

Анализ временных рядов

Максим Карпов



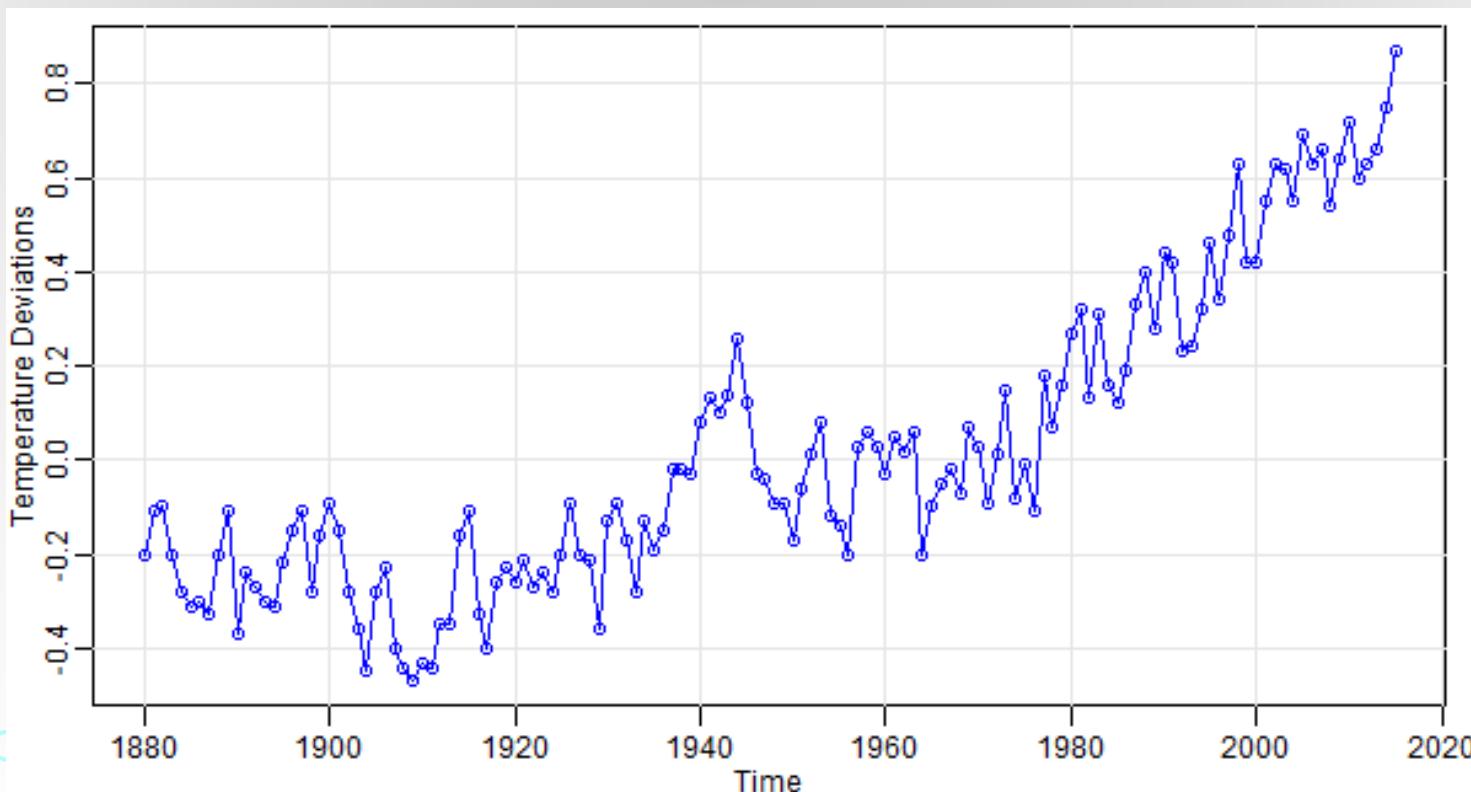
Центр непрерывного образования ФКН ВШЭ

ПЛАН ЛЕКЦИИ

- Общие понятия временных рядов
- Эконометрические модели
- Адаптивные модели
- Модели машинного обучения

ВРЕМЕННОЙ РЯД

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



ЗАДАЧА ПРОГНОЗИРОВАНИЯ

$y_0, y_1, \dots, y_t, \dots$ - временной ряд, $y_i \in \mathbb{R}$.

Задача: построить функцию

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

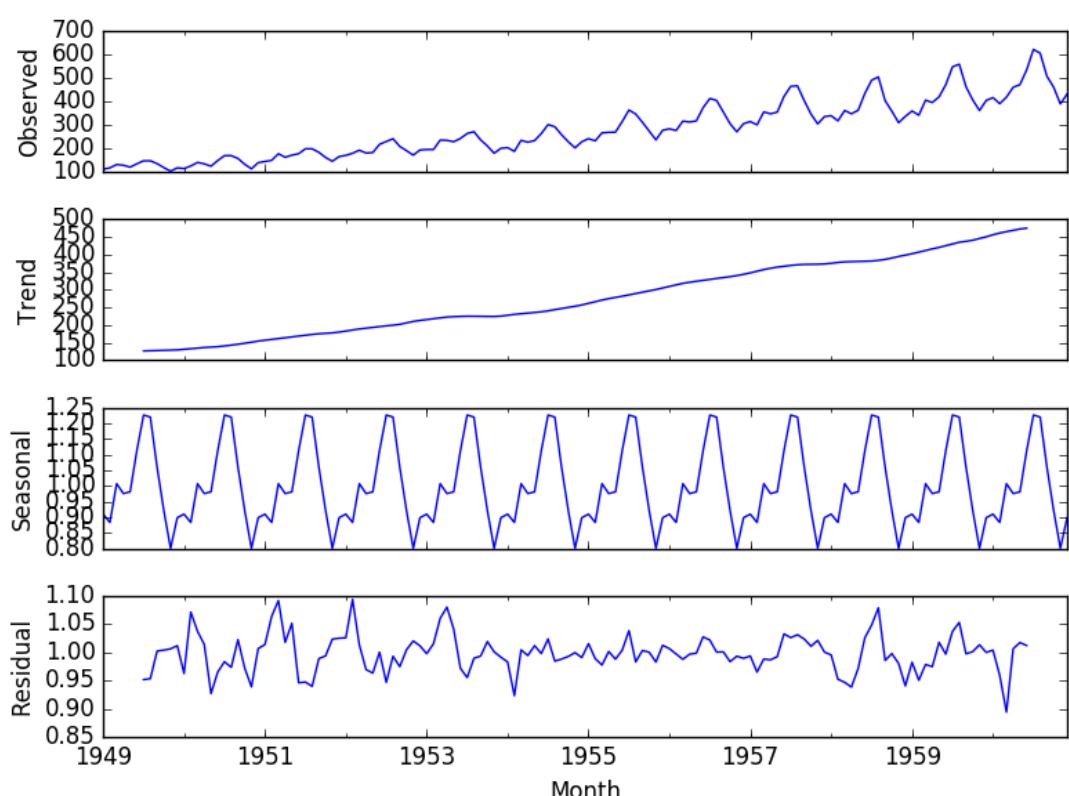
- $d = 1, \dots, D$, где D – горизонт прогнозирования
- w – вектор параметров модели

ОСОБЕННОСТЬ ВРЕМЕННЫХ РЯДОВ

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

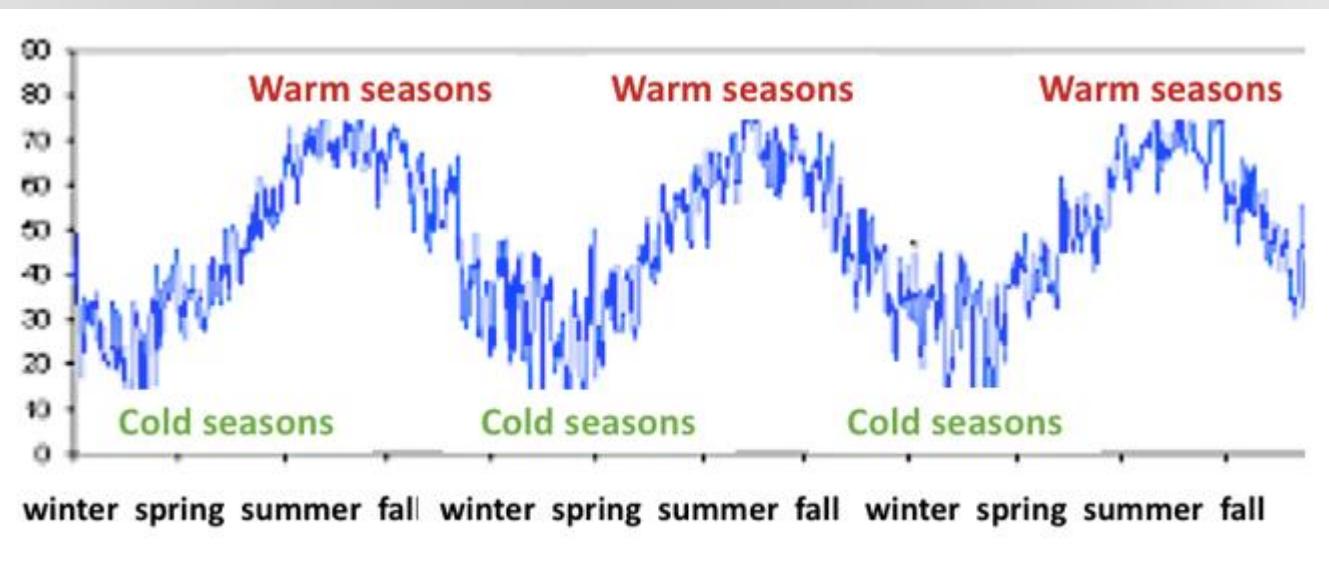
КОМПОНЕНТЫ ВРЕМЕННОГО РЯДА

- Тренд – плавное долгосрочное изменение уровня ряда
- Сезонность – циклические изменения уровня ряда с постоянным периодом
- Циклы – изменения уровня ряда с переменным периодом (цикл жизни товара, экономические волны, периоды солнечной активности)
- Ошибка (шум) – непрогнозируемая случайная компонента ряда

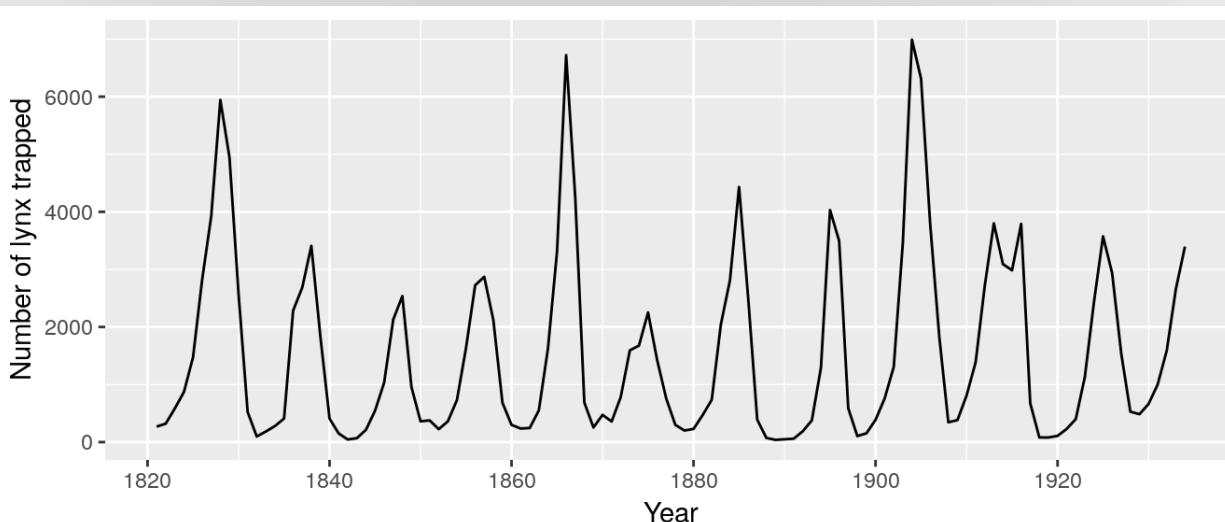


ЦИКЛЫ И СЕЗОННОСТЬ

Сезонность:



Цикл:



СТАЦИОНАРНОСТЬ

Ряд y_1, \dots, y_T **стационарен**, если для любого s распределение y_t, \dots, y_{t+s} не зависит от t , то есть его свойства не зависят от времени.

