

Прогнозирование временного ряда



*Расскажите нам, что ждет нас в будущем,
чтобы мы могли знать, что вы – боги.*

Исайя 41:23

Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Анализ остатков
- Построение прогноза

Прогноз (forecast) vs. предсказание (prediction)



Вычисление/оценка будущего события по систематизированным данным из прошлого



Вычисление/оценка прошлого/будущего события по субъективным соображениям

Постановка задачи

- Дан временной ряд $y_1, \dots, y_T, \dots, y_t \in \mathbb{R}$ (измерения выполнены через равные промежутки)
- Найти функцию прогнозирования $f_T: y_{T+h} \approx f(y_T, \dots, y_1, h) \equiv \hat{y}_{T+h|T}$
отсрочка прогноза $h \in \{1, \dots, H\}$, H – горизонт прогнозирования



Прогноз vs. другие задачи машинного обучения

Данные	Задачи	
	Поиск шаблонов, классификация, кластеризация, поиск аномалий	Прогноз временных рядов
Прецеденты обучающей выборки	значения независимы	будущие значения зависят от прошлых

Предсказательный интервал (Prediction interval)

- Оценка интервала, в который будущее значение попадет с вероятностью не меньше заданной
- Наводнение в Гранд-Форкс, Сев. Дакота, США, апрель 1997 г.: 50000 жителей эвакуировано, 75% зданий повреждено
 - Прогноз высоты паводка: 49 футов (15 м)
 - Построенная защитная дамба: 51 фут (15.5 м)
 - Истинная высота паводка: 54 фута (16.5 м)
 - Точность прогнозов NWS (Нац. метеослужба) на исторических данных: ± 9 футов (2.7 м)
 - Дамба выше на 7 футов (2 м) помогла бы избежать наводнения



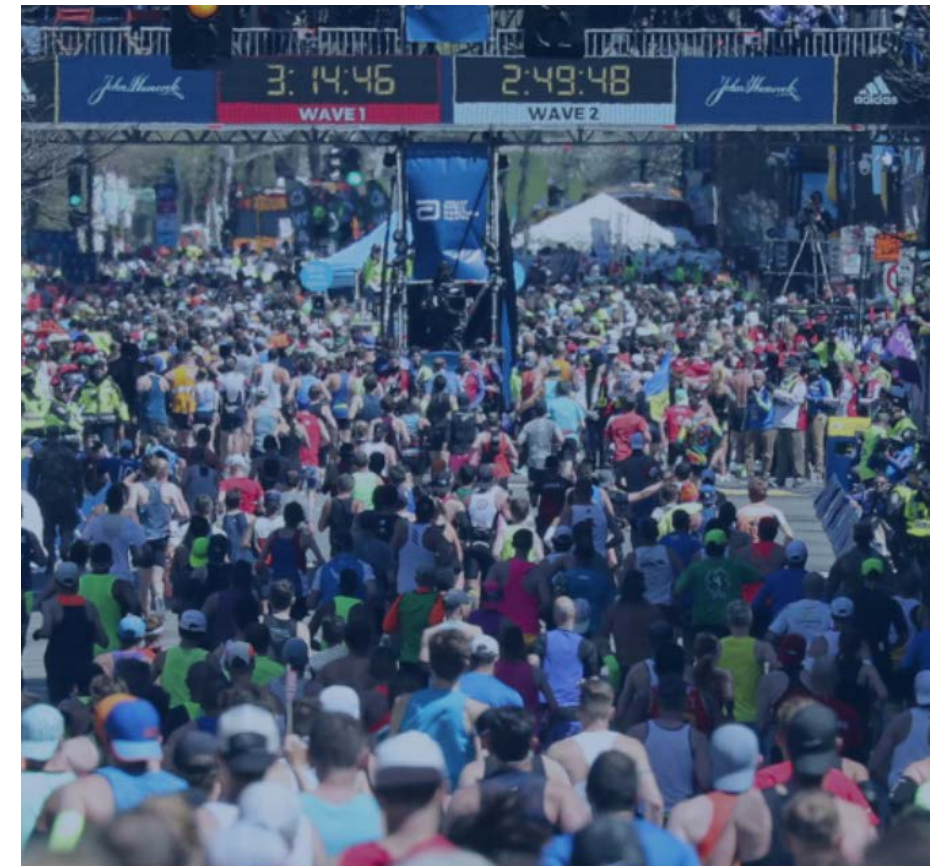
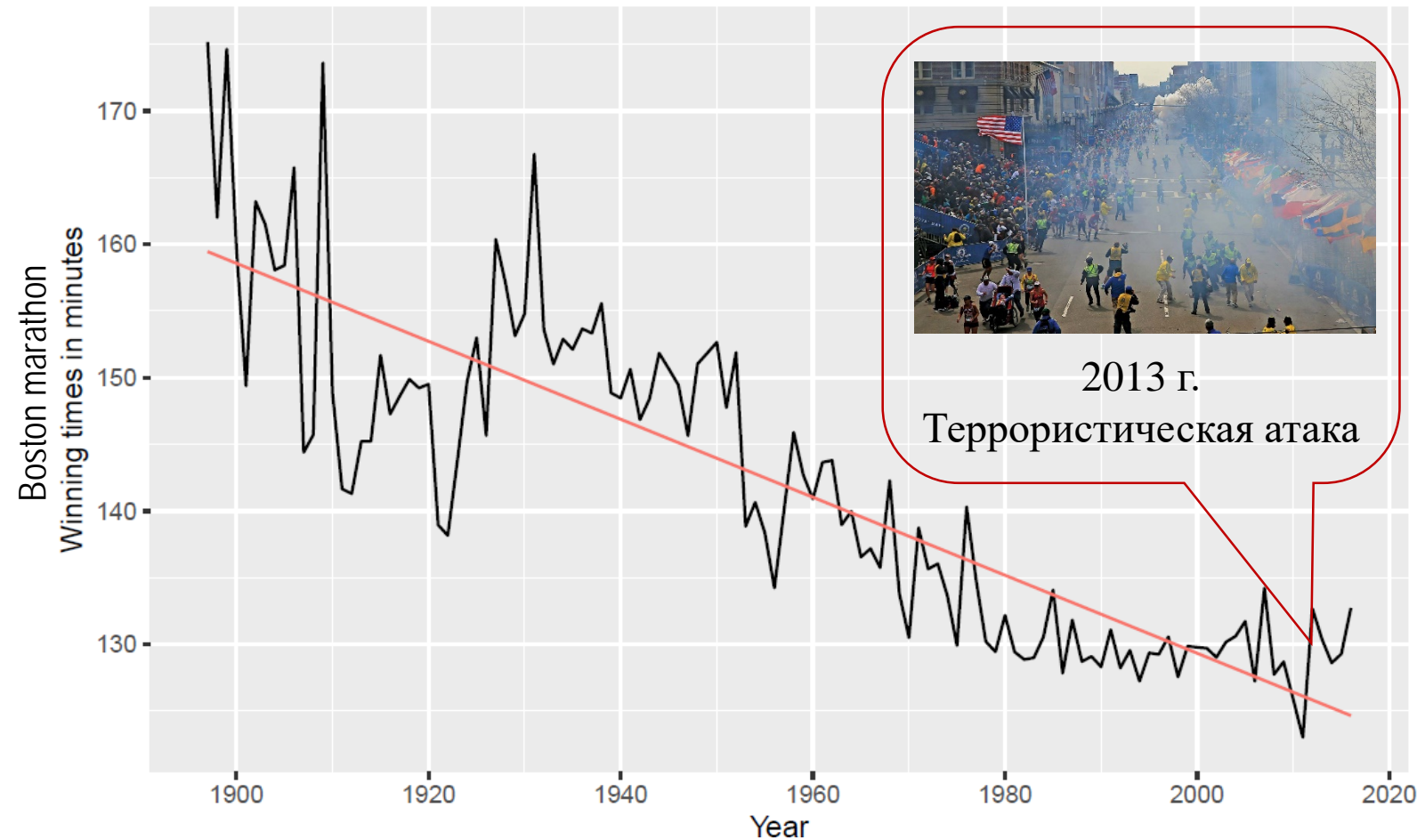
Содержание

- Постановка задачи прогнозирования
- **Компоненты временного ряда**
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Анализ остатков
- Построение прогноза

Компоненты временного ряда

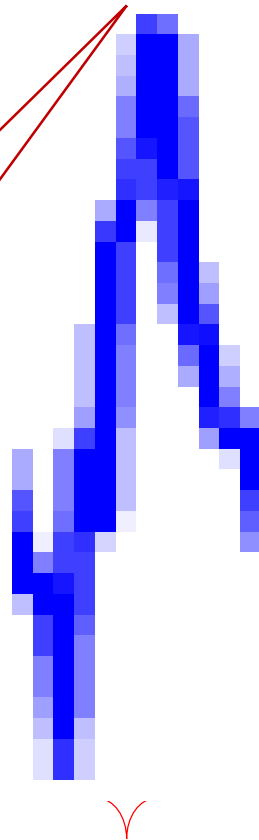
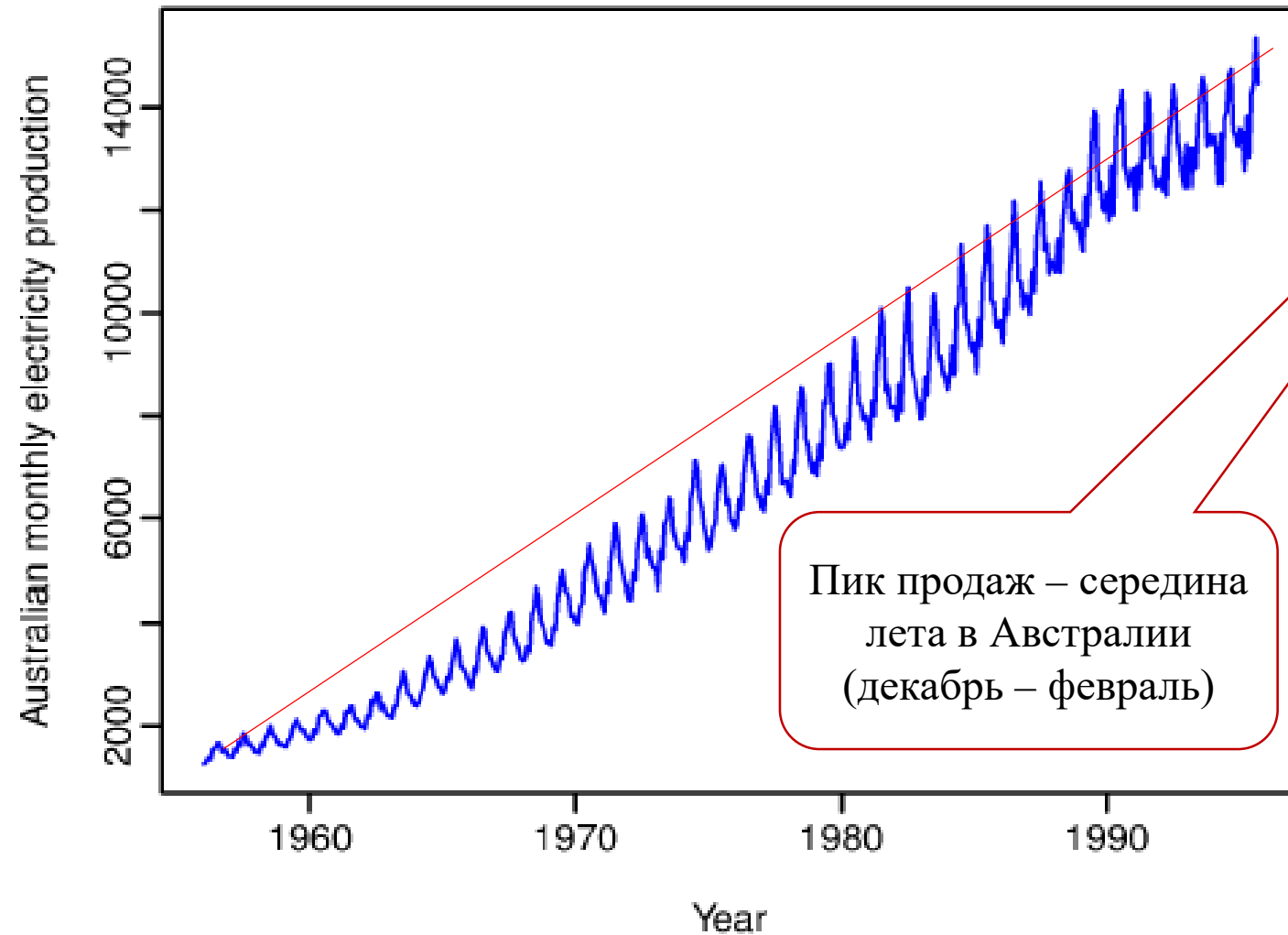
Компонент	Определение	Пример ряда
Тренд	плавное долгосрочное изменение уровня ряда	цена на недвижимость в регионе (повышение)
Сезонность	циклические изменения уровня ряда с постоянным периодом	средняя месячная зарплата на предприятии (годовая сезонность)
Цикл	изменения уровня ряда с переменным периодом	количество солнечных пятен (период солнечной активности – от 5 до 7 лет)
Ошибка	непрогнозируемая случайная компонента ряда	выбросы, аномалии

