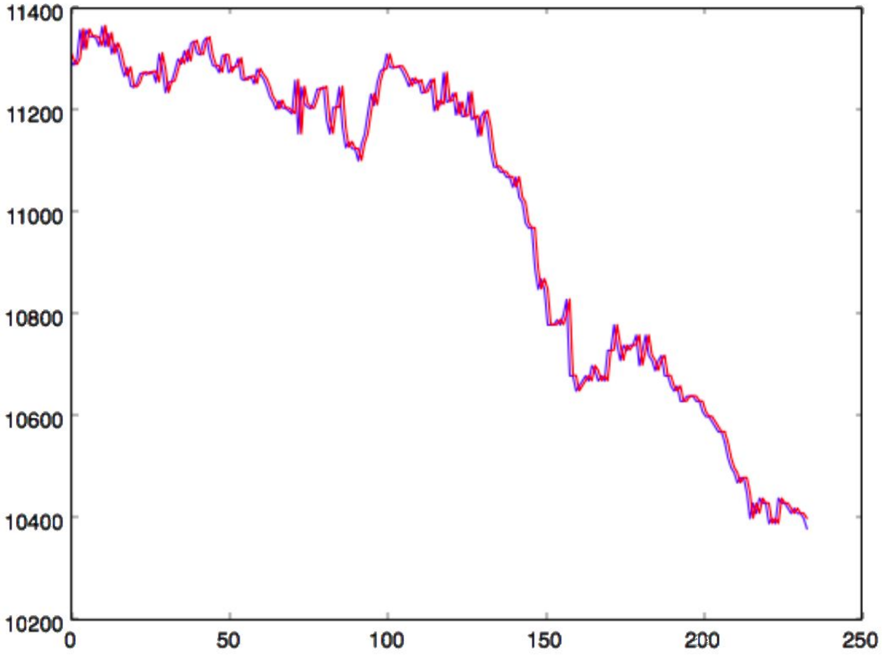
$$y_t = \alpha + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q}$$

Теорема Вольда: Любой стационарный ряд можно приблизить моделью ARMA(p, q) сколь угодно точно.



Модель ARIMA

- Модель ARIMA(p, q, d) – модель ARMA(p, q) для d раз продифференцированного ряда.