- тренд \Rightarrow нестационарность
- сезонность \Rightarrow нестационарность
- цикл – заранее неизвестно

СТАЦИОНАРНОСТЬ

Ряд y_1, \dots, y_T *стационарен*, если для любого s распределение y_t, \dots, y_{t+s} не зависит от t , то есть его свойства не зависят от времени.

- тренд \Rightarrow нестационарность
- сезонность \Rightarrow нестационарность
- цикл – заранее неизвестно

По стационарному ряду просто построить прогноз, так как мы полагаем, что его будущие статистические характеристики не будут отличаться от наблюдаемых текущих.

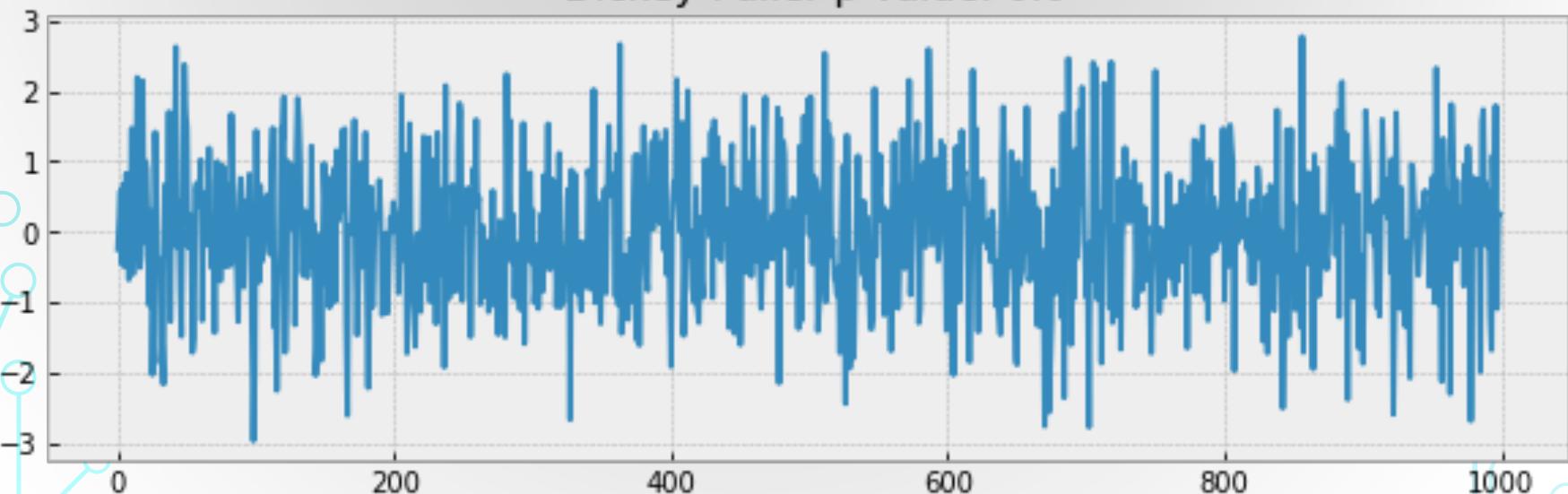
Для проверки стационарности ряда можно использовать критерий Дики-Фуллера.

ЕДИНИЧНЫЙ КОРЕНЬ

Рассмотрим модель временного ряда $X_t = \rho \cdot X_{t-1} + \varepsilon_t$,
где ε_t - ошибка, не зависящая от значений временного ряда.

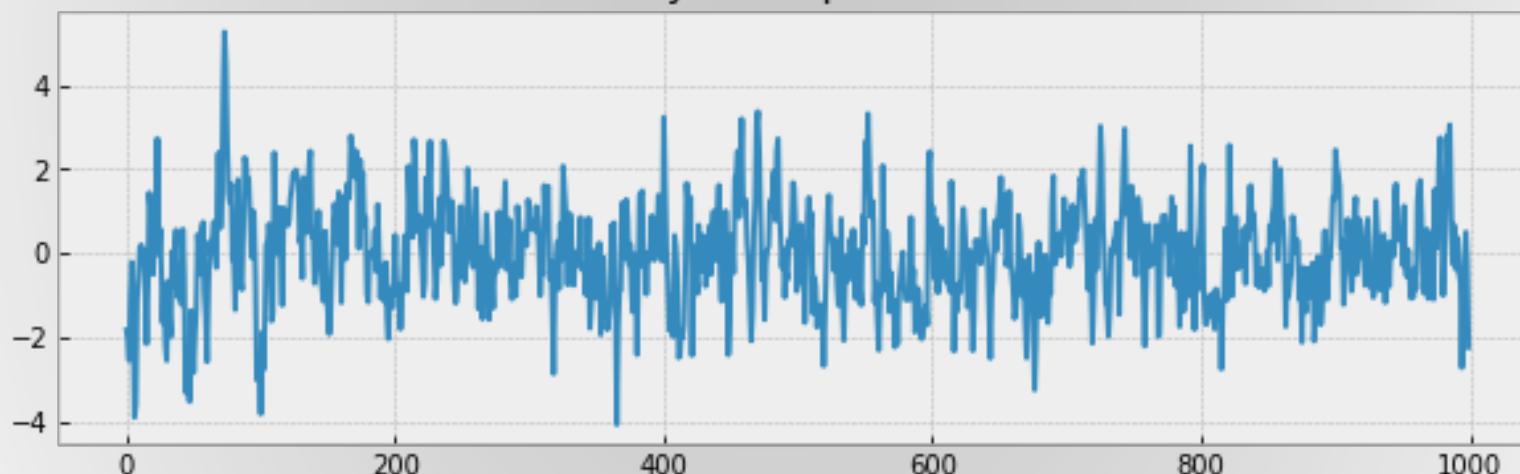
Определение. Если $\rho = 1$, то говорят, что ряд имеет единичный корень.