Пример: тренд без сезонности и цикла



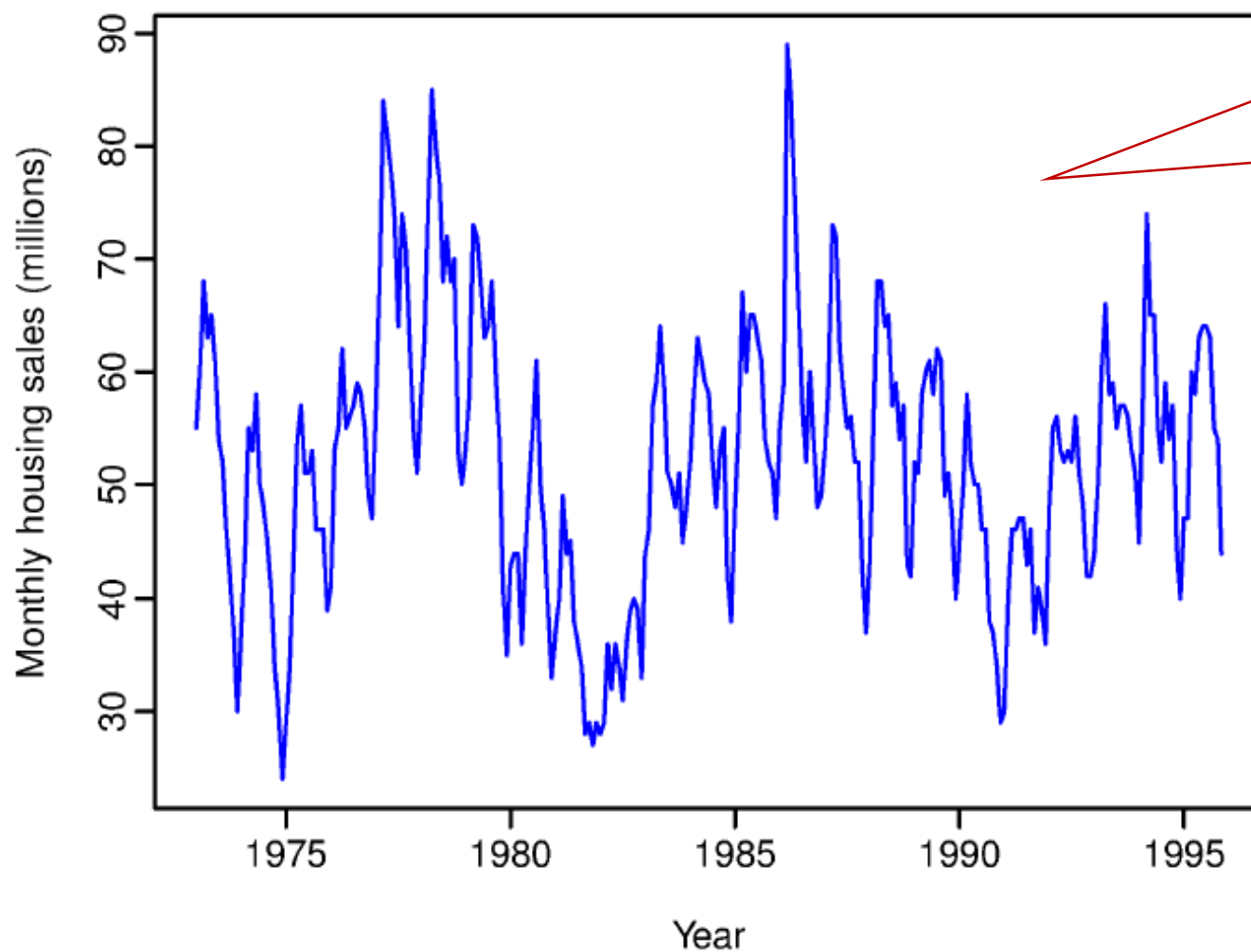
* The Boston Marathon Results. URL: <https://www.baa.org/sites/default/files/2019-07/BostonMarathonHistoricalResults.pdf>