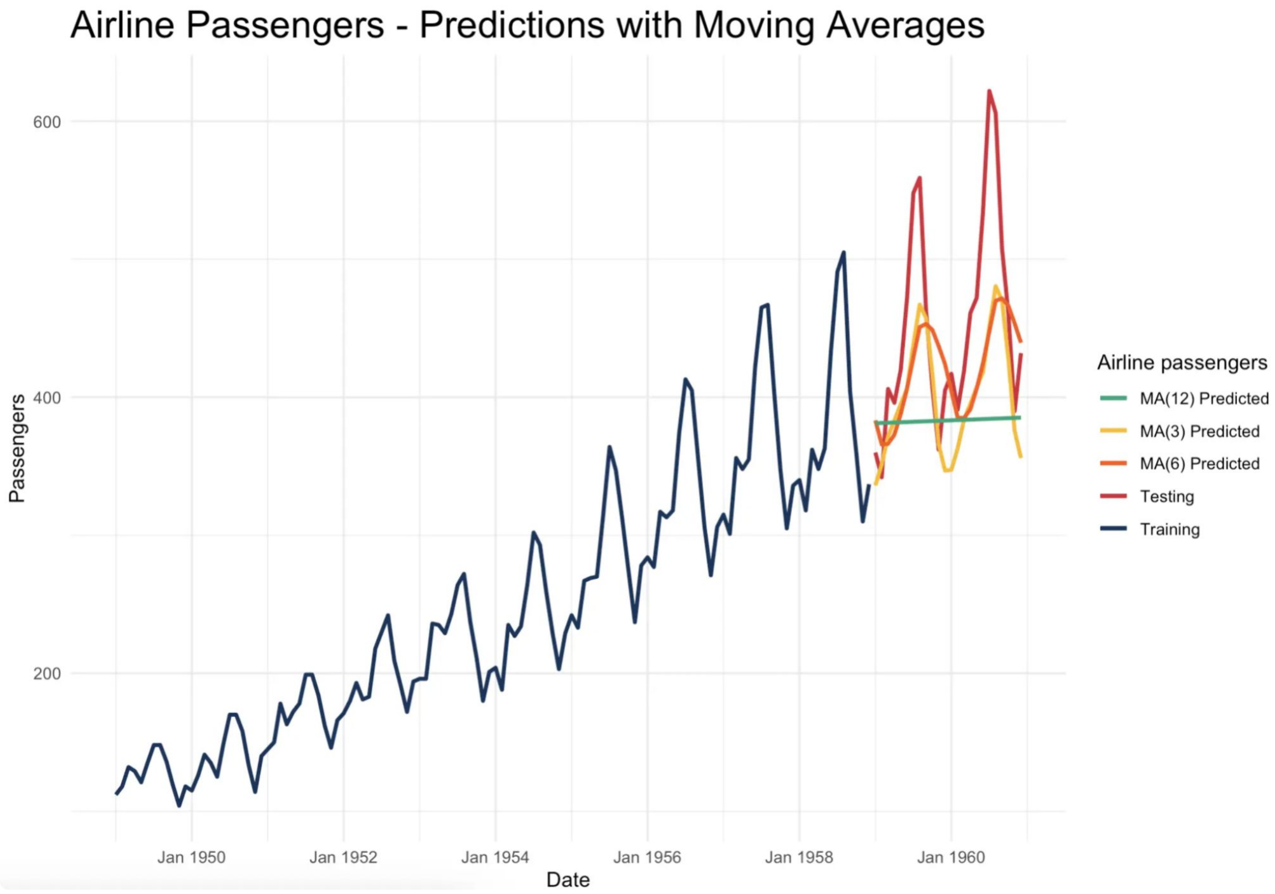
Адаптивные методы. Скользящее среднее

$$\hat{y}_{t+1} = \frac{1}{k} \sum\nolimits_{i=0}^{k-1} y_{t-i}$$

- Чтобы сделать прогноз на следующий период времени, надо знать значение на текущий (т.е. долгосрочный прогноз невозможен)

- Сглаживает данные и уменьшает влияние выбросов

Влияние размера окна на прогнозы



Экспоненциальное скользящее среднее

Идея метода: На значение ряда в данный момент времени больше влияет значение в предыдущий момент времени, затем – значение в предпредыдущий момент и т.д. (то есть более поздние данные – более важные).

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}$$

Пример:

$$Y_t = \frac{1}{2}y_{t-1} + \frac{1}{4}y_{t-2} + \frac{1}{8}y_{t-3} + \ldots$$

Обоснование метода

- Будем делать прогноз в момент времени t+1 простейшей прогнозной моделью – константой ($c = c_t$): $\hat{y}_{t+1} = c$
- Минимизируем среднеквадратичную ошибку с весами β^i, убывающим в прошлое:

$$\sum_{t=0}^{T} \beta^{T-t} (y_t - c)^2 \to min_c$$

- Приравниваем производную к 0:

$$2 \sum_{t=0}^{T} \beta^{T-t} (y_t - c) = 0$$

Обоснование метода

- Выразим c и воспользуемся формулой разложения 1/(1-x) в ряд Тейлора:

$$c = \frac{\sum_{t=0}^{T} \beta^{T-t} y_t}{\sum_{t=0}^{T} b^t} \approx \frac{\sum_{t=0}^{T} \beta^{T-t} y_t}{1/(1-\beta)} = (1-\beta) \sum_{t=0}^{T} \beta^{T-t} y_t =$$

$$= (1-\beta) y_T + (1-\beta)\beta \sum_{t=0}^{T-1} \beta^{T-t-1} y_t = (1-\beta) y_t + \beta \hat{y}_t$$

- Получили модель экспоненциального сглаживания с α = (1 - β)