Rho 0
Dickey-Fuller p-value: 0.0

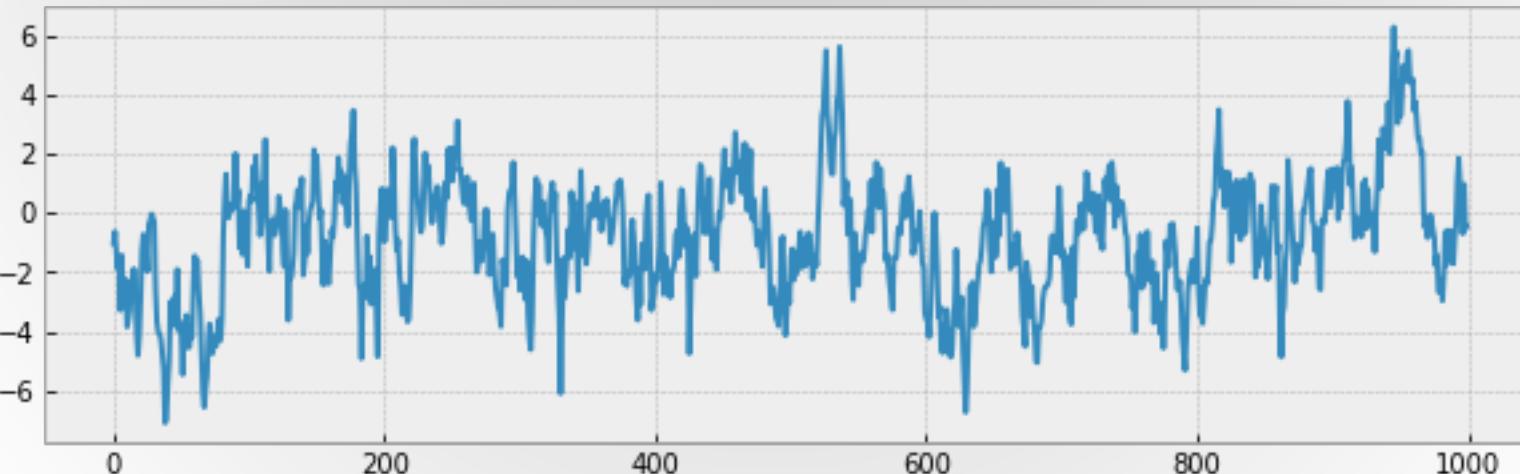


ЕДИНИЧНЫЙ КОРЕНЬ

Rho 0.6
Dickey-Fuller p-value: 0.0

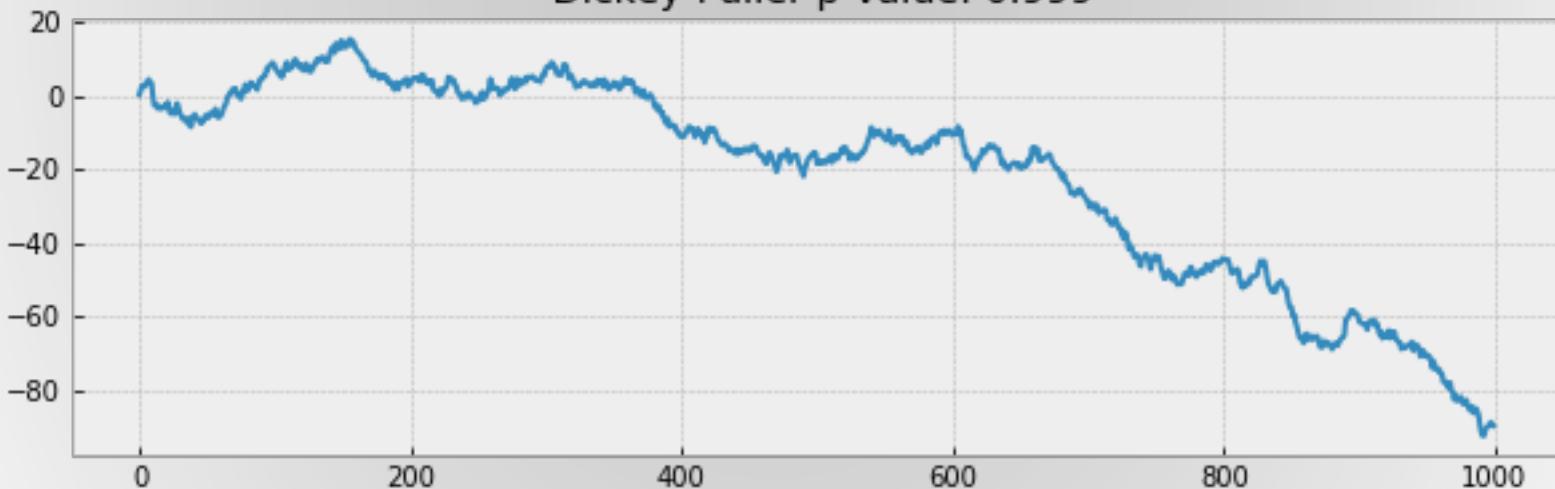


Rho 0.9
Dickey-Fuller p-value: 0.0



ЕДИНИЧНЫЙ КОРЕНЬ

Rho 1
Dickey-Fuller p-value: 0.999



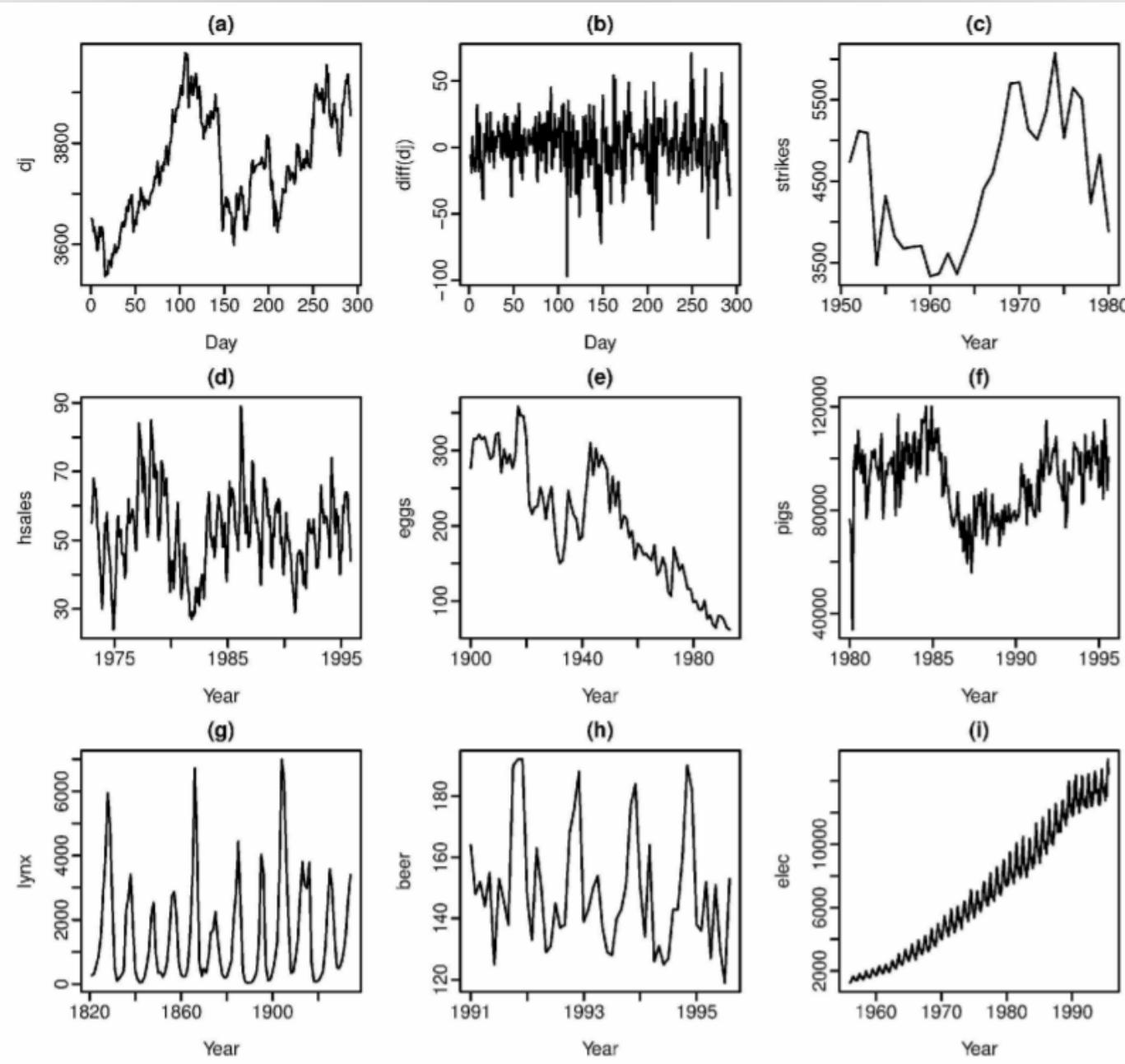
Видно, что при $\rho = 1$ процесс не возвращается к своему среднему, а значит, не является стационарным.

ПРОВЕРКА СТАЦИОНАРНОСТИ РЯДА

Проверку стационарности ряда можно осуществлять с помощью критерия Дики-Фуллера.

- Критерий Дики-Фуллера проверяет гипотезу $\rho = 1$.

ПРИМЕРЫ ВРЕМЕННЫХ РЯДОВ



МЕТОДЫ ИЗБАВЛЕНИЯ ОТ НЕСТАЦИОНАРНОСТИ

1. Стабилизация дисперсии

- для рядов с монотонно меняющейся дисперсией можно использовать стабилизирующее **преобразование Бокса-Кокса** (λ – параметр метода):

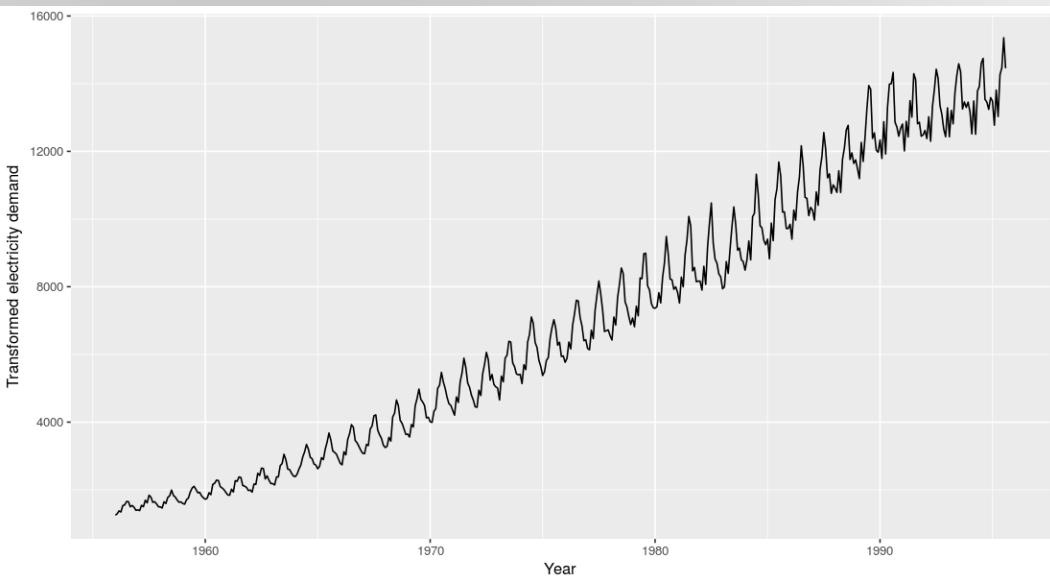
$$y'_t = \begin{cases} \ln y_t, & \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases}$$

- логарифмирование – частный случай

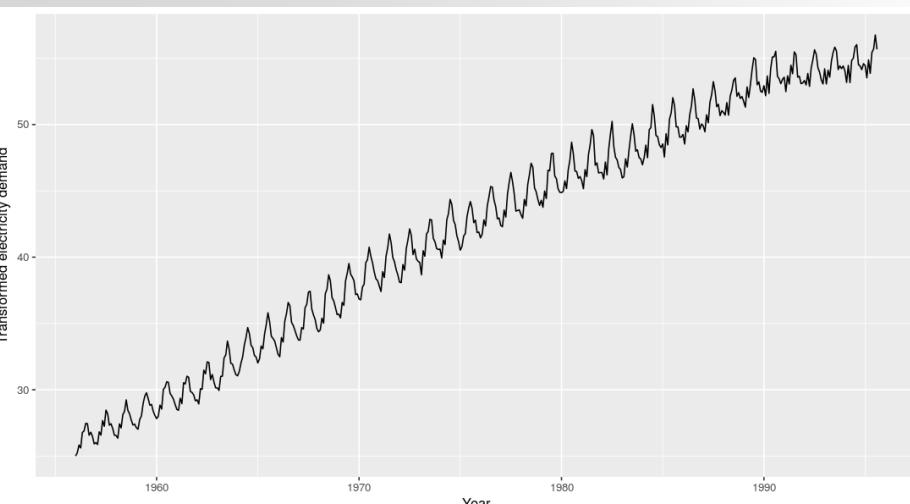
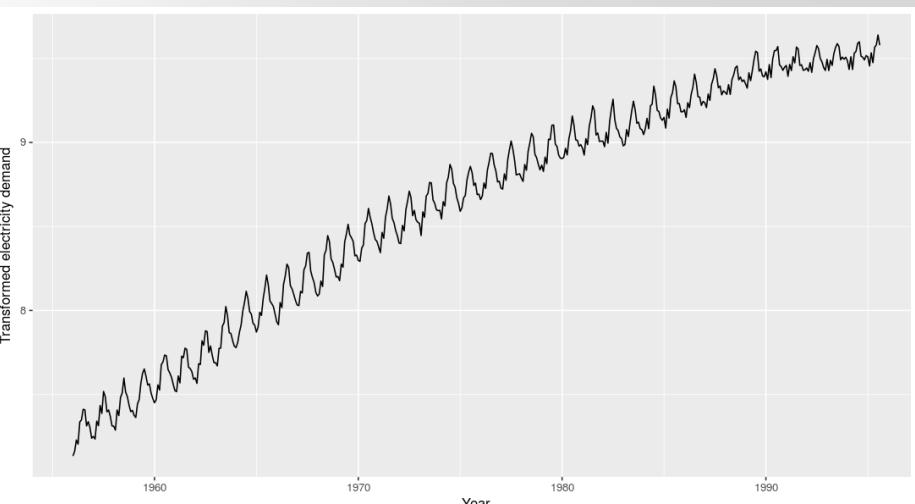
Параметр λ подбирается так, чтобы сделать дисперсию как можно более однородной.

СТАБИЛИЗАЦИЯ ДИСПЕРСИИ (ПРИМЕР)

Исходный ряд:



Преобразование Бокса-Кокса с $\lambda = 0$ (слева) и $\lambda = 0.3$ (справа):

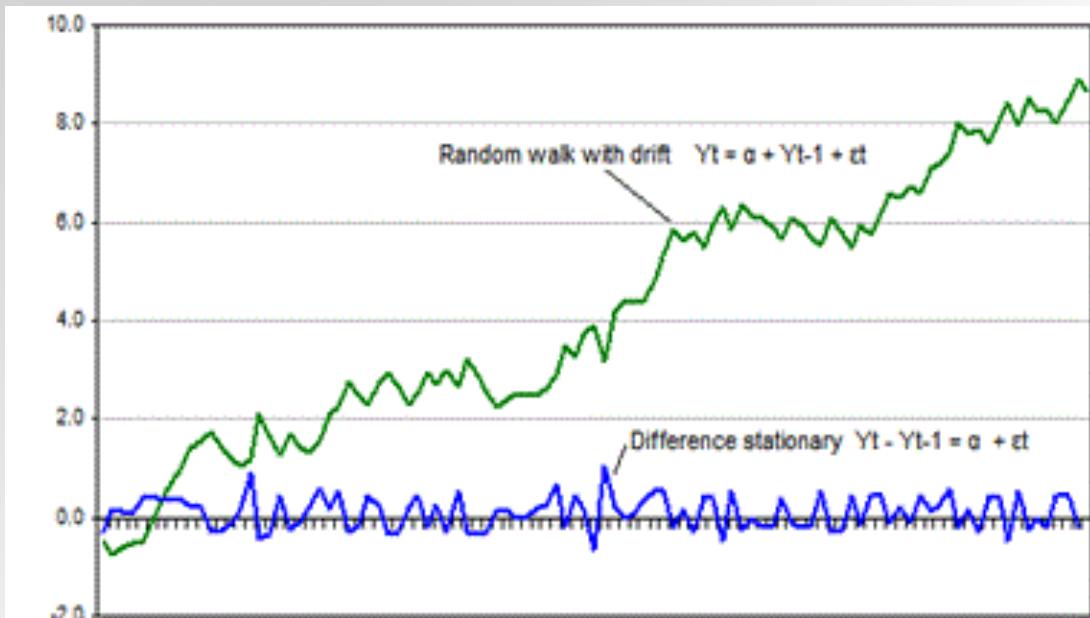


МЕТОДЫ ИЗБАВЛЕНИЯ ОТ НЕСТАЦИОНАРНОСТИ

2. 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



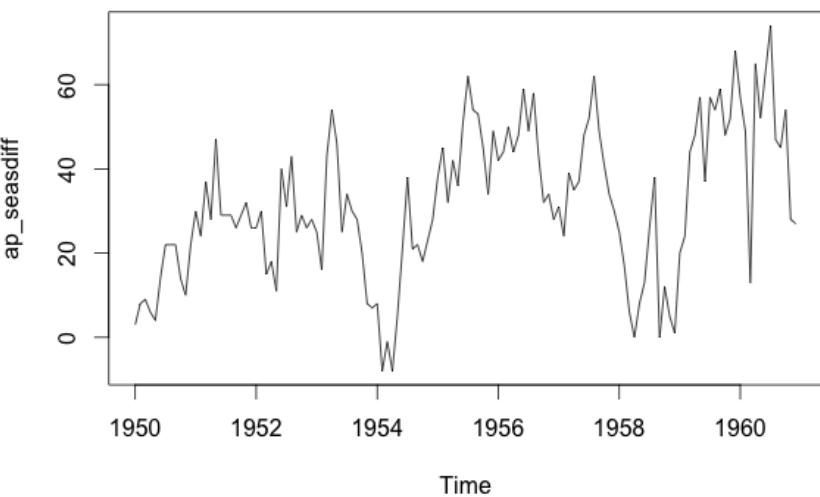
МЕТОДЫ ИЗБАВЛЕНИЯ ОТ НЕСТАЦИОНАРНОСТИ

3. Сезонное дифференцирование – переход к попарным разностям значений в соседних сезонах