Пример: тренд и сезонность



12 месяцев

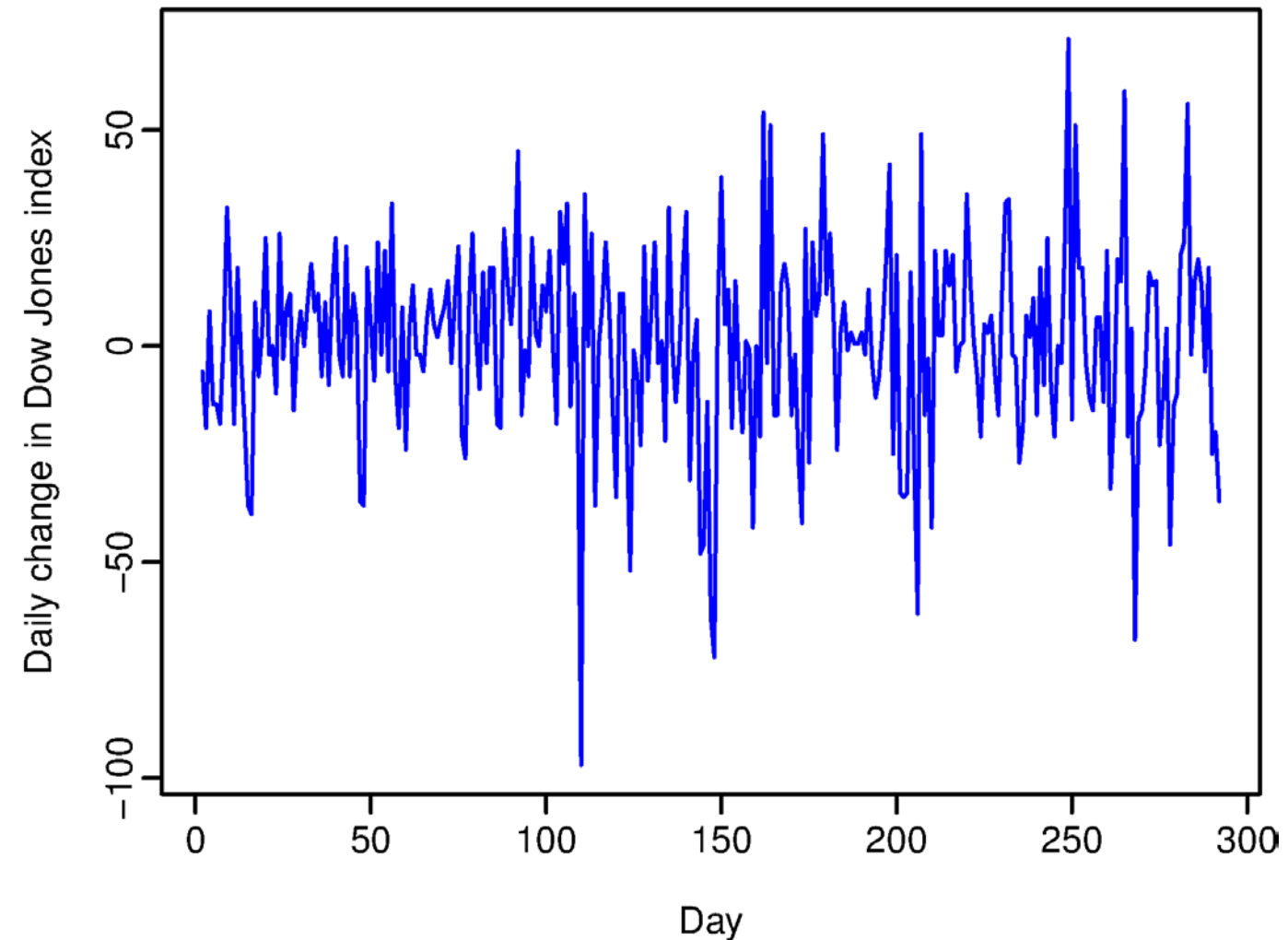
Пример: сезонность, циклы



Годовая сезонность
(ежегодные пики),
недетерминированные
экономические спады



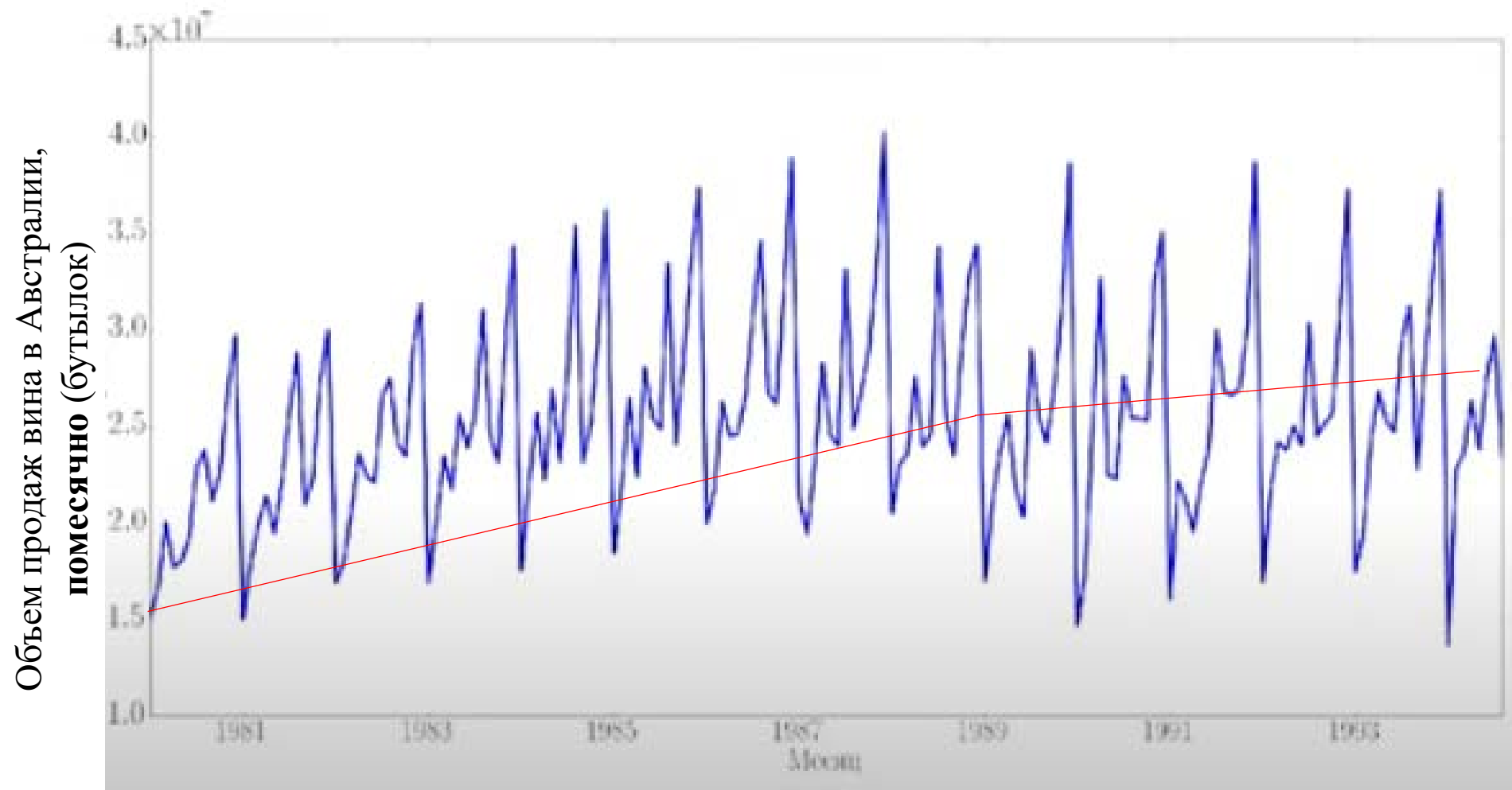
Пример: отсутствие тренда, сезонности, циклов



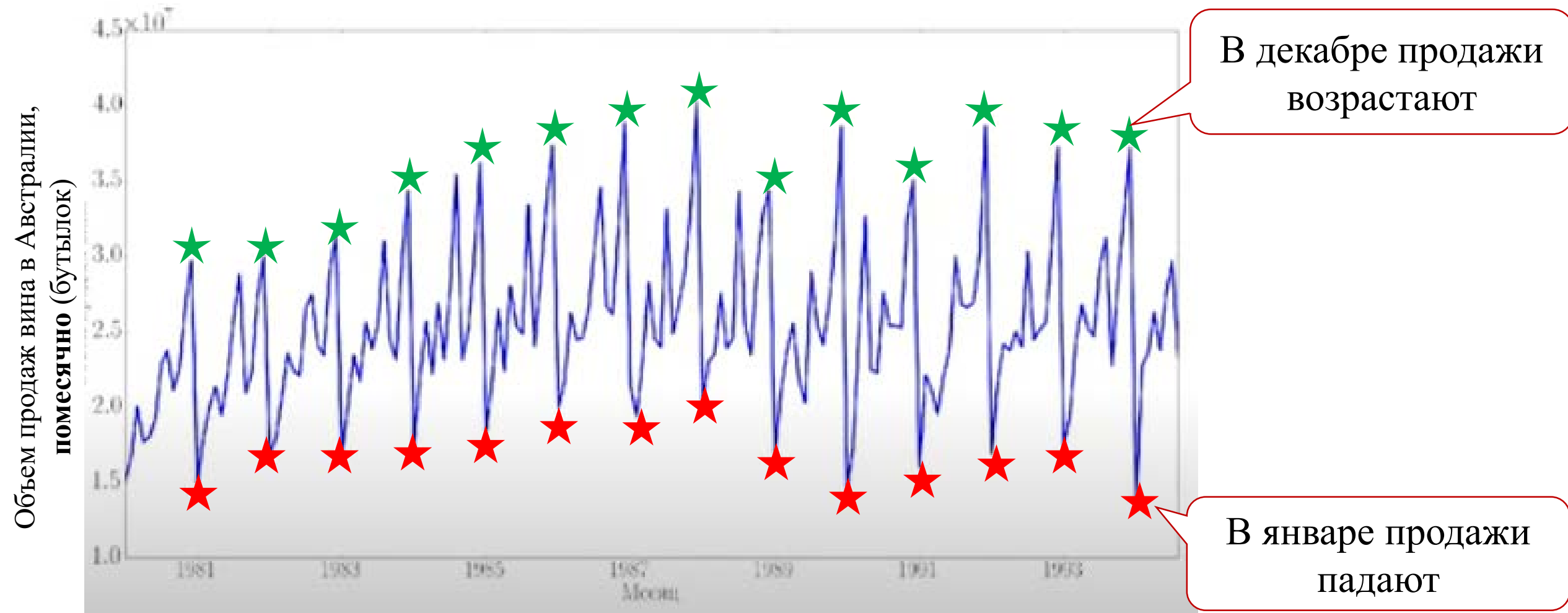
Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- **Автокорреляция временного ряда**
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Анализ остатков
- Построение прогноза