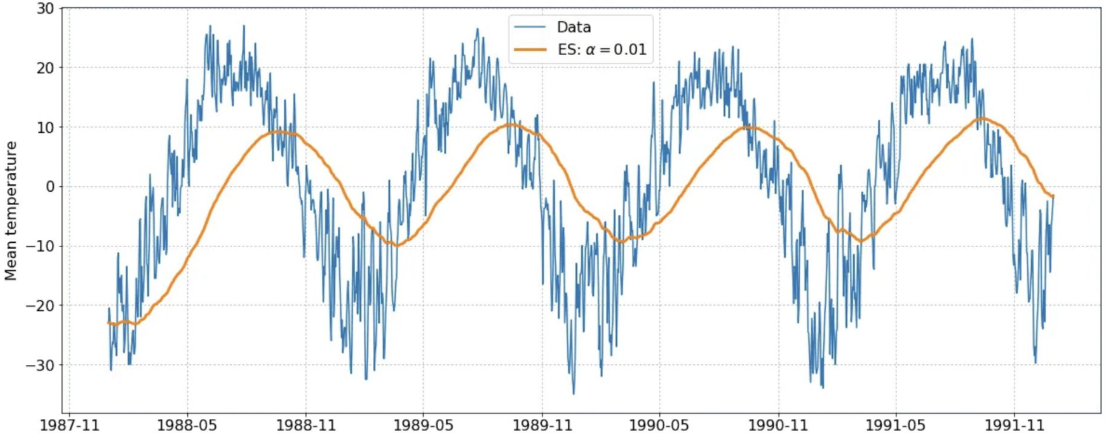
Свойства

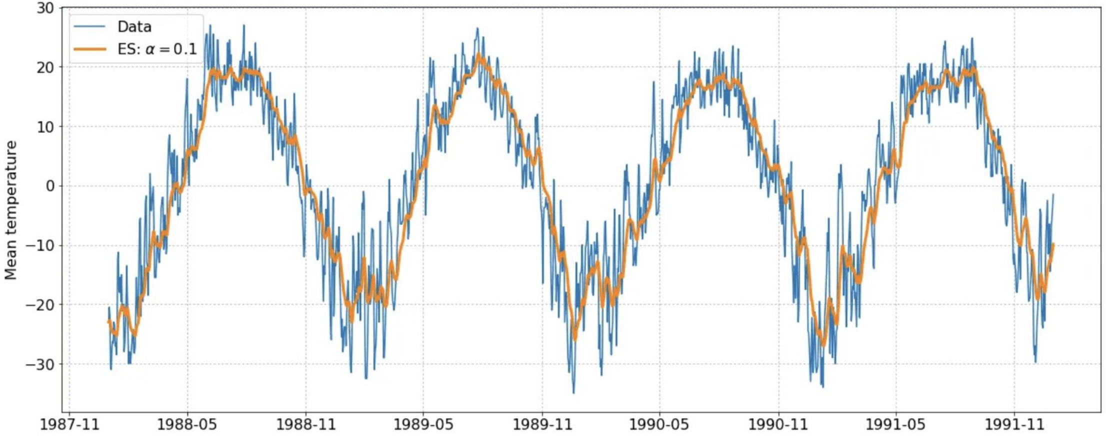
$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}$$

- Чем больше α тем больше вес последним значениям ряда, поэтому получается слабое сглаживание
- При малых α больший вес отдаётся сглаженному значению и сглаживание становится сильнее
- Оптимальное значение α можно получить по графику либо оптимизируя:

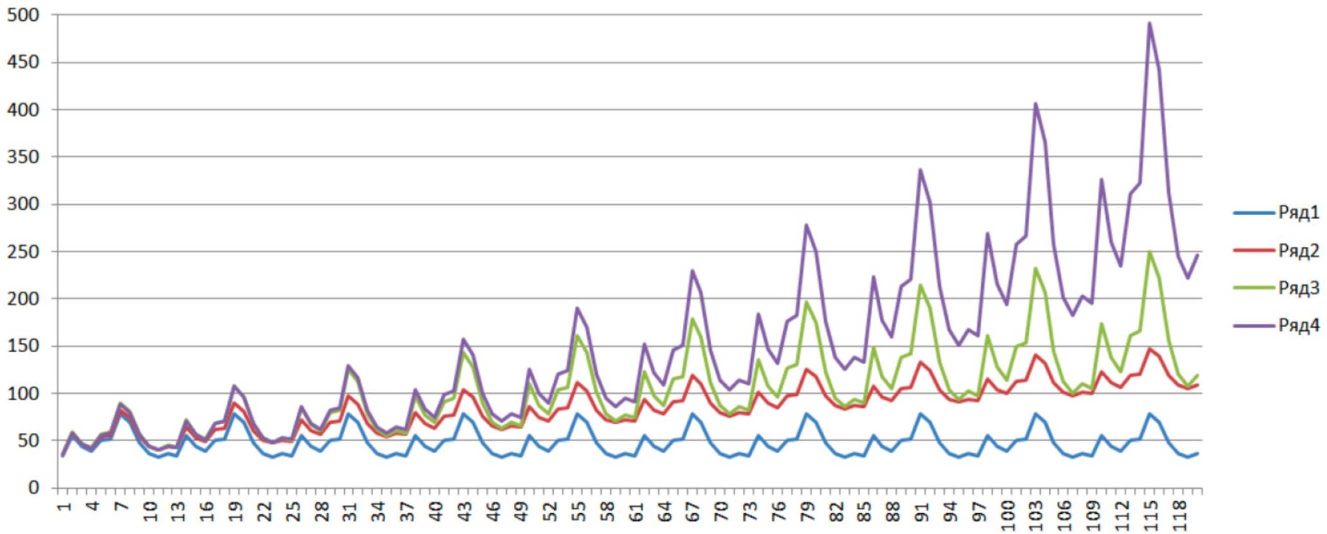
$$\sum_{t=0}^{T}(\hat{y}_t(\alpha) - y_t)^2 \rightarrow min_\alpha$$

Влияние параметра на прогнозы





Модели с трендом и сезонностью



- Ряд 1 – сезонность без тренда
- Ряд 2 – линейный тренд аддитивная сезонность
- Ряд 3 – линейный тренд мультипликативная сезонность
- Ряд 4 – линейный тренд экспоненциальная сезонность

Модели с трендом и сезонностью

- **Модель Хольта** – модель линейного тренда:

$$\hat{y}_{t+d} = a_t + b_t d$$

- a_t, b_t – адаптивные компоненты линейного тренда, рассчитываемые как

$$a_t = \alpha_1 y_t + (1 - \alpha_1)(a_{t-1} + b_{t-1})$$
$$b_t = \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1}$$

- где α_1, α_2 – параметры сглаживания

Модели с трендом и сезонностью

- **Модель Тейла-Вейджа** – модель линейного тренда с аддитивной сезонностью:

$$\hat{y}_{t+d} = (a_t + b_t d) + \theta_{t+(d \bmod s)-s}$$

- $a_t + b_t d$ – тренд, очищенный от сезонных колебаний $\Theta_0, \ldots, \Theta_{s-1}$ – сезонный профиль периода без тренда

$$a_t = \alpha_1 y_t + (1 - \alpha_1)(a_{t-1} + b_{t-1})$$
$$b_t = \alpha_2 (a_t - a_{t-1}) + (1 - \alpha_2) b_{t-1}$$
$$\theta_t = \alpha_3 (y_t - a_t) + (1 - \alpha_3) \theta_{t-s}$$

Адаптивная модель авторегрессии

Линейная модель авторегрессии (линейный фильтр):

$$\hat{y}_{t+1} = \sum_{j=0}^{d} w_j y_{t-j+1}$$

- Обучающее множество

$$\underset{\ell \times n}{F} = \begin{pmatrix} y_{t-1} & y_{t-2} & y_{t-3} & \cdots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \cdots & y_{t-n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n & y_{n-1} & y_{n-2} & \cdots & y_1 \\ y_{n-1} & y_{n-2} & y_{n-3} & \cdots & y_0 \end{pmatrix}, \quad \underset{\ell \times 1}{y} = \begin{pmatrix} y_t \\ y_{t-1} \\ \cdots \\ y_{n+1} \\ y_n \end{pmatrix}$$