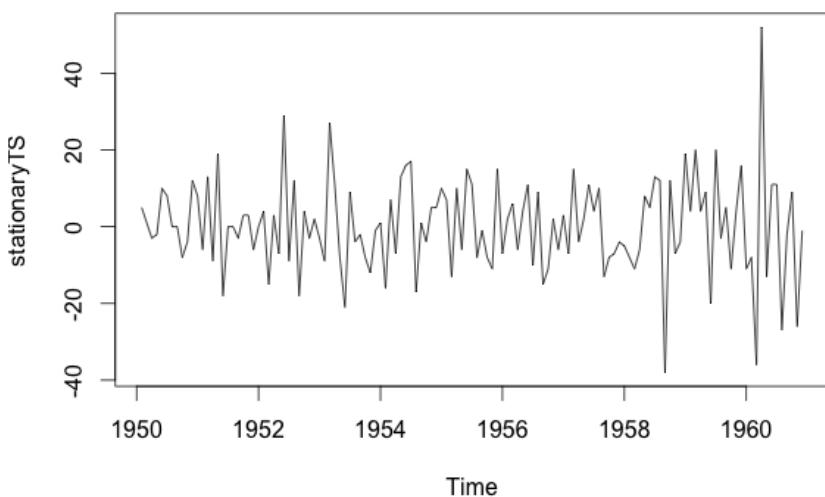
$$y'_t = y_t - y_{t-s}$$

- убирает сезонность

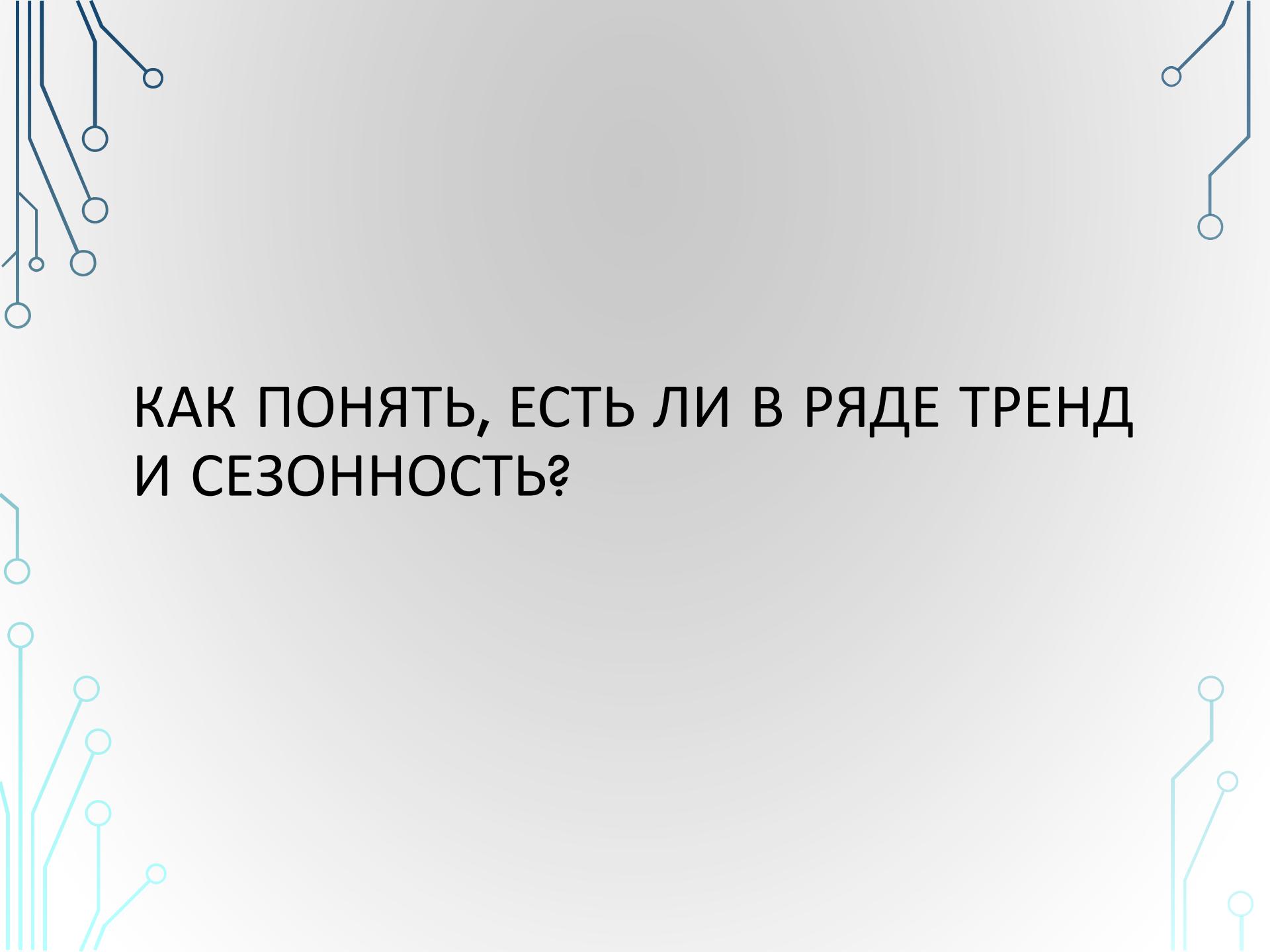
Seasonally Differenced



Differenced and Stationary



Сезонное дифференцирование лучше применять в начале
– возможно, после него ряд уже станет стационарным.



КАК ПОНЯТЬ, ЕСТЬ ЛИ В РЯДЕ ТРЕНД
И СЕЗОННОСТЬ?

АВТОКОРРЕЛЯЦИЯ

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

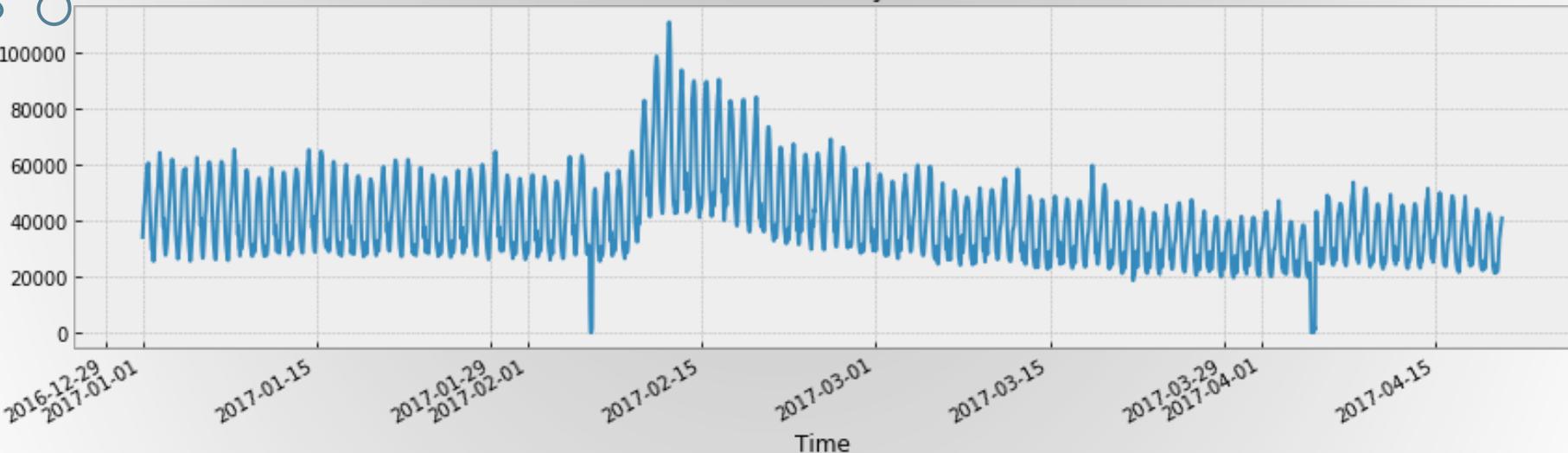
Определение. Корреляция (зависимость) между последовательными элементами временного ряда – это автокорреляция.

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

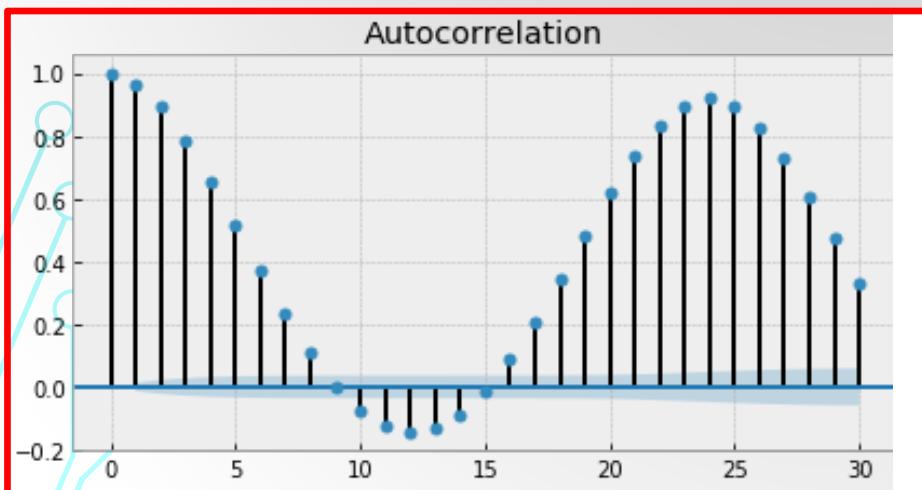
АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

Множество коэффициентов корреляции для различных смещений и есть *автокорреляционная функция (ACF)*.

Time Series Analysis Plots



Autocorrelation

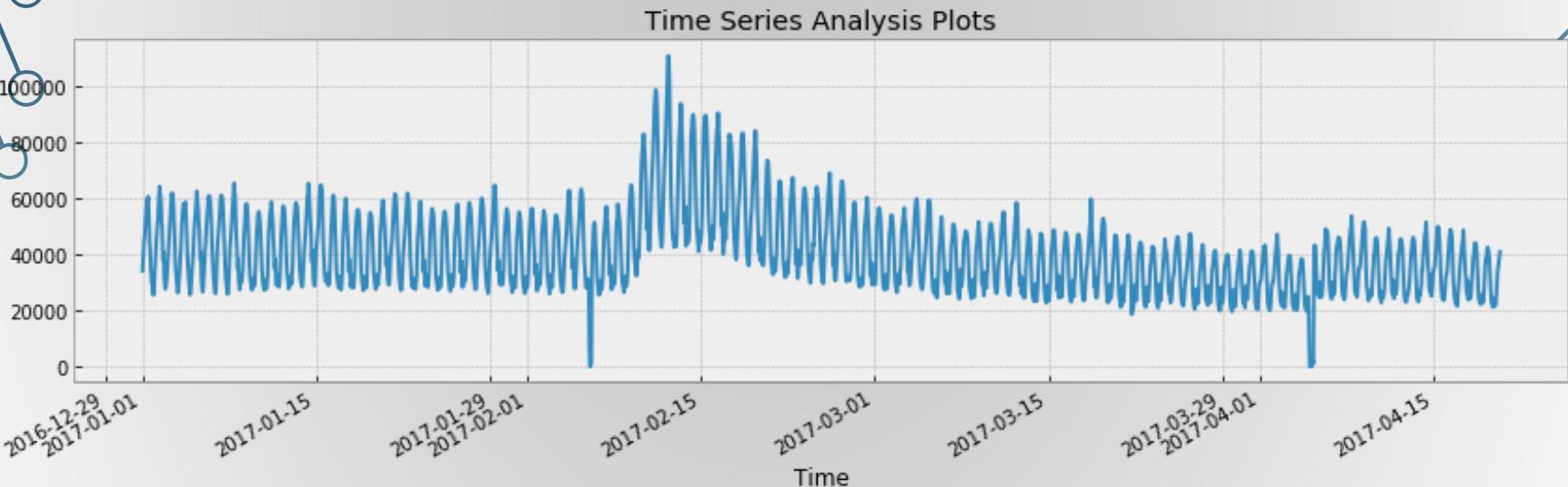


ЧАСТНАЯ АВТОКОРРЕЛЯЦИЯ

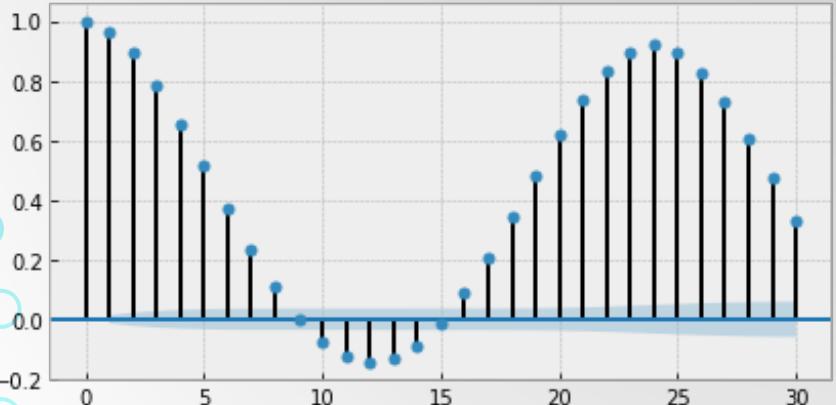
Частная (частичная) автокорреляция – это часть корреляции между моментами времени X_{t-k} и X_t , которая не объясняется промежуточными корреляциями.