Автокорреляция временного ряда

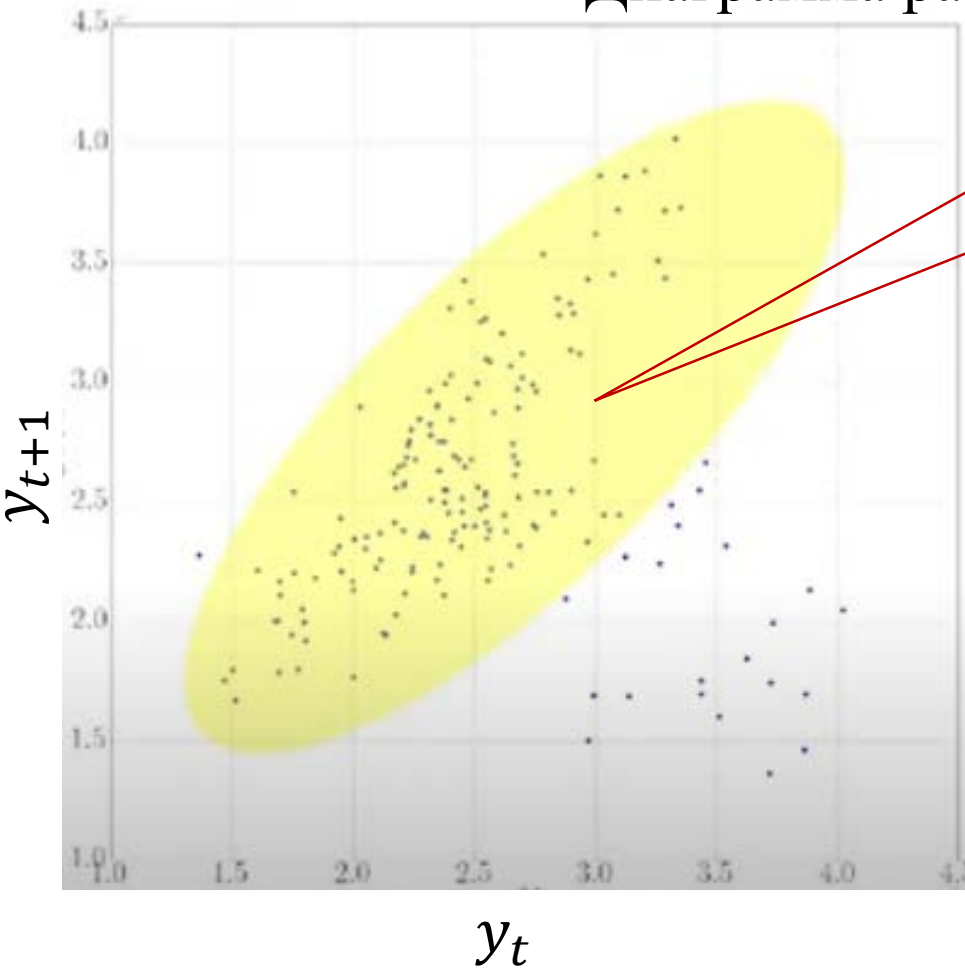


Автокорреляция временного ряда



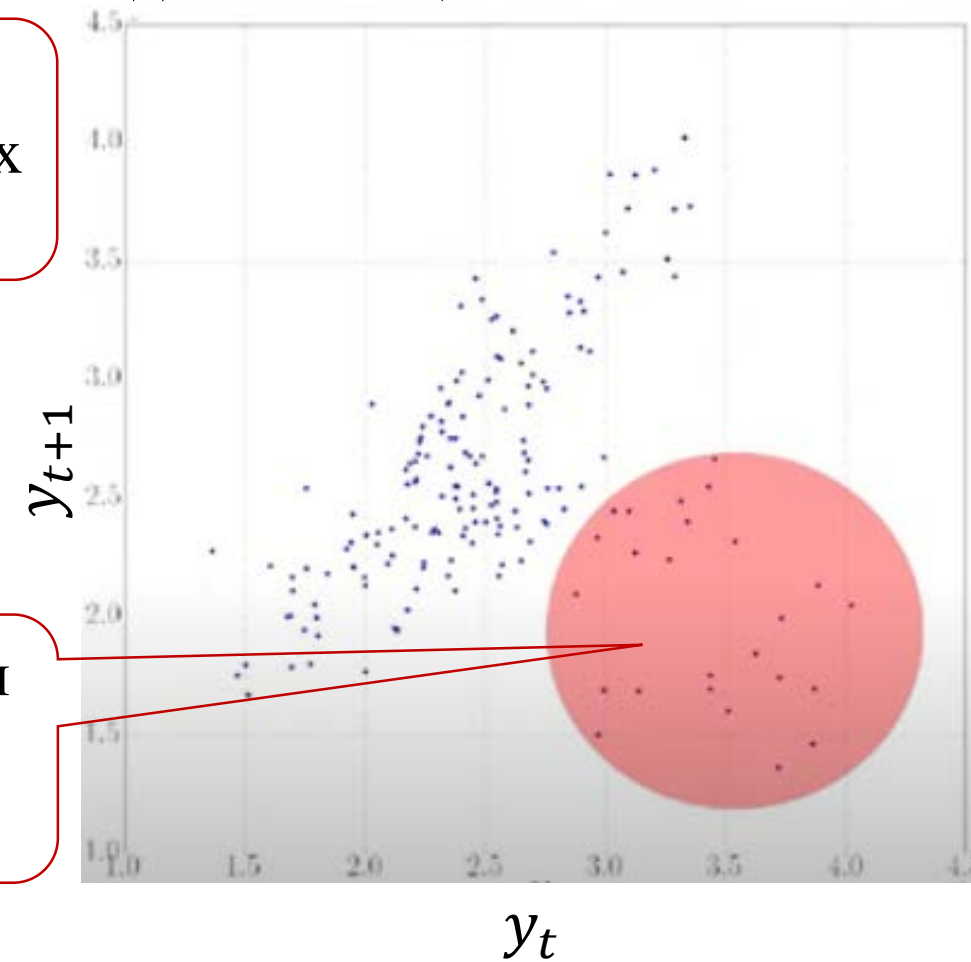
Автокорреляция временного ряда

Диаграмма рассеяния для продаж в соседних месяцах



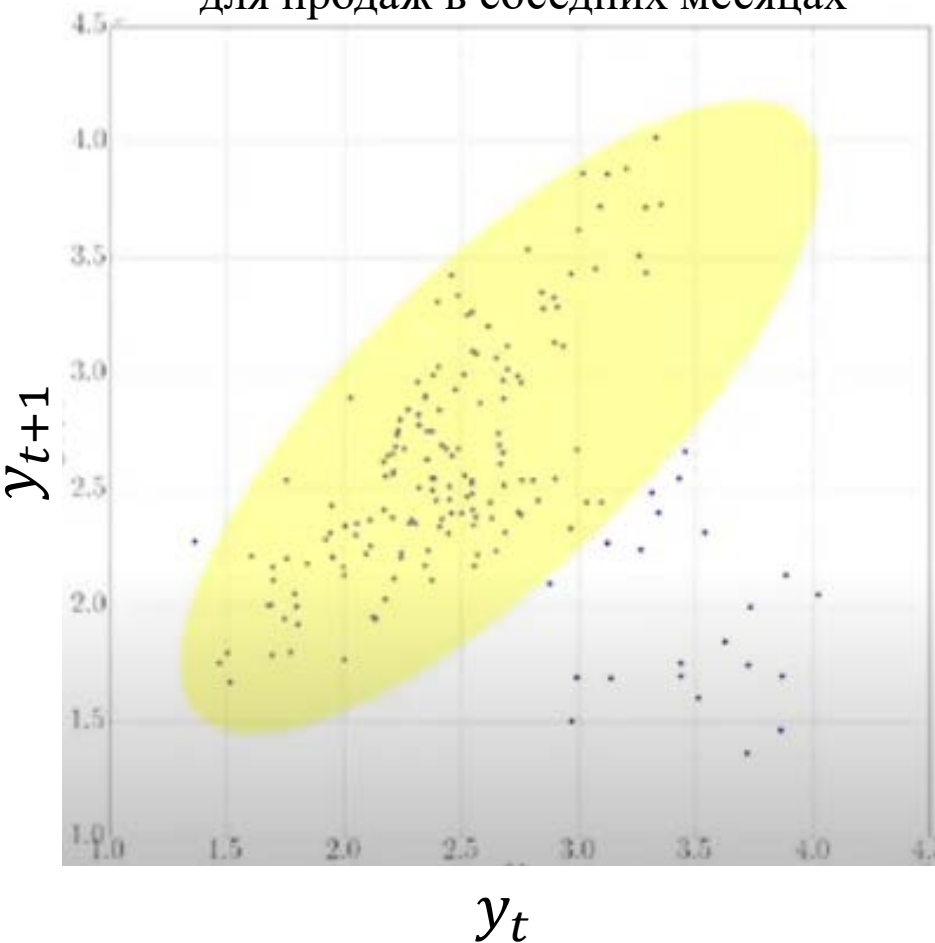
В феврале-ноябре
продажи в соседних
месяцах похожи

В декабре продажи
возрастают,
в январе – падают



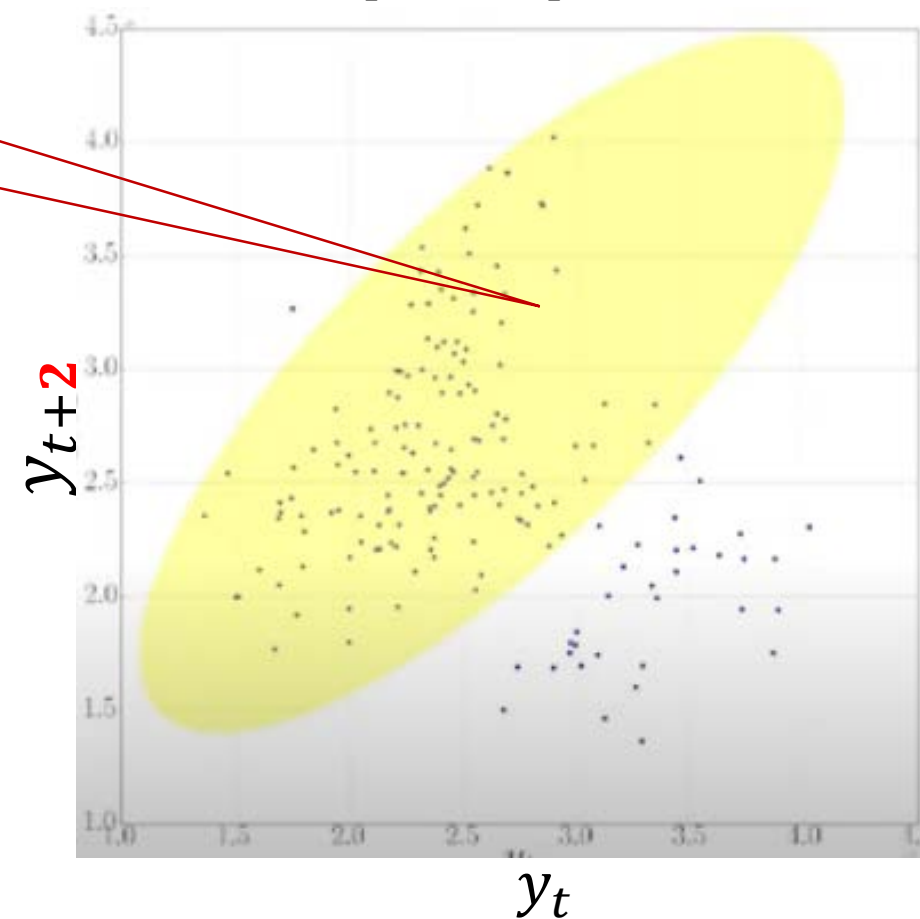
Автокорреляция временного ряда

Диаграмма рассеяния
для продаж в соседних месяцах



Кластер
укрупняется

Диаграмма рассеяния
для продаж через месяц



Автокорреляция временного ряда

Диаграмма рассеяния
для продаж в соседних месяцах

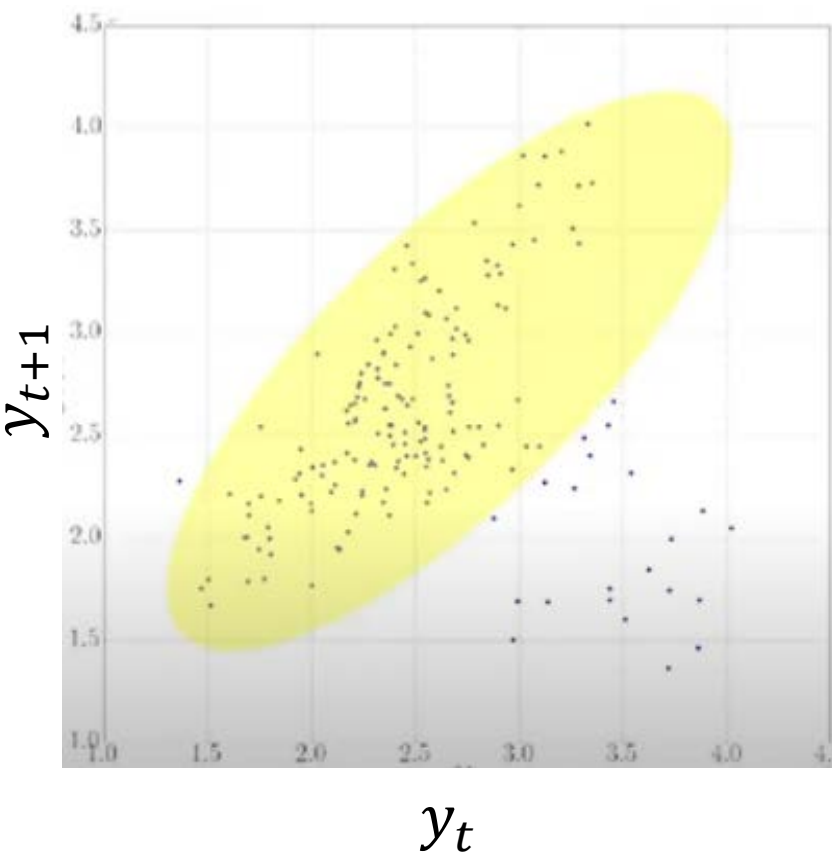


Диаграмма рассеяния
для продаж через месяц

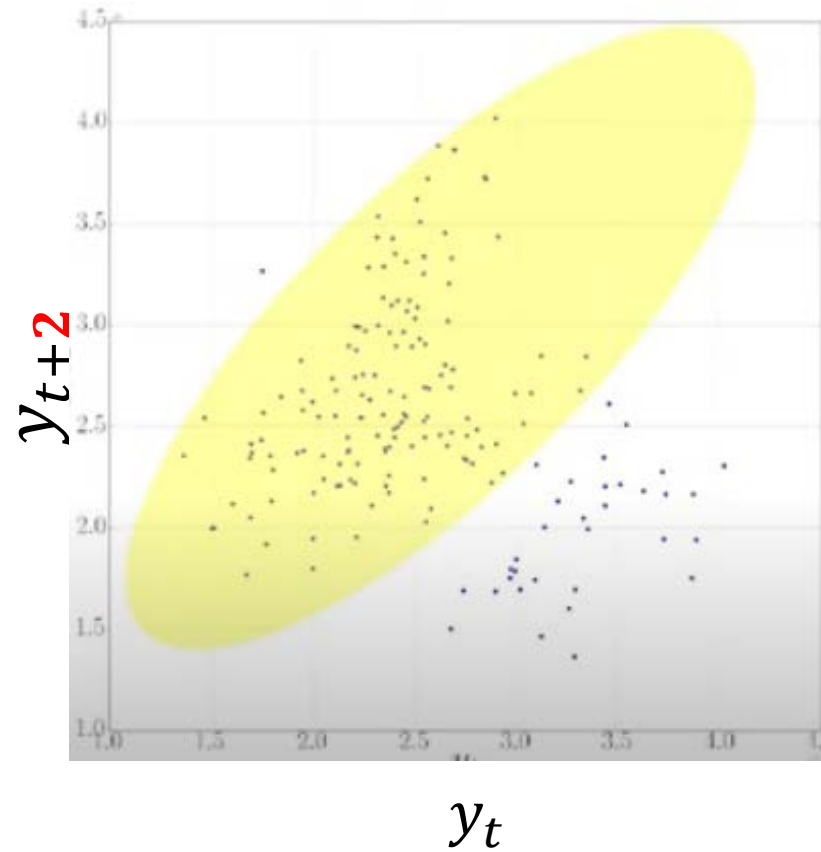
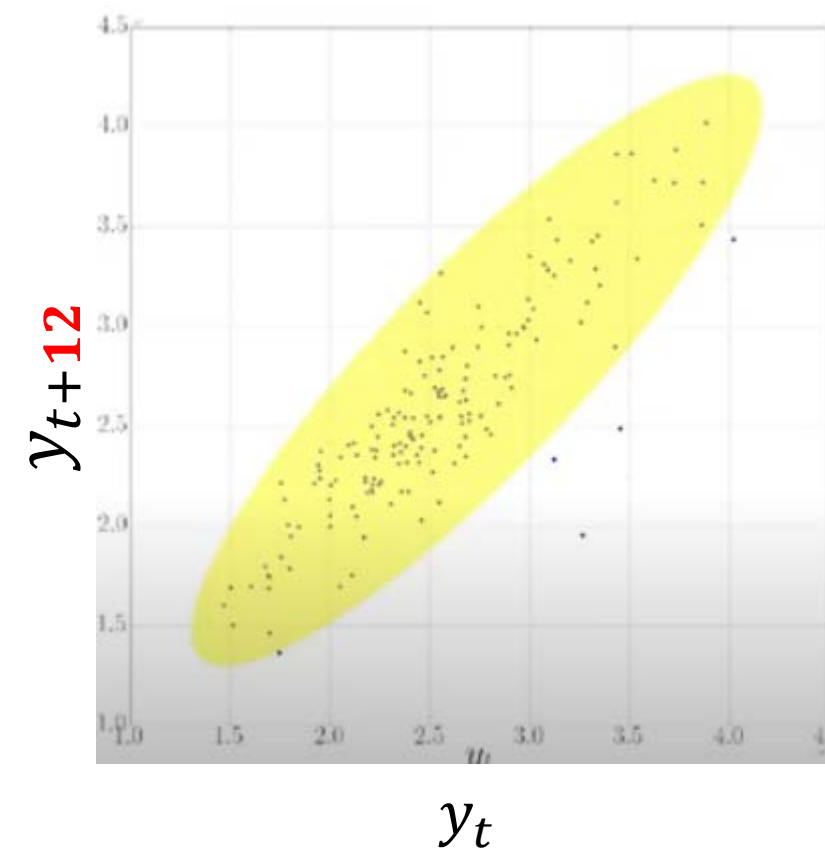
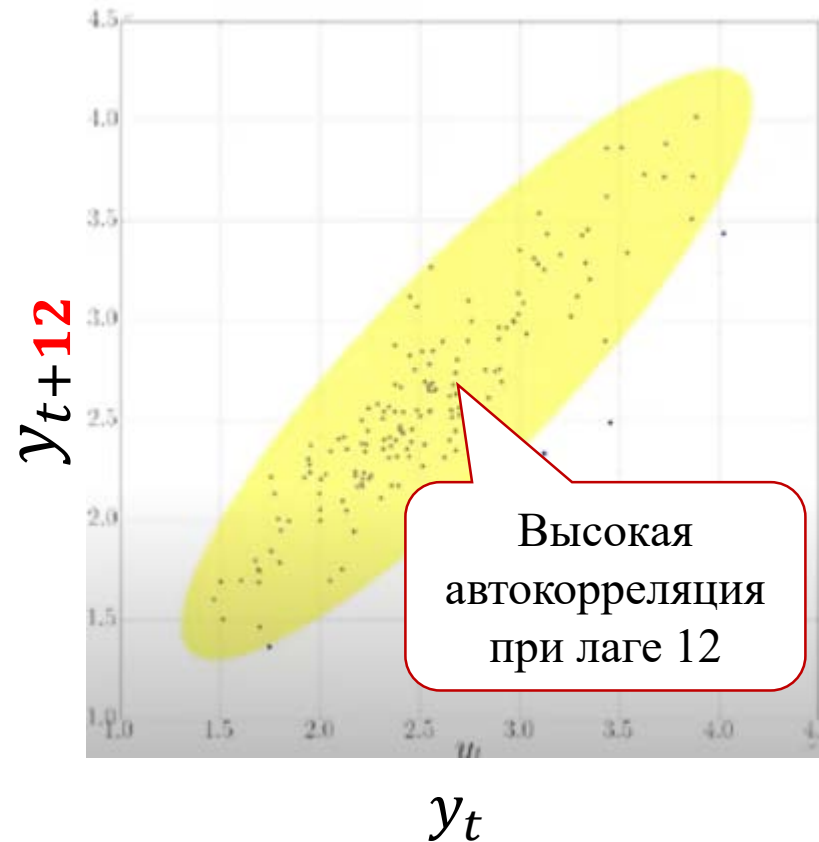