- Функционал качества

$$Q(w) = \sum_{i=n}^{t} (\hat{y}_i(t) - y_i)^2 = ||Fw - y||^2 \to \min_w$$

Адаптивная модель авторегрессии

- Ошибка на прогнозе для момента времени t:

$$\varepsilon_t = y_t - \hat{y}_t$$

- Один шаг градиентного спуска в момент времени t:

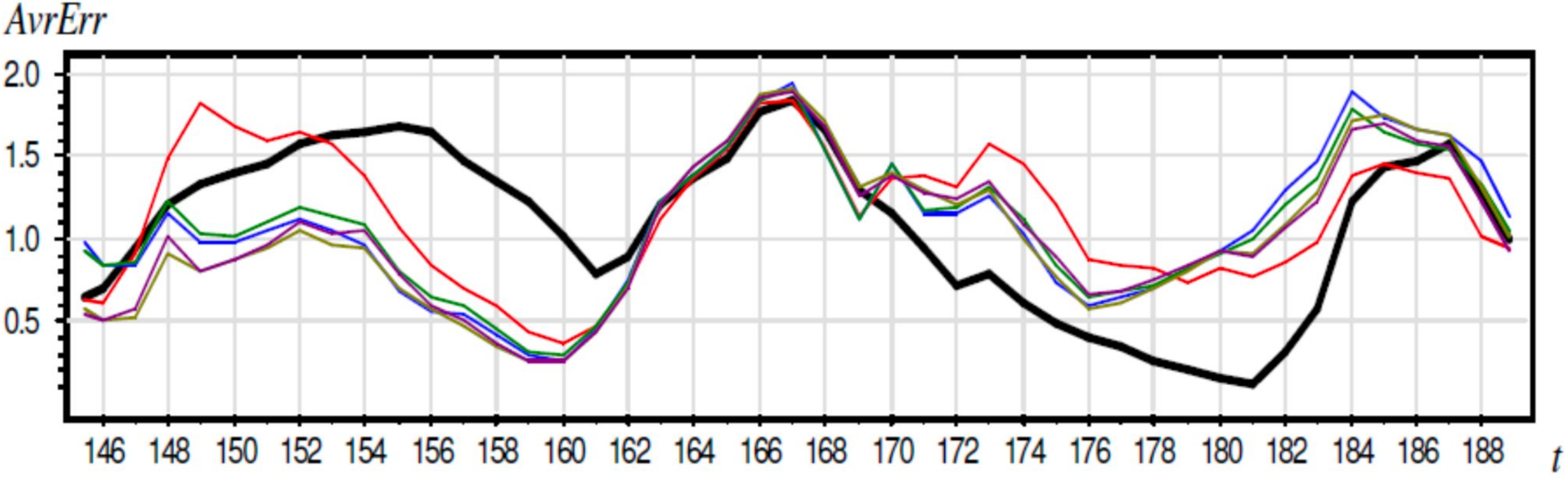
$$w_j = w_j + h_t \varepsilon_t y_{t-j+1}$$

- Градиентный шаг:

$$h_t = \frac{\alpha}{\sum_{j=1}^{n} y_{t-j+1}^2}$$

Адаптивная селекция и композиция моделей

Пример: $|\varepsilon_t|$ для нескольких моделей временного ряда



Адаптивная селекция

- Пусть имеется k моделей прогнозирования, $\hat{y}_{j,t+d}$ – прогноз j-ой модели на момент t+d
- $\varepsilon_{jt} = y_t - \hat{y}_{jt}$ – ошибка прогноза в момент времени t
- $\tilde{\varepsilon}_{jt} = \gamma|\varepsilon_{jt}| + (1-\gamma)\tilde{\varepsilon}_{jt}$ – сглаженная ошибка
- Лучшая модель в момент времени t:

$$j_t^* = \underset{j=1,\dots,k}{argmin}\, \tilde{\varepsilon}_{jt}$$

- Итоговый прогноз

$$\hat{y}_{t+d} = \hat{y}_{j_t^*,t+d}$$

Адаптивная комбинация

- В каждый момент времени итоговое предсказание – это сумма предсказаний с весами, обратно пропорциональными их ошибкам
- Предсказание:

$$\hat{y}_{t+d} = \sum_{j=1}^{k} w_{jt} \hat{y}_{j.t+d}$$

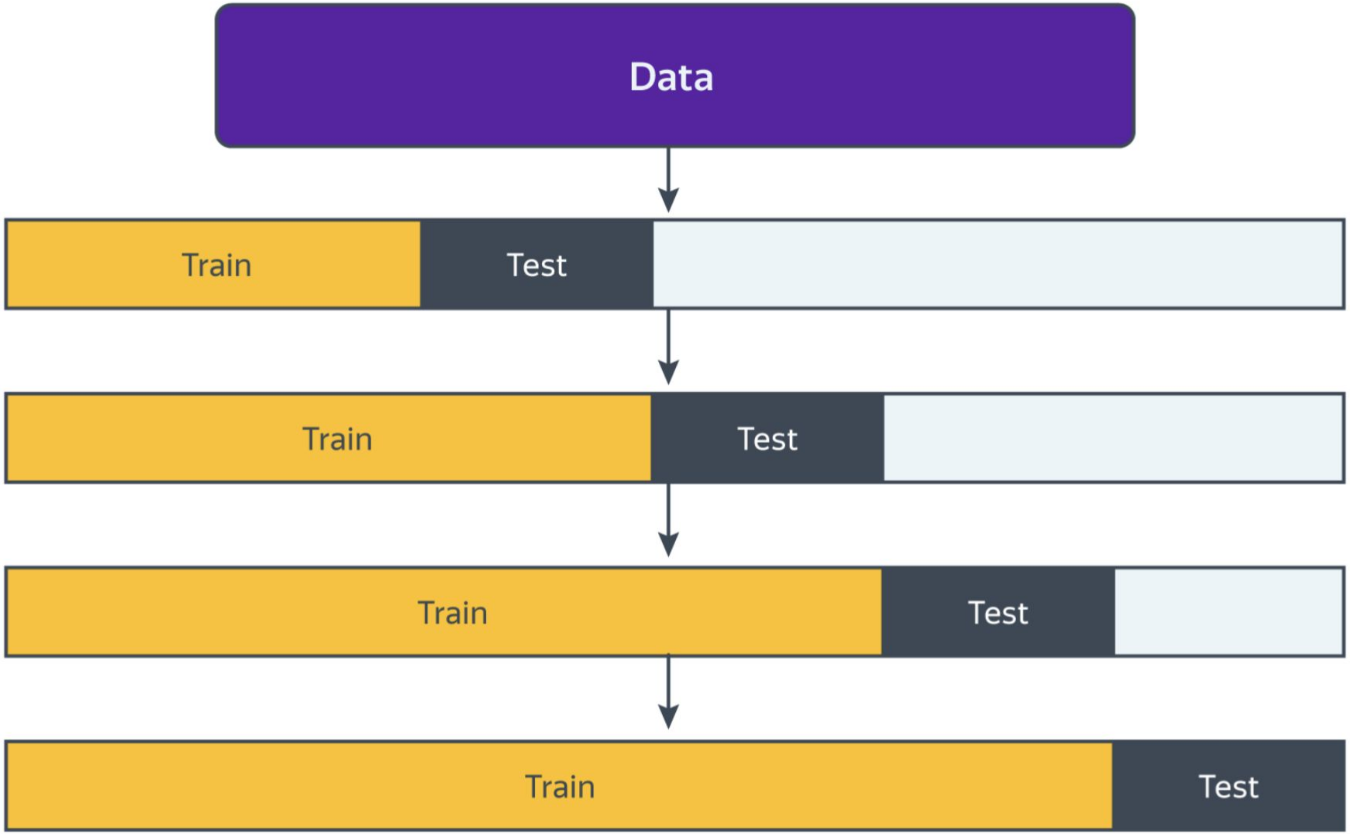
- Вес для отдельной модели:

$$w_{jt} = \frac{(\tilde{\varepsilon}_{jt})^{-1}}{\sum_{s=1}^{k} (\tilde{\varepsilon}_{jt})^{-1}}$$

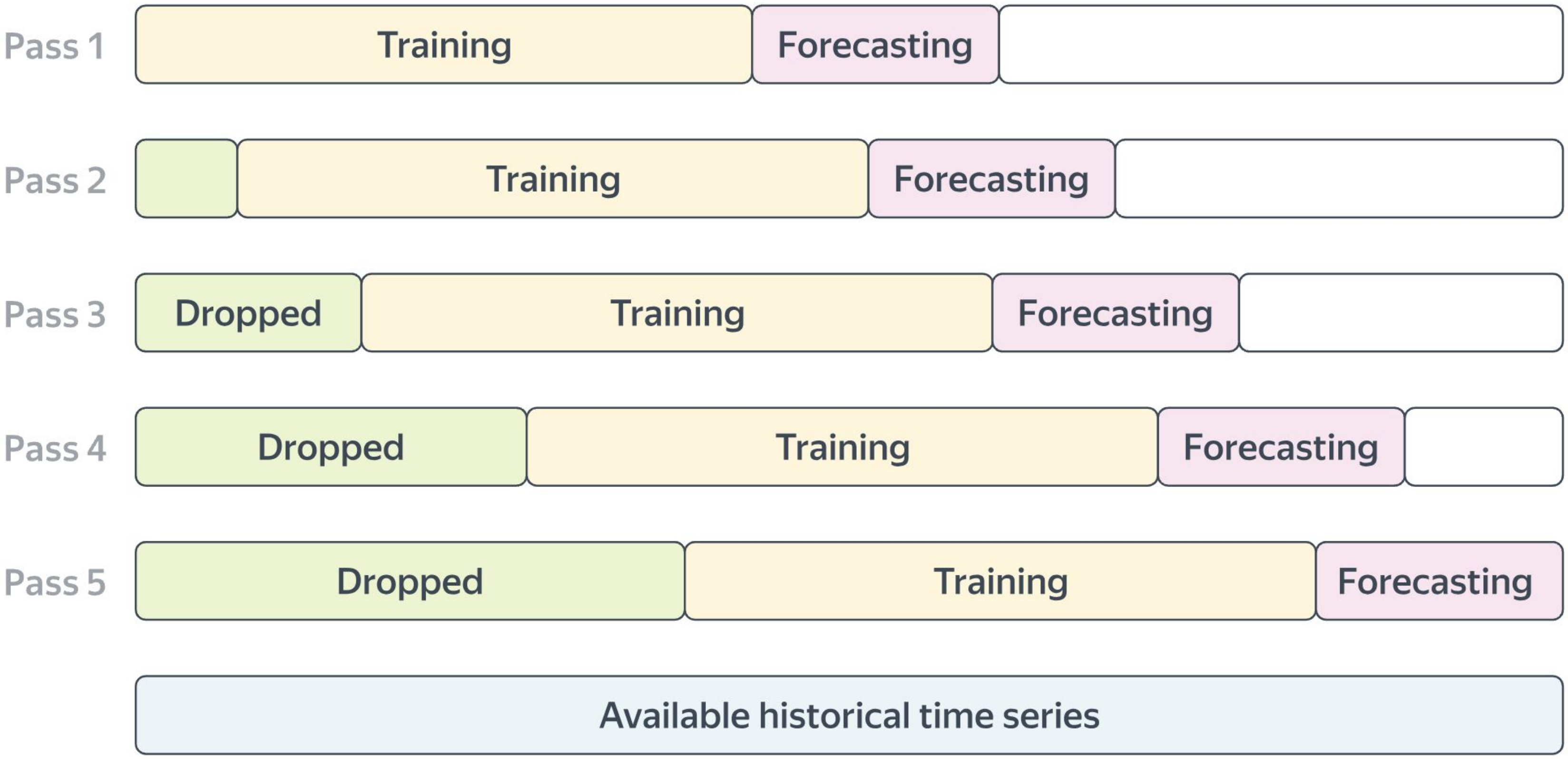
Требования к задаче прогнозирования рядов