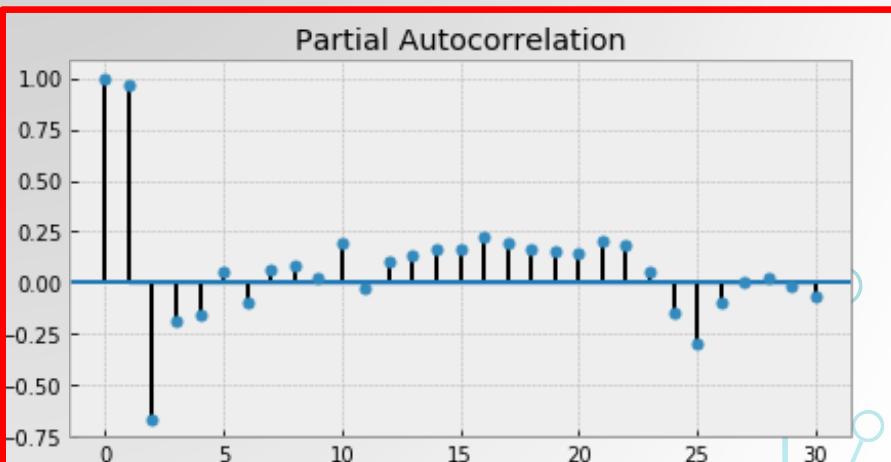
ЧАСТНАЯ АВТОКОРРЕЛЯЦИЯ



Autocorrelation



Partial Autocorrelation



ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

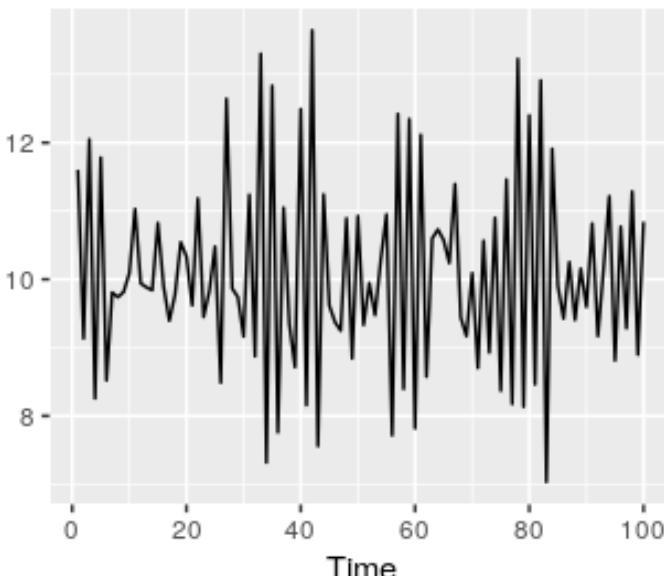
МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

- Модель авторегрессии $AR(p)$:

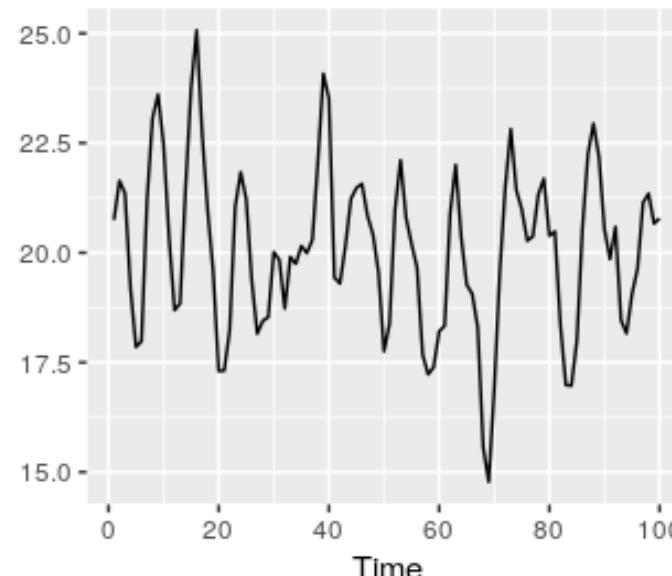
$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t,$$

$a_p \neq 0$, ε_t - процесс белого шума, $E \varepsilon_t = 0$, $D \varepsilon_t = \sigma_\varepsilon^2$, $cov(\varepsilon_t, \varepsilon_s) = 0$, некоррелируемый с y_t .

AR(1)



AR(2)



ЛИНЕЙНАЯ МОДЕЛЬ АВТОРЕГРЕССИИ

- признаки - n предыдущих наблюдений ряда:

$$\hat{y}_{t+1}(w) = \sum_{j=1}^n w_j y_{t-j+1}, \quad w \in \mathbb{R}^n$$

- объекты - $t - n + 1$ моментов в истории ряда:

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

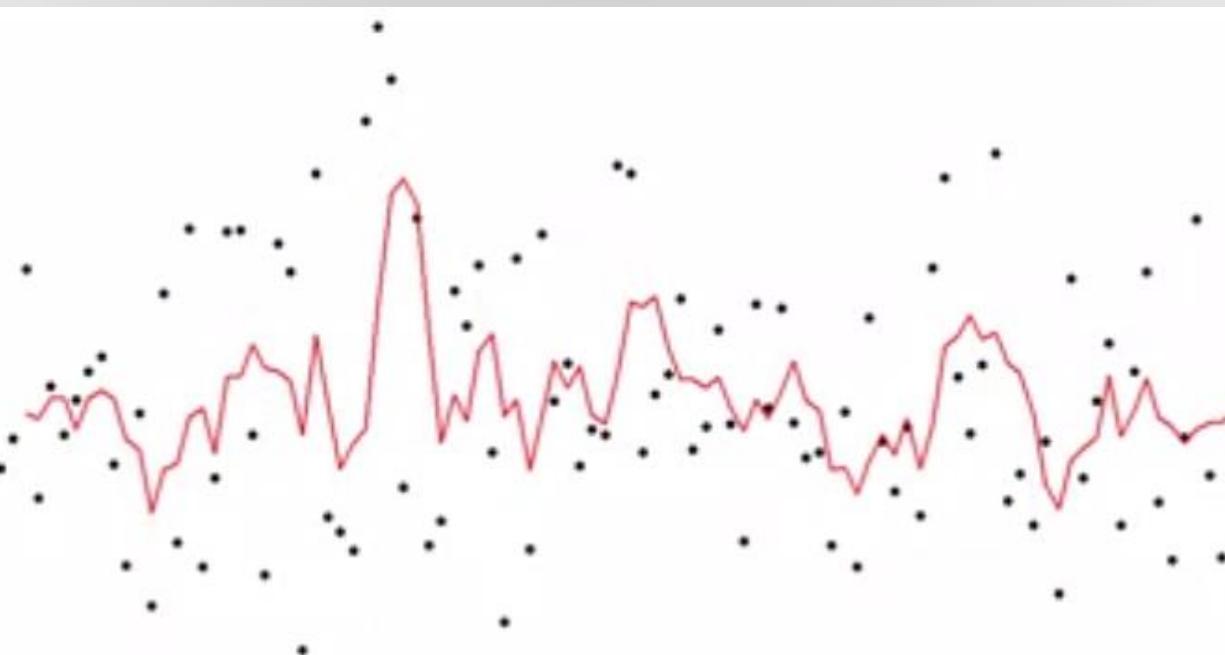
$$Q(w) = \sum_{i=n}^t (\hat{y}_i(w) - y_i)^2 = \|Fw - y\|^2 \rightarrow \min_w$$

МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

- Процесс скользящего среднего порядка q ($MA(q)$):

$$y_t = \alpha + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q},$$

$b_q \neq 0$, ε_t - процесс белого шума, $E \varepsilon_t = 0$, $D \varepsilon_t = \sigma_\varepsilon^2$, $cov(\varepsilon_t, \varepsilon_s) = 0$, некоррелируемый с y_t . Такой процесс всегда стационарен.

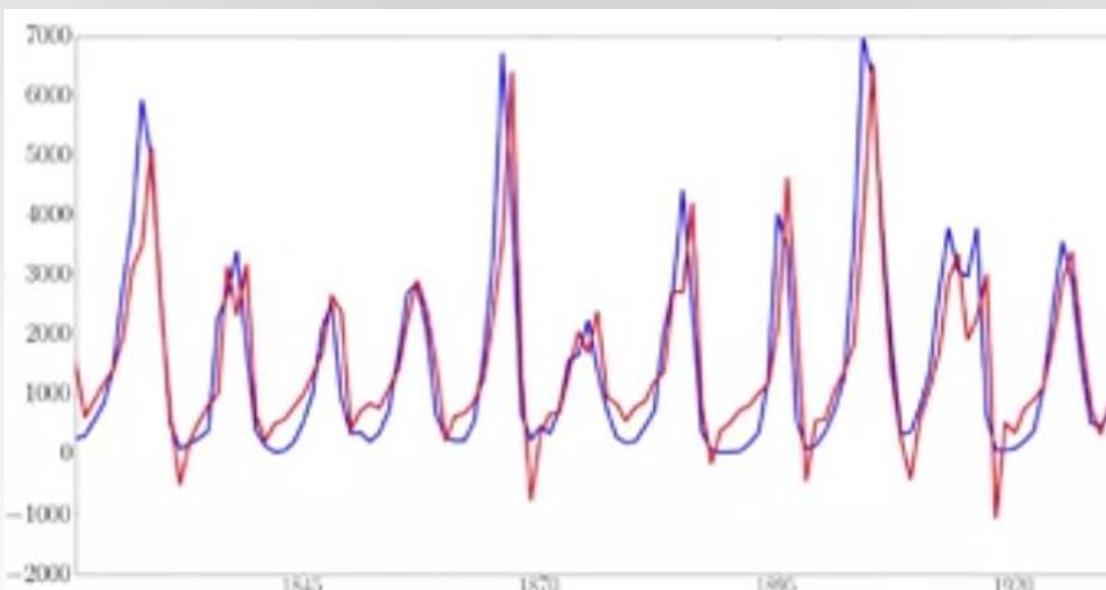


МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

- Смешанный процесс авторегрессии $ARMA(p, q)$:

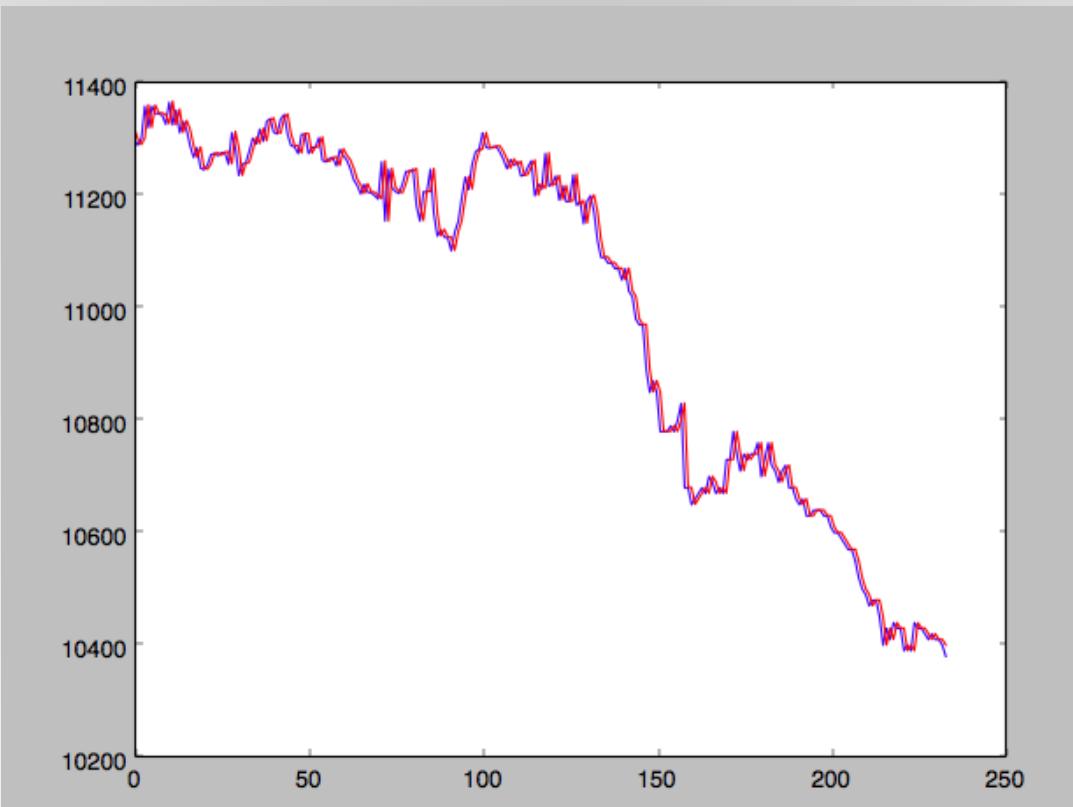
$$y_t = \alpha + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q},$$
$$a_p, b_q \neq 0.$$