Диаграмма рассеяния
для продаж через год (12 мес.)



Автокорреляция временного ряда

Диаграмма рассеяния
для продаж через год (12 мес.)



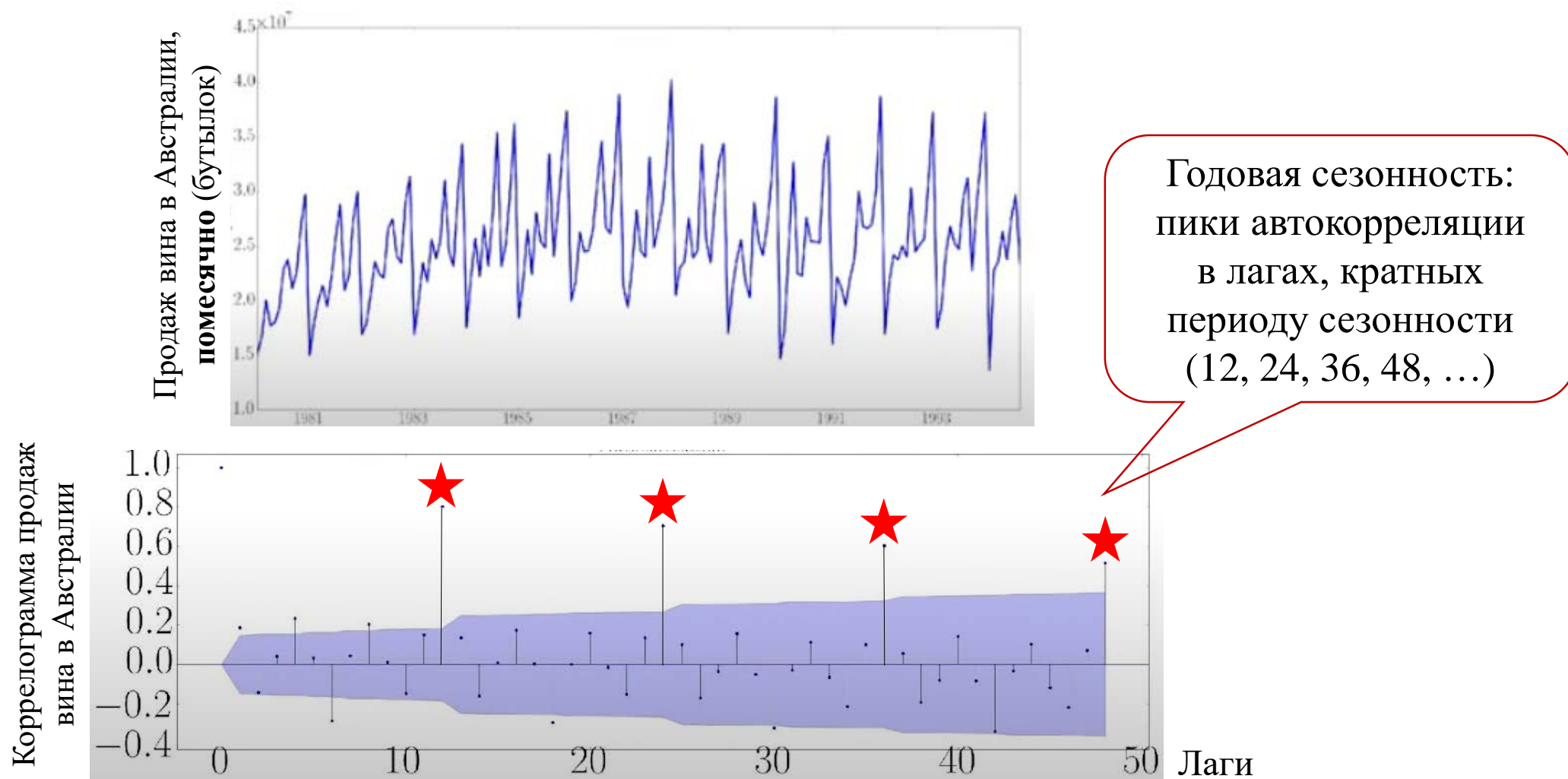
- Функция автокорреляции вычисляет корреляцию Пирсона между рядом и его копией со значениями, отстоящими на заданный лаг:
$$ACF(lag) = corr(y_t, y_{t+lag})$$
- Функция частичной (частной) автокорреляции вычисляет корреляцию Пирсона между рядом и его копией со значениями, отстоящими на заданный лаг, дополнительно удаляя линейную зависимость между сдвинутыми рядами

Вычисление автокорреляции

• $Y = [2, 7, 3, 1, 0, 6, 8, 9, 5, 10]$

lag	Y_t	Y_{t+lag}	$ACF(lag)$
0	[2, 7, 3, 1, 0, 6, 8, 9, 5, 10]	[2, 7, 3, 1, 0, 6, 8, 9, 5, 10]	1
1	[2, 7, 3, 1, 0, 6, 8, 9, 5, 10]	[2 , 7, 3, 1, 0, 6, 8, 9, 5, 10]	0.331354
2	[2, 7, 3, 1, 0, 6, 8, 9, 5 , 10]	[2 , 7 , 3, 1, 0, 6, 8, 9, 5, 10]	0.158205
3	[2, 7, 3, 1, 0, 6, 8, 9 , 5 , 10]	[2 , 7 , 3 , 1, 0, 6, 8, 9, 5, 10]	-0.171137

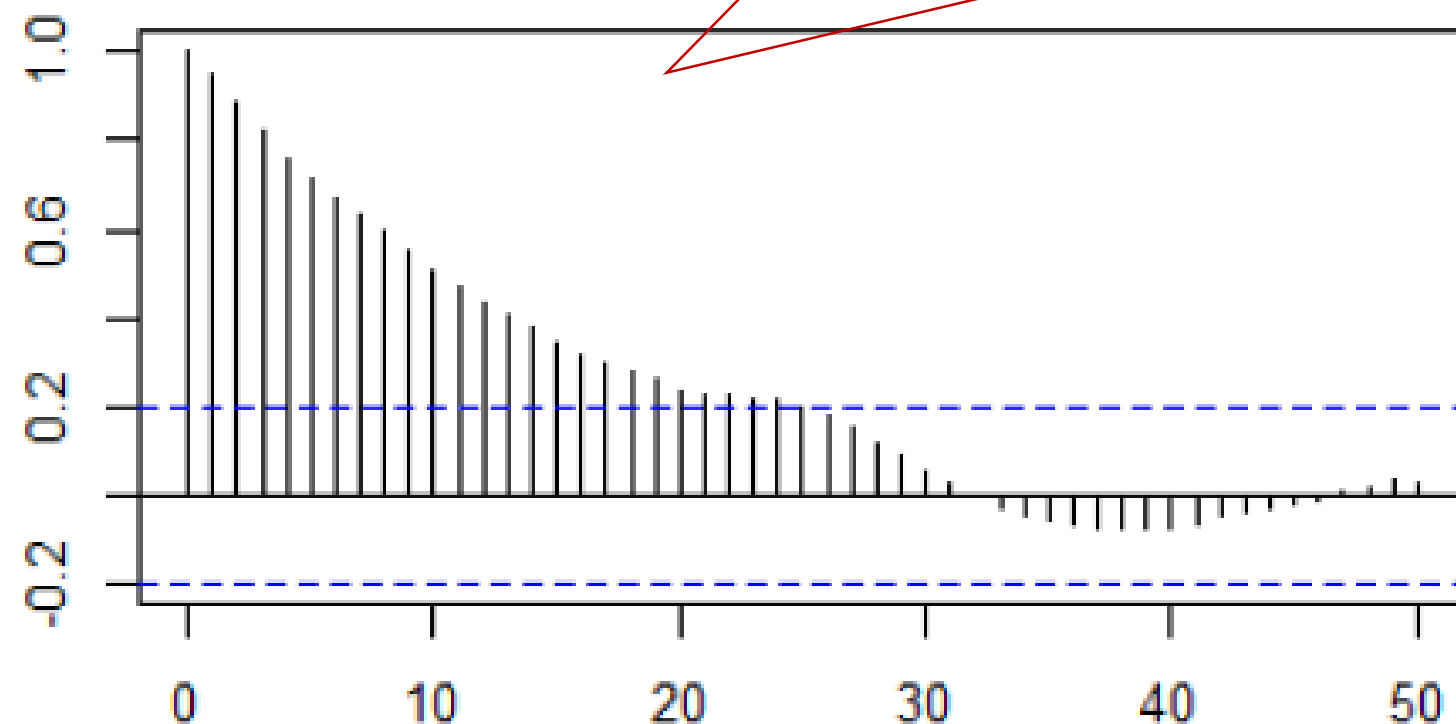
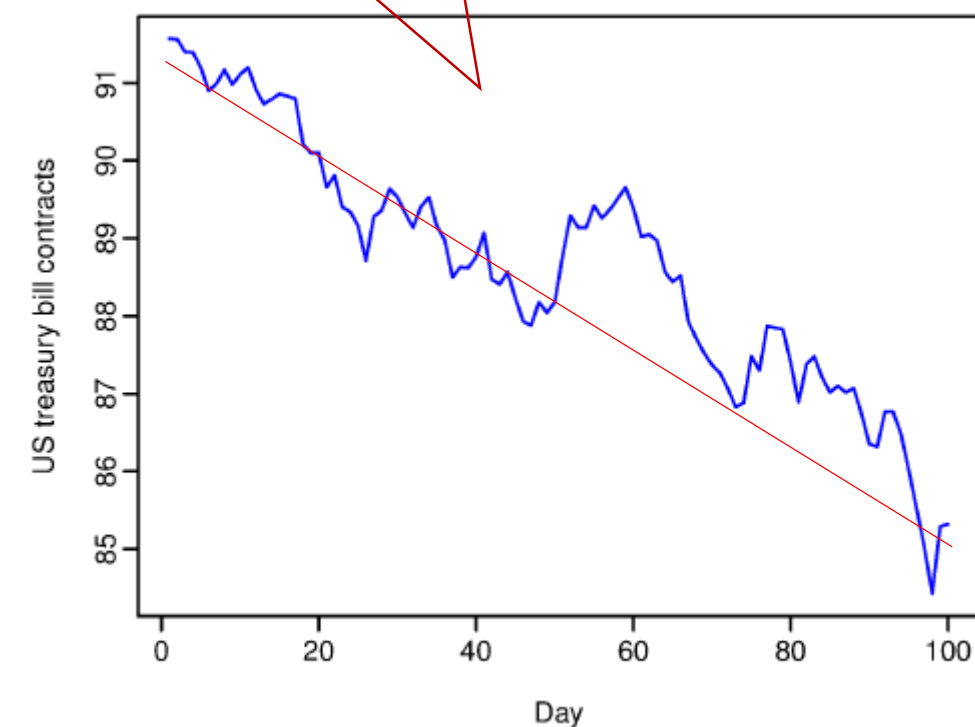
Коррелограмма (график функции автокорреляции)