- Визуализация данных, анализ распределения признаков
- Подготовка данных: избавление от выбросов, подготовка признаков
- Настройка эконометрических моделей (ARMA, ARIMA и тд.)
- Настройка адаптивных моделей – автоматический подбор модели, проверка соответствия модели особенностям ряда
- Настройка регрессионных моделей машинного обучения
- Визуальный анализ
- Сравнение и выбор наилучшей модели
- Выводы

Кросс-валидация для временных рядов



Кросс-валидация для временных рядов



Модели машинного обучения

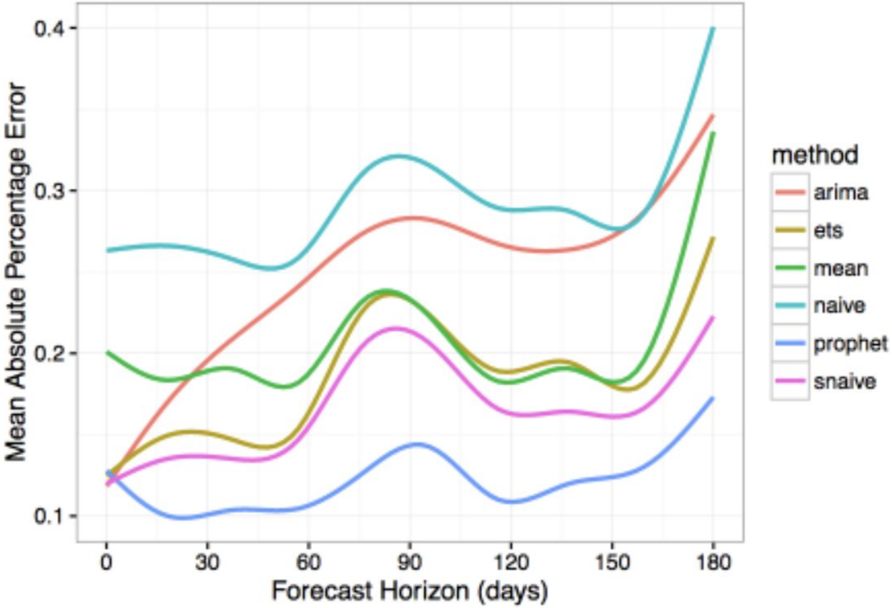
- Часто затраты на настройку моделей (ARMA, ARIMA и тд.) не окупаются, поэтому имеет смысл попробовать применить методы машинного обучения к предсказанию временных рядов
- Можно использовать линейную регрессию, в качестве признаков для которой использовать лаговые признаки (значения признака предыдущие периоды времени).

Машинное обучение: признаки

1. Лаги – значения временного ряда 1, 2, … периодов назад
2. Агрегированные признаки по дате: среднее значение таргета для каждого дня недели месяца, года и тд.
3. Другие характеристик

Facebook Prophet

- Facebook prophet от компании Facebook для задач предсказания временных рядов



Что ещё почитать по временным рядам

- Пост на хабре по основам анализа и обработки врменных рядов https://habr.com/ru/companies/ods/articles/327242/
- Лекция К.В. Воронцова http://www.machinelearning.ru/wiki/images/archive/c/cb/20160412121749!voron-ml-forecasting-slides.pdf
- Учебник от Яндекса, главы по работе с временными рядами: https://education.yandex.ru/handbook/ml/article/vremennye-ryady