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



МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

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



- Модель $SARIMA$ – модель ARMA с учетом наличия тренда и сезонности

АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

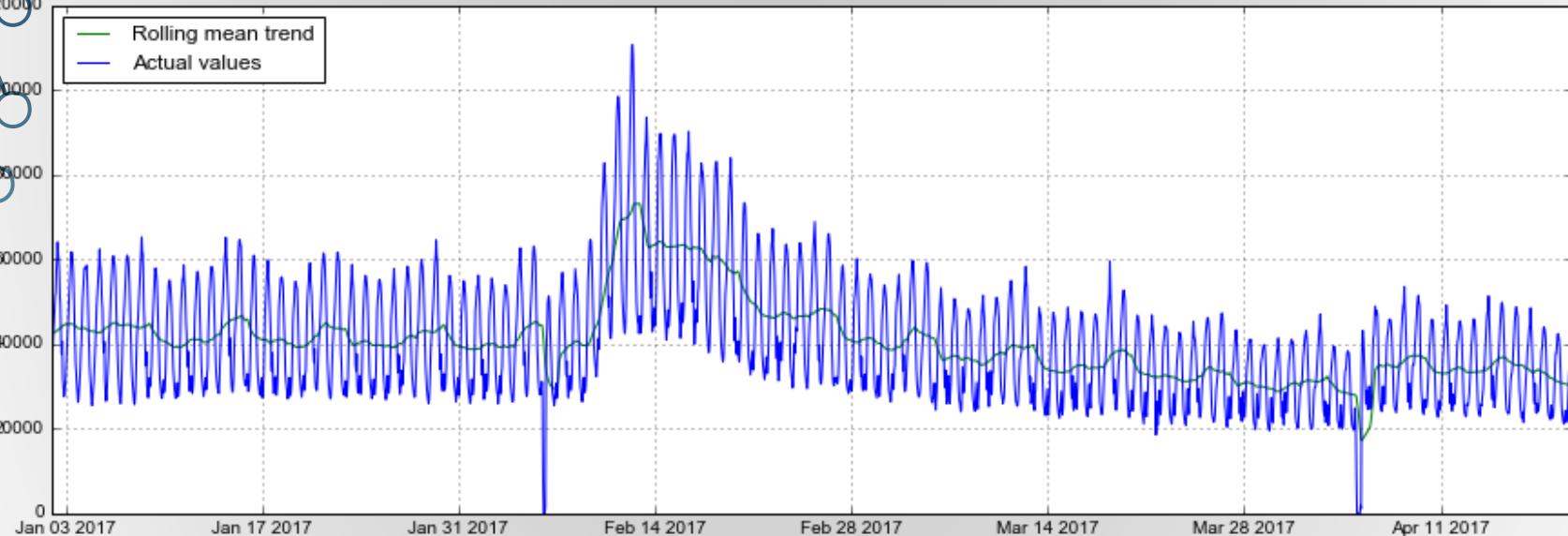
СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

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

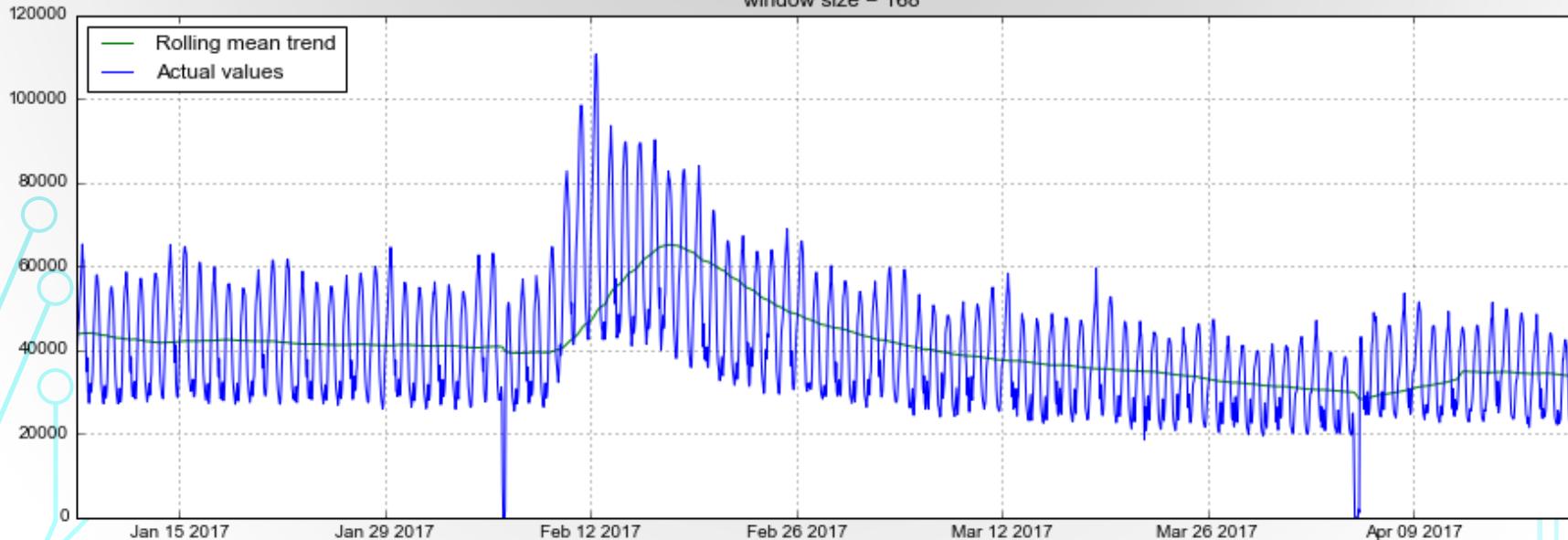
- чтобы сделать прогноз на следующий период времени, надо знать значение на текущий период (т.е. долгосрочный прогноз невозможен)
- + сглаживает данные

СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

Moving average
window size = 24



Moving average
window size = 168



ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

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

Пример:

$$EMA(t) = \frac{1}{2} p_t + \frac{1}{4} p_{t-1} + \frac{1}{8} p_{t-2} + \dots$$

ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

- Простейшая регрессионная модель – константа ($c = c_t$):

$$\hat{y}_{t+1} = c$$

Минимизируем квадратичную ошибку с весами β^i , убывающими в прошлое:

$$\sum_{i=0}^t \beta^i (y_{t-i} - c)^2 \rightarrow \min_c$$

Аналитическое решение (формула Надаля-Ватсона):

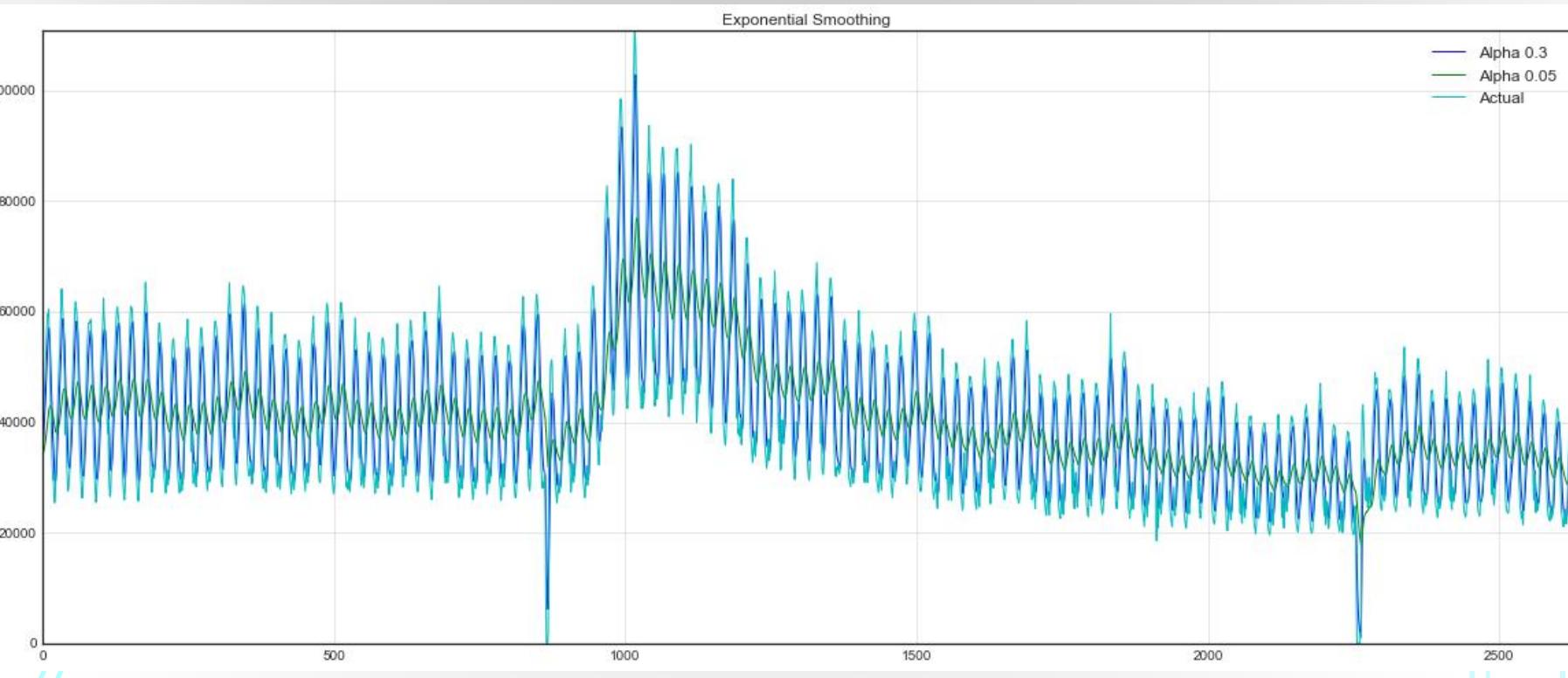
$$\hat{y}_{t+1} = \frac{\sum_{i=0}^t \beta^i y_{t-i}}{\sum_{i=0}^t \beta^i}$$

ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ (ЭСС)

Модель ЭСС можно записать в виде

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t, \alpha \in (0, 1)$$

- чем больше α , тем больше вес последних точек
- чем меньше α , тем сильнее сглаживание