Коррелограмма: сильный тренд

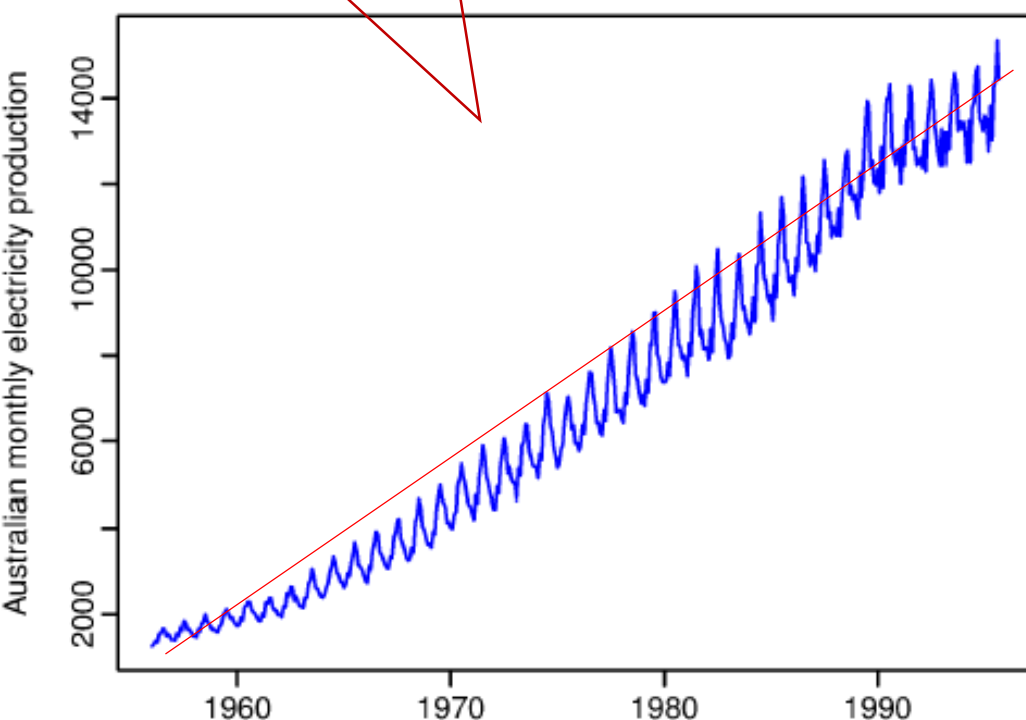
Ряд с сильным трендом

Типичная структура автокорреляции для ряда с сильным трендом: большие значения при малых лагах, убывание, колебания вокруг нуля

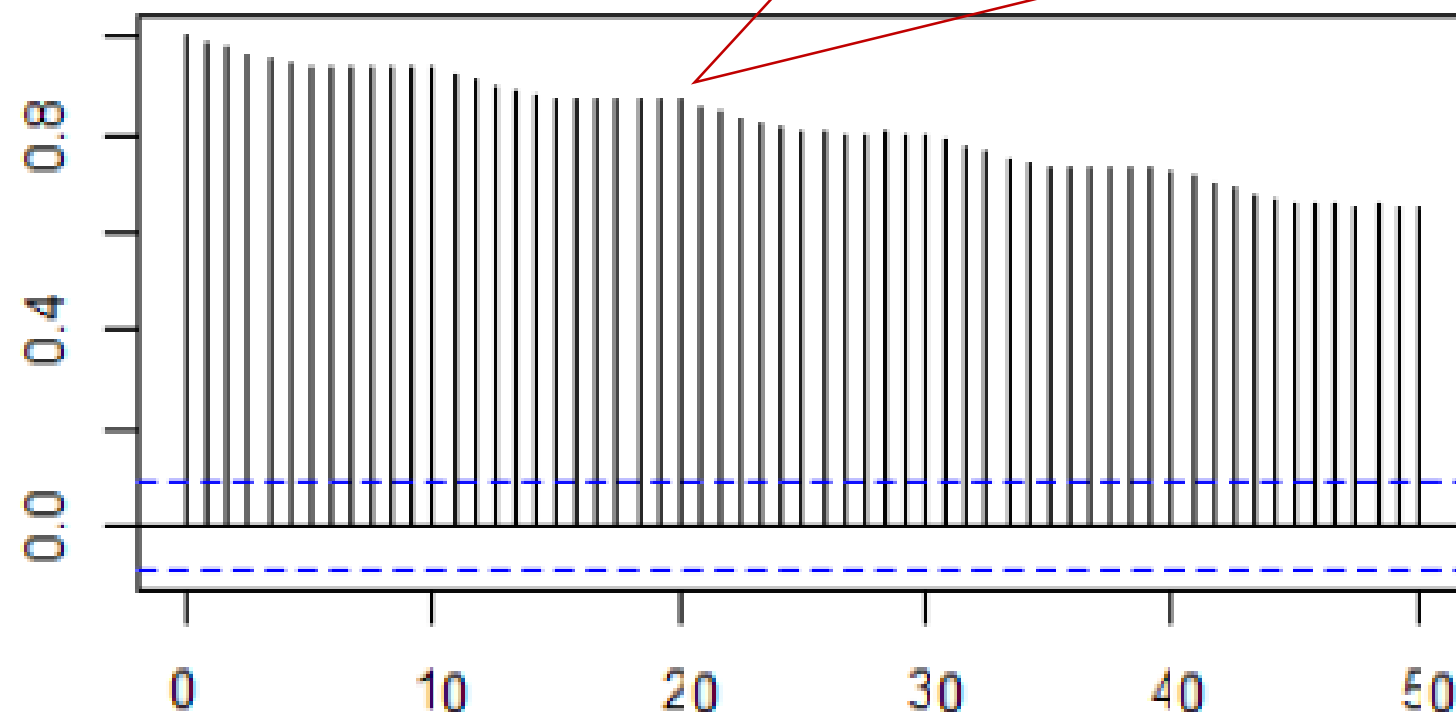


Коррелограмма: сильный тренд и сезонность

Ряд с сильным трендом
и сезонностью

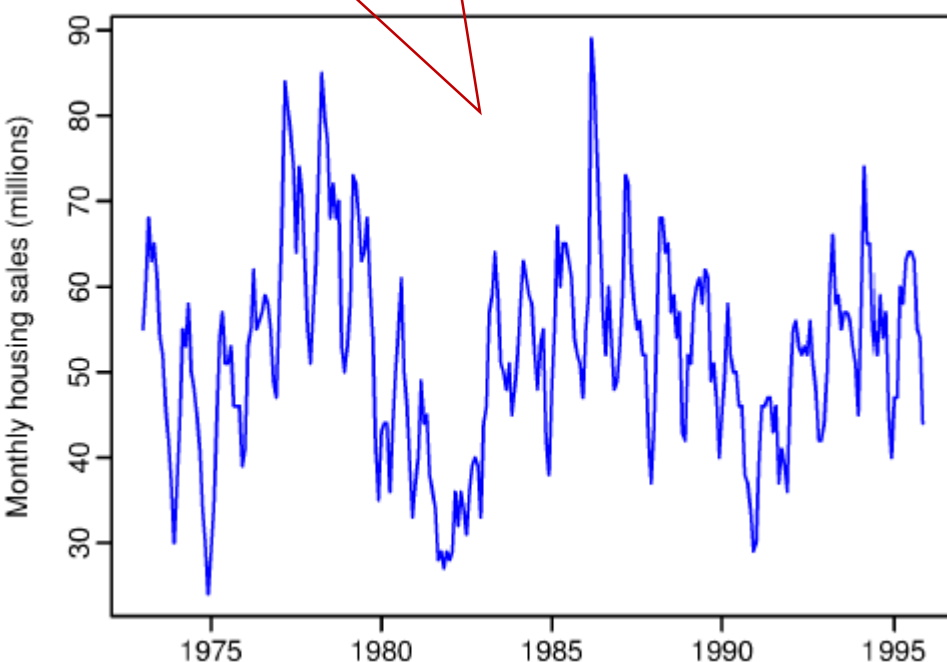


Типичная структура автокорреляции для ряда с сильным трендом
и сезонностью: слабо различимые пики автокорреляции
в лагах, кратных периоду сезонности (12, 24, 36, 48, ...)