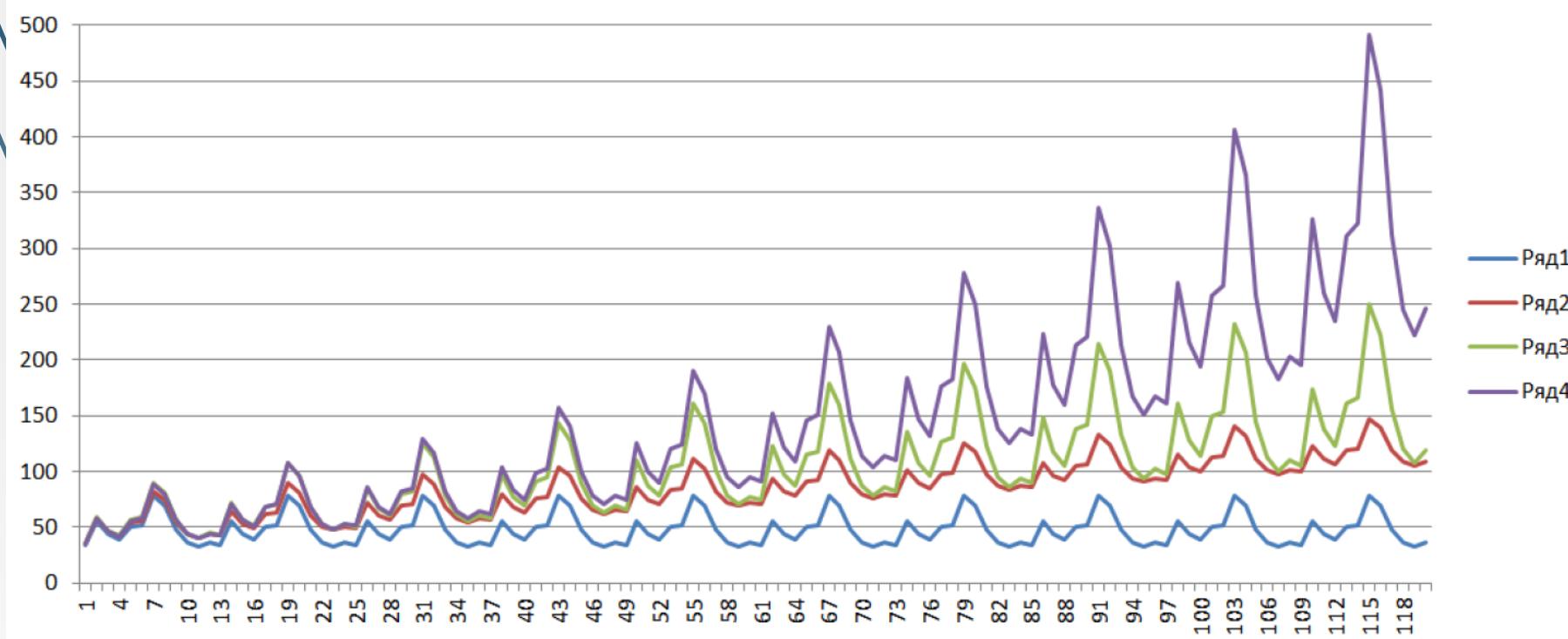
ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ (ЭСС)

Оптимальное значение α подбираем по скользящему контролю:

$$Q(\alpha) = \sum_{t=t_0}^T (\hat{y}_t(\alpha) - y_t)^2 \rightarrow \min_{\alpha}$$

- при $\alpha \in (0,0.3)$ ряд стационарен, модель ЭСС работает
- при $\alpha \in (0.3,1)$ ряд нестационарен, нужна модель тренда

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ



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

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Хольта

- модель линейного тренда

$$\hat{y}_{t+d} = a_t + \mathbf{b}_t \mathbf{d},$$

где a_t, b_t - адаптивные компоненты линейного тренда.

- формулы для a_t, b_t :

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

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

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,

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

формулы для a_t, b_t, θ_t :

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

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

$$\theta_t = \alpha_3(y_t - a_t) + (1 - \alpha_3)\theta_{t-s},$$

где $\alpha_1, \alpha_2, \alpha_3$ - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса

- модель мультипликативной сезонности периода s

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

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

формулы для a_t, b_t, θ_t :

$$a_t = \alpha_1(y_t/\theta_{t-s}) + (1 - \alpha_1) a_{t-1}$$

$$\theta_t = \alpha_2(y_t/a_t) + (1 - \alpha_2)\theta_{t-s},$$

где α_1, α_2 - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса с линейным трендом

- модель мультипликативной сезонности периода s с линейным трендом

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

$a_t + b_t d$ – тренд, очищенный от сезонных колебаний,

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

формулы для a_t, b_t, θ_t :

$$a_t = \alpha_1(y_t/\theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1})$$

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

$$\theta_t = \alpha_3(y_t/a_t) + (1 - \alpha_3)\theta_{t-s},$$

где $\alpha_1, \alpha_2, \alpha_3$ - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса с экспоненциальным трендом

- модель мультипликативной сезонности периода s с экспоненциальным трендом

$$\hat{y}_{t+d} = a_t(r_t)^d \cdot \theta_{t+(d \bmod s)-s},$$

$a_t(r_t)^d$ – тренд, очищенный от сезонных колебаний,

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

формулы для a_t, b_t, θ_t :

$$a_t = \alpha_1(y_t/\theta_{t-s}) + (1 - \alpha_1)a_{t-1}r_{t-1}$$

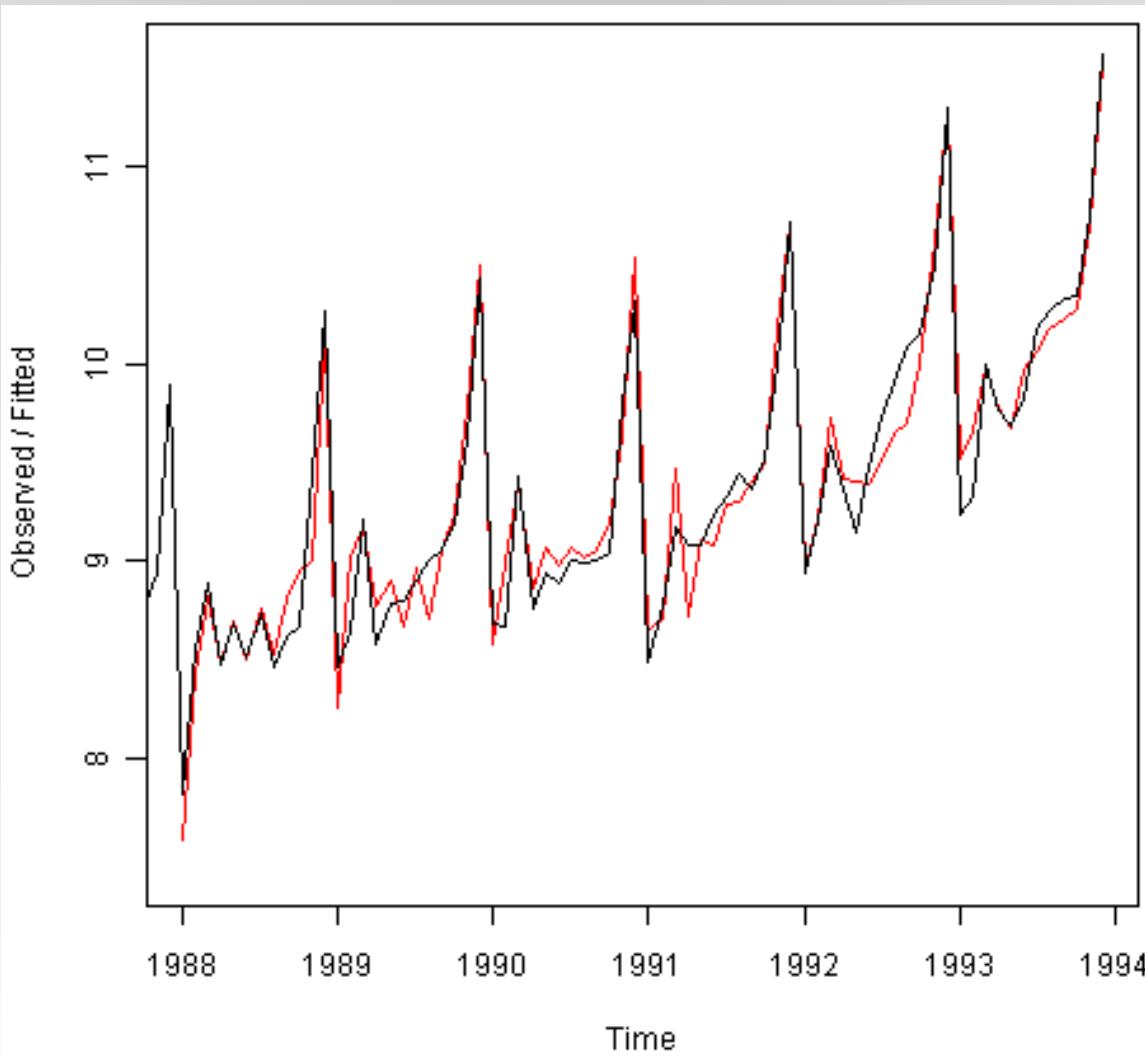
$$r_t = \alpha_2(a_t/a_{t-1}) + (1 - \alpha_2)r_{t-1}$$

$$\theta_t = \alpha_3(y_t/a_t) + (1 - \alpha_3)\theta_{t-s},$$

где $\alpha_1, \alpha_2, \alpha_3$ - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса с линейным трендом (модель Хольта-Уинтерса)



АДАПТИВНАЯ МОДЕЛЬ АВТОРЕГРЕССИИ

Линейная модель авторегрессии (линейный фильтр):

$$\hat{y}_{t+1}(w) = \sum_{j=1}^n w_j y_{t-j+1}, w \in \mathbb{R}^n$$

$\varepsilon_t = y_t - \hat{y}_t$ - ошибка прогноза \hat{y}_t , сделанного на $t - 1$ шаге

Метод наименьших квадратов: $\varepsilon_t^2 \rightarrow \min_w$

Один шаг градиентного спуска в момент времени t :

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

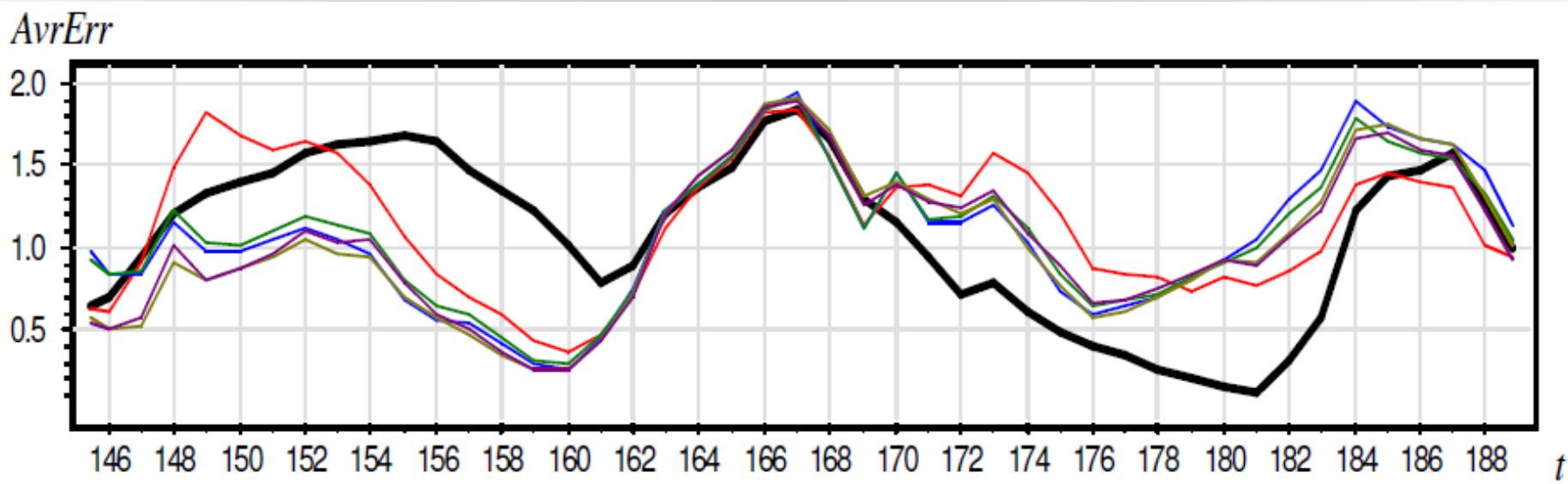
Градиентный шаг:

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

α – аналог параметра сглаживания.