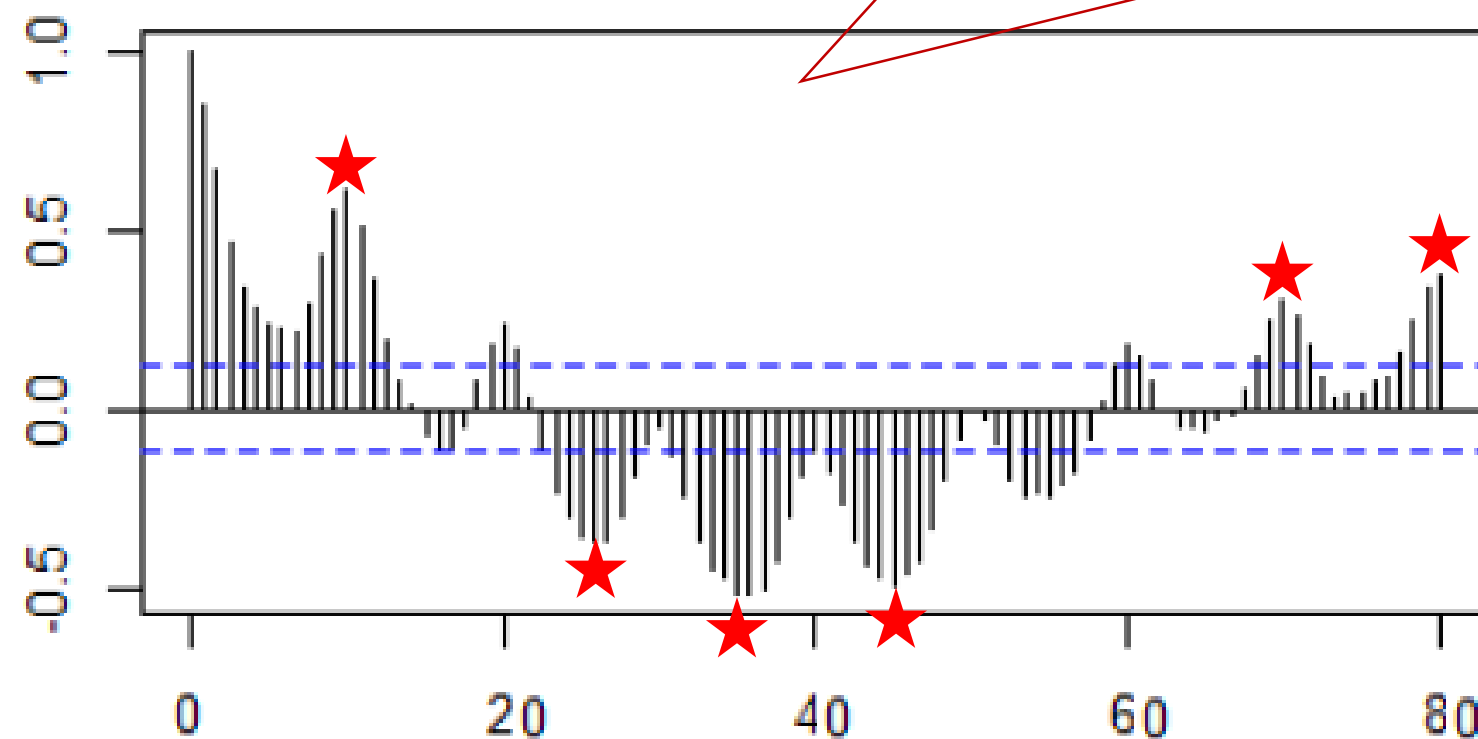


Коррелограмма: сезонность и циклы

Ряд с сезонностью
и циклами

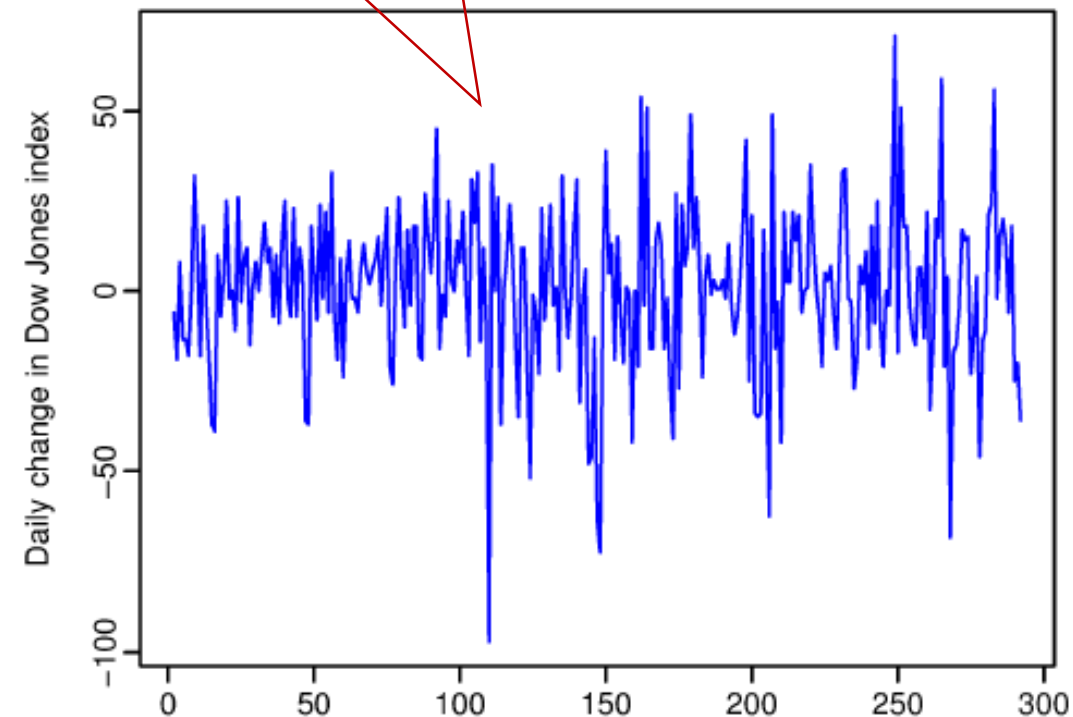


Типичная структура автокорреляции для ряда с сезонностью и циклами: циклы приводят к сдвигу пиков автокорреляции в лагах, кратных периоду сезонности, в иные лаги

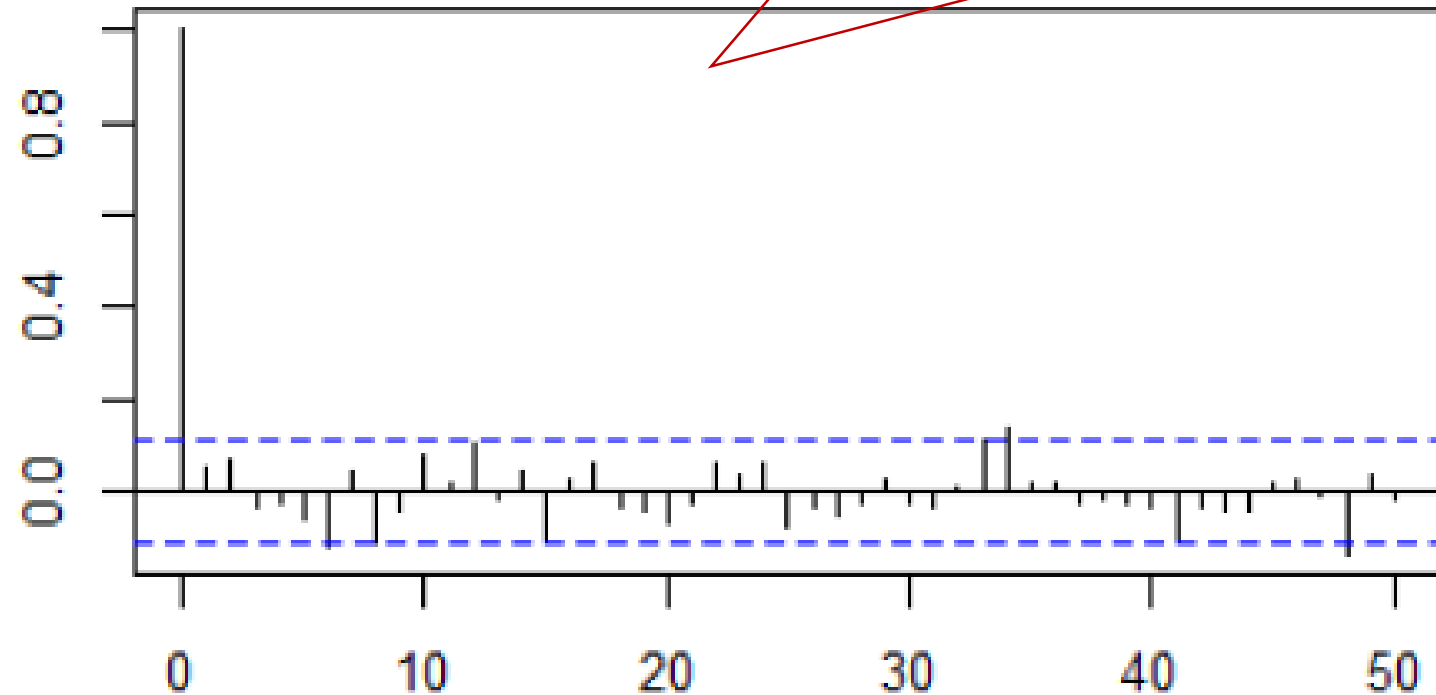


Коррелограмма: отсутствие тренда, сезонности и циклов

Ряд без тренда,
сезонности и циклов



Типичная структура автокорреляции для ряда без тренда,
сезонности и циклов: значения слабо колеблются вокруг нуля



Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- **Стационарность и стабилизирующие преобразования ряда**
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Построение прогноза
- Анализ остатков

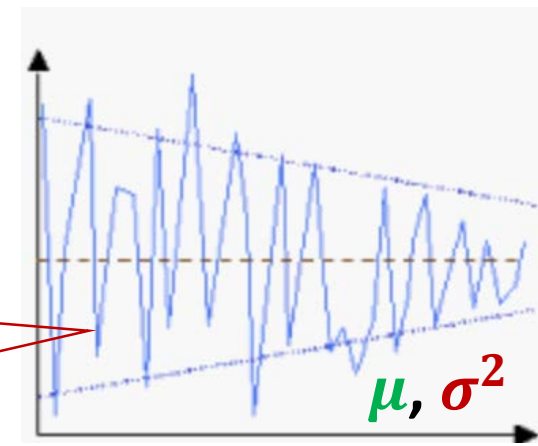
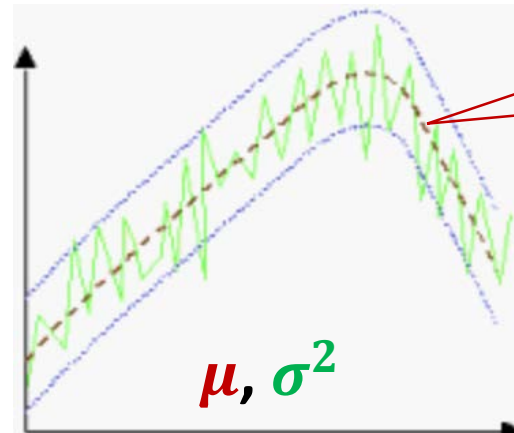
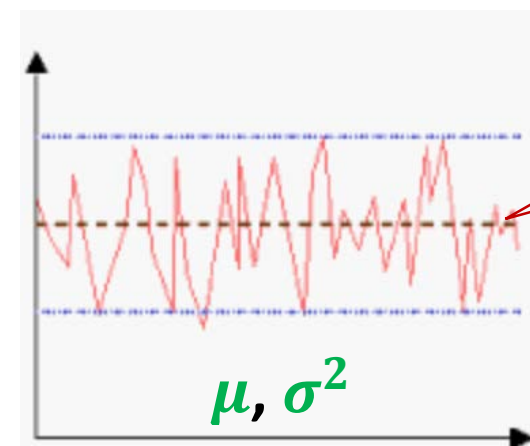
Стационарный (stationary) ряд: стабильность характеристик

- Среднее арифметическое
$$\mu(T_{i,m}) = \frac{1}{m} \sum_{k=1}^m t_k$$

- Дисперсия
$$\sigma^2(T_{i,m}) = \frac{1}{m} \sum_{k=1}^m (t_k - \mu)^2$$

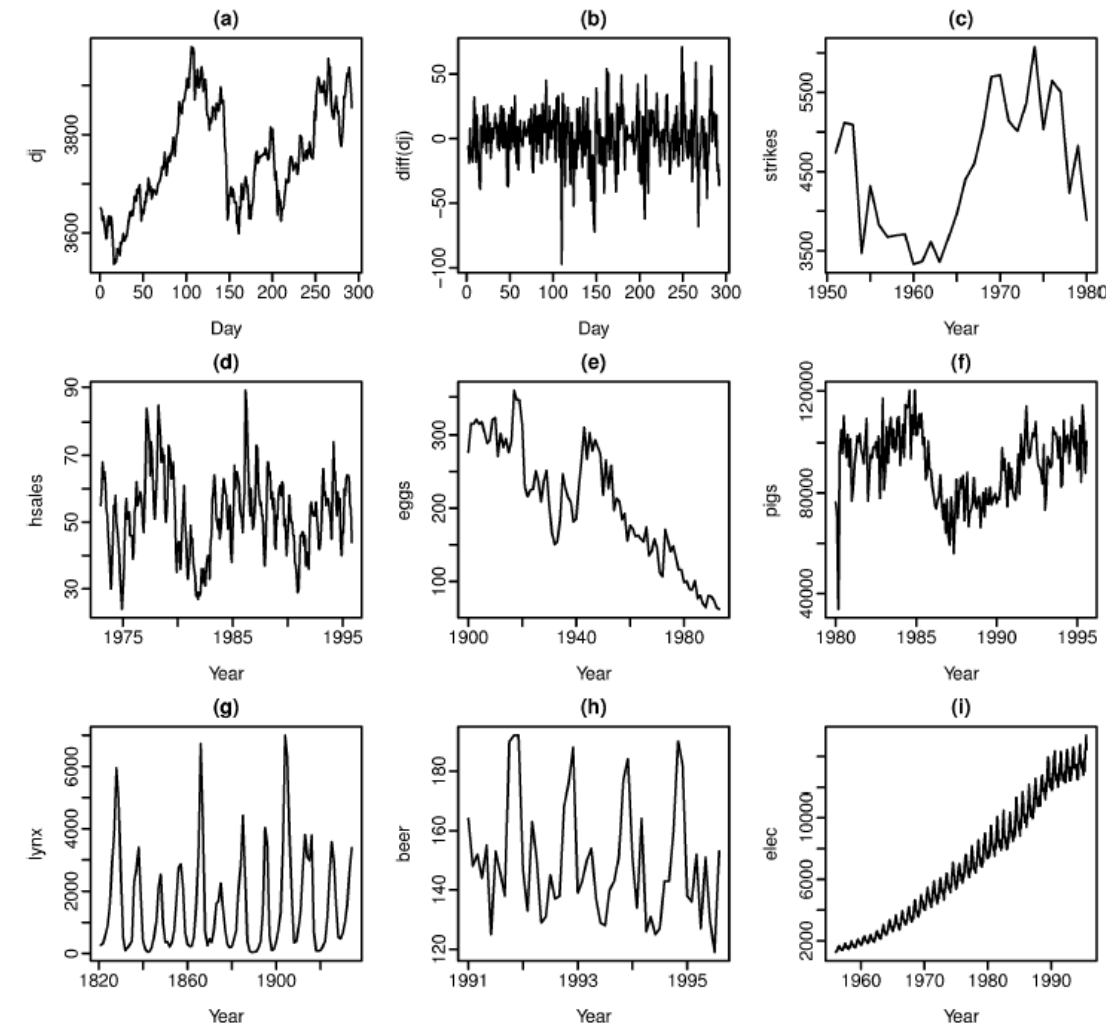
- Ковариация
$$\text{cov}(T_{i,m}, T_{j,m}) = \frac{1}{n} \sum_{k=1}^m (t_{i+k-1} - \mu(T_{i,m}))(t_{j+k-1} - \mu(T_{j,m}))$$

Ряд y_1, \dots, y_T стационарен, если
 $\forall s$ распределение y_t, \dots, y_{t+s}
не зависит от времени t

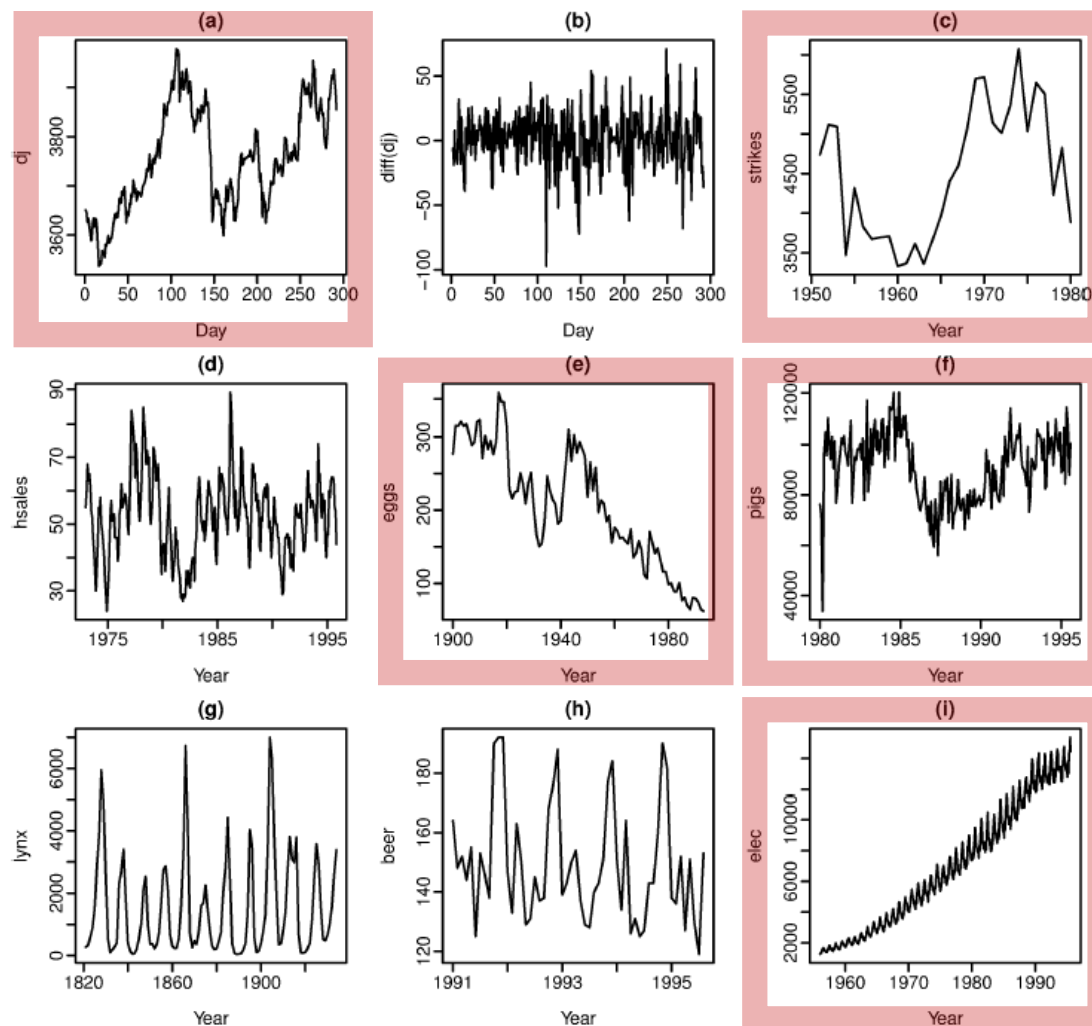


Влияние компонентов ряда на стационарность

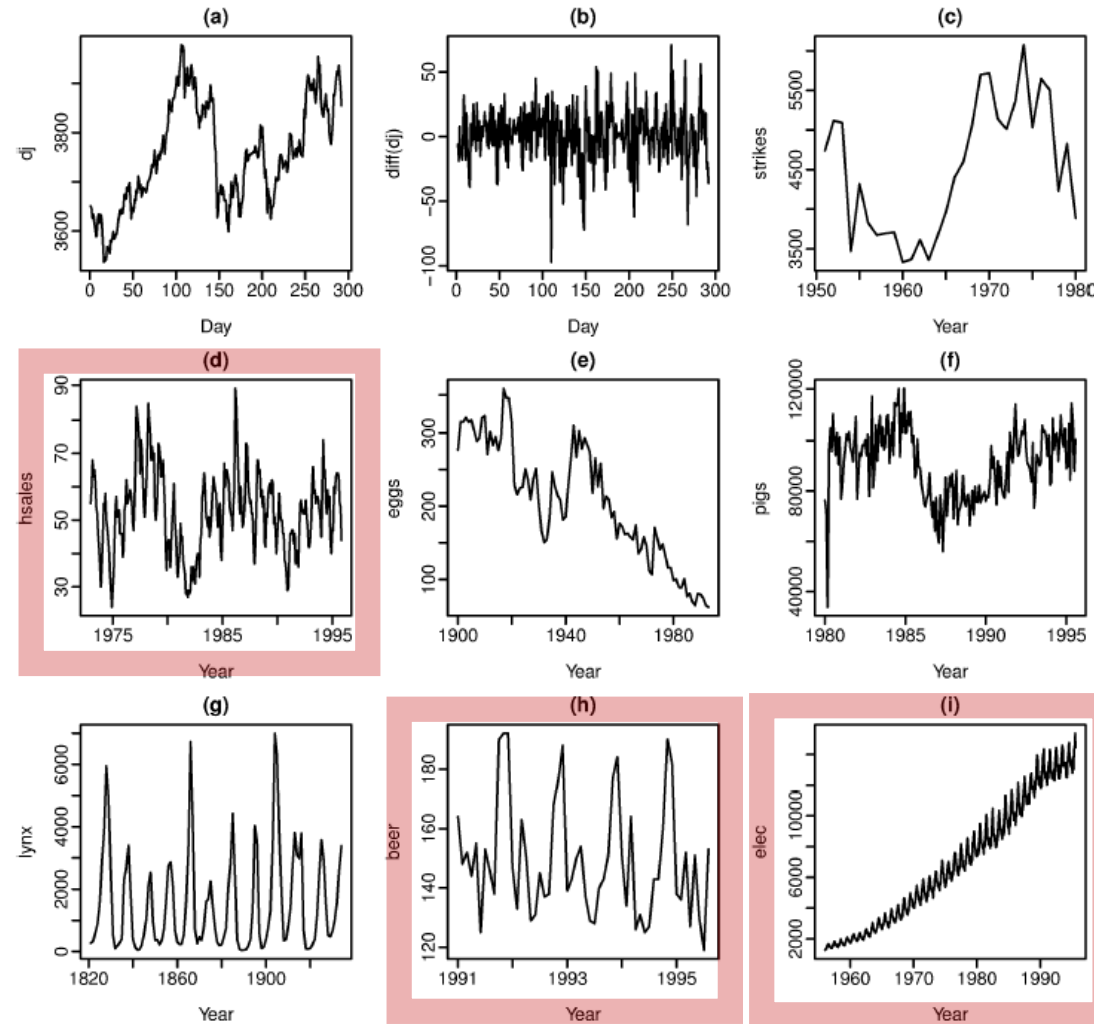
- Ряды с трендом не стационарны
 - распределения в окнах в начале и в конце ряда существенно отличаются
- Ряды с сезонностью не стационарны
 - распределения в окнах длины менее сезона в периоды минимумов и максимумов существенно отличаются
- Ряды с непериодическими циклами стационарны
 - нельзя предсказать заранее локацию минимумов и максимумов