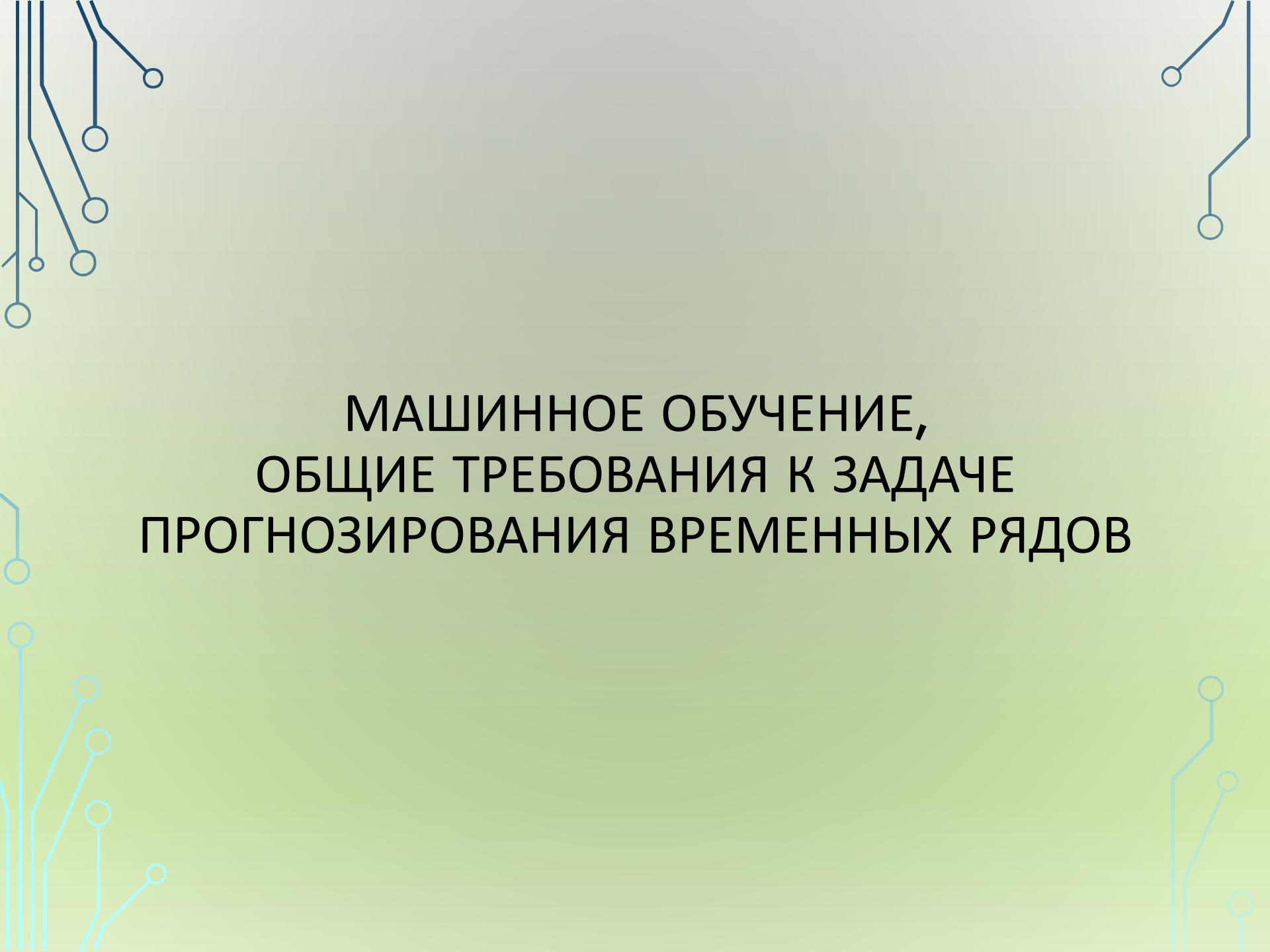
АДАПТИВНАЯ СЕЛЕКЦИЯ И КОМПОЗИЦИЯ МОДЕЛЕЙ

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



АДАПТИВНАЯ СЕЛЕКЦИЯ И КОМПОЗИЦИЯ МОДЕЛЕЙ

- Адаптивная селекция (иdea): в каждый момент времени выбираем предсказание модели, наиболее успешной на предыдущем шаге
- Адаптивная композиция (иdea): в каждый момент времени итоговое предсказание – это сумма предсказаний моделей с весами, обратно пропорциональными их ошибкам на предыдущем шаге

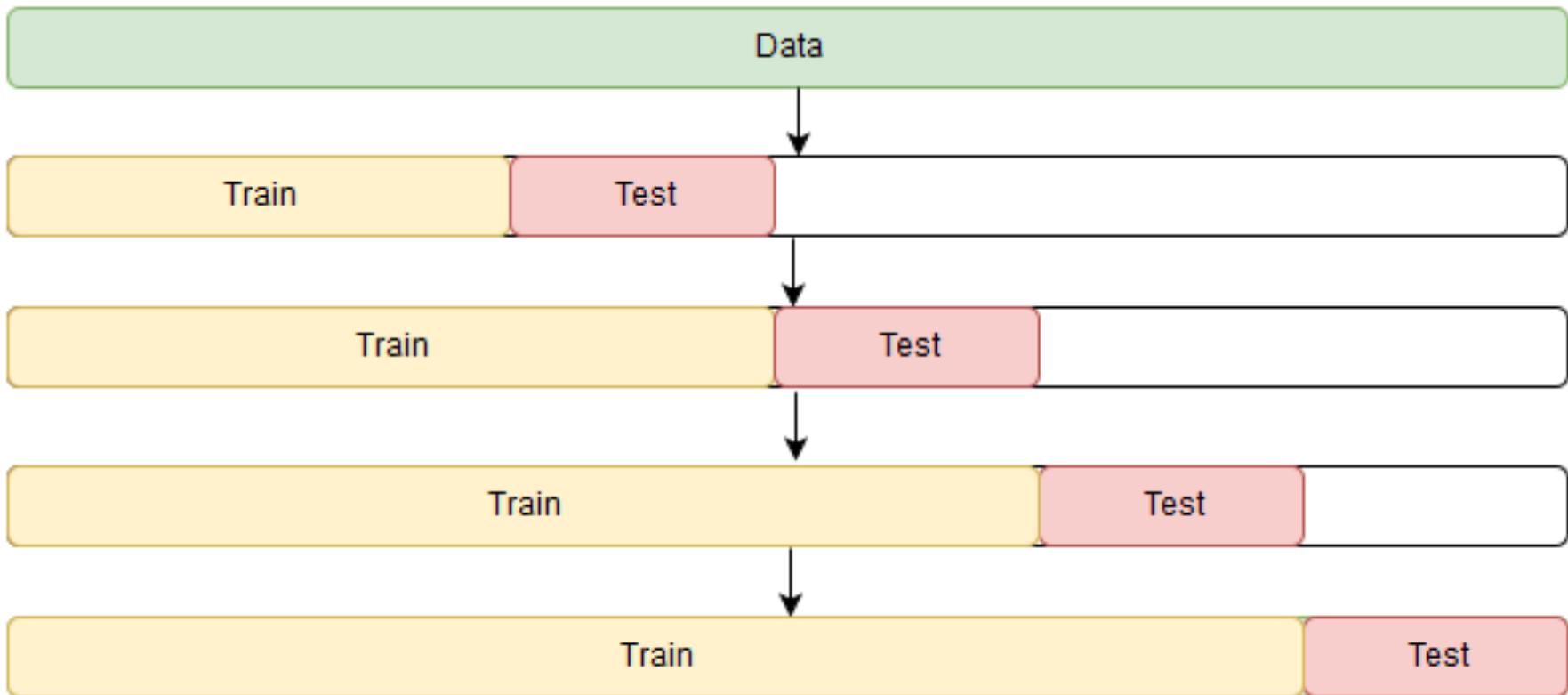


МАШИННОЕ ОБУЧЕНИЕ, ОБЩИЕ ТРЕБОВАНИЯ К ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

ТРЕБОВАНИЯ К ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

- визуализация данных, анализ распределения признака (оценка необходимости трансформации), оценка наличия выбросов, преобразования для снятия календарных эффектов
- настройка эконометрических моделей (ARIMA и пр.)
- настройка адаптивных моделей – автоматический подбор модели, проверка соответствия модели особенностям ряда, корректировка, анализ остатков
- настройка регрессионных моделей машинного обучения
 - визуальный анализ
 - сравнение и выбор наилучшей модели
 - выводы

КРОСС-ВАЛИДАЦИЯ НА ВРЕМЕННЫХ РЯДАХ



МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

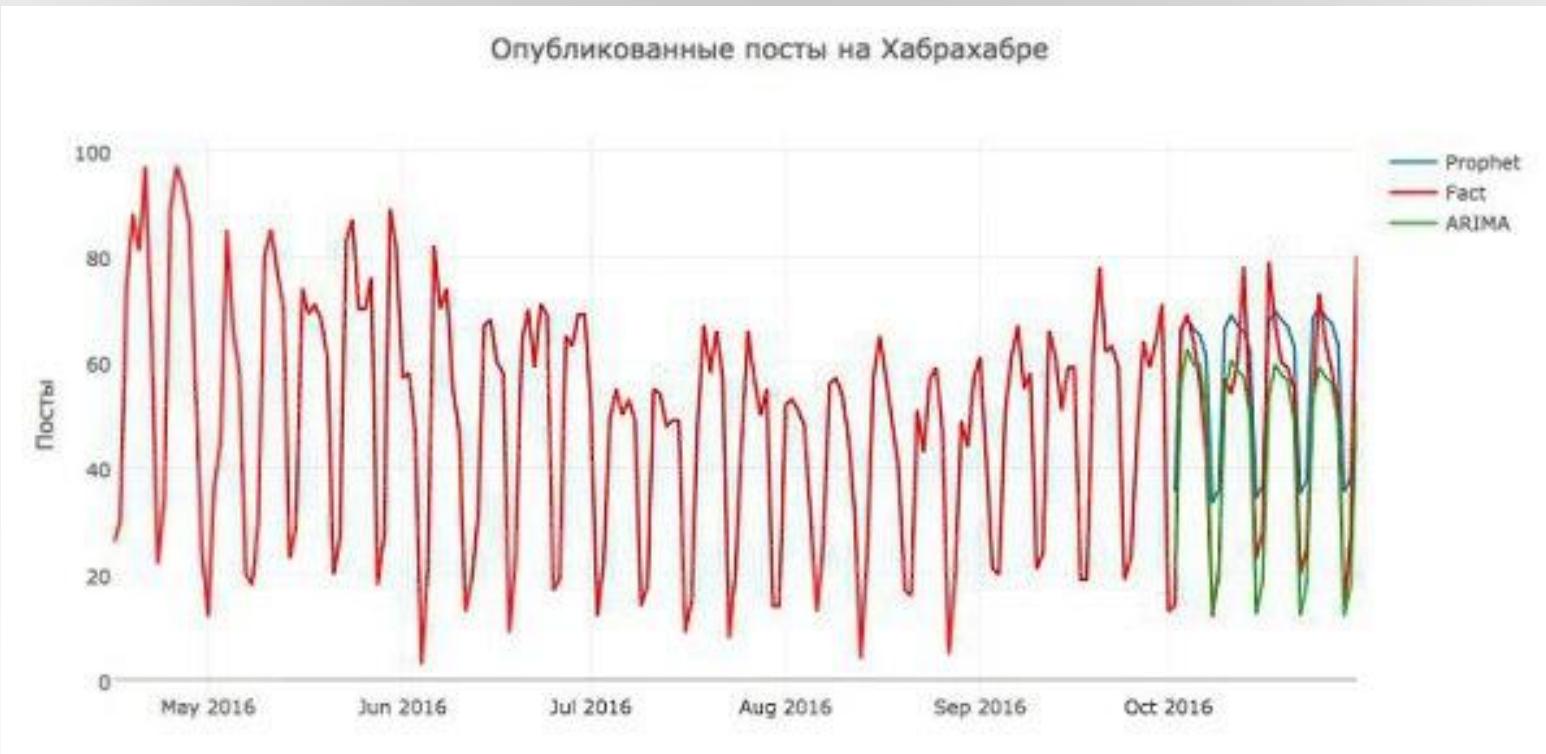
- Часто затраты на настройку моделей (ARMA, ARIMA, SARIMA и др.) не окупаются, поэтому имеет смысл попробовать применить методы машинного обучения к предсказанию временных рядов.
- Можно, например, использовать линейную регрессию, в качестве признаков для которой использовать лаговые признаки (значения признака в предыдущие периоды времени). Кроме того, из признака времени можно выделить признаки дней недели, часов и т.д. Модель приобретает больший смысл, если кроме самих значений временного ряда у нас есть и другие признаки.

МАШИННОЕ ОБУЧЕНИЕ: ПРИЗНАКИ

- 1) **Лаги** – значение временного ряда 1, 2, ... периодов назад
- 2) **Агрегированные признаки** по дате (среднее значение таргета для каждого дня недели, часа и т.д.)
- 3) Другие характеристики

FACEBOOK PROPHET

- Библиотека Prophet от компании Facebook для задач предсказания временных рядов.



<https://github.com/facebook/prophet>

ЛИТЕРАТУРА

- www.machinelearning.ru – всё по временным рядам, в частности:
http://www.machinelearning.ru/wiki/images/archive/e/e7/20150323154210%21Psad_corr.pdf
- <https://www.coursera.org/lecture/data-analysis-applications/arma-fXTrB> и остальные лекции этого курса по теме
- Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003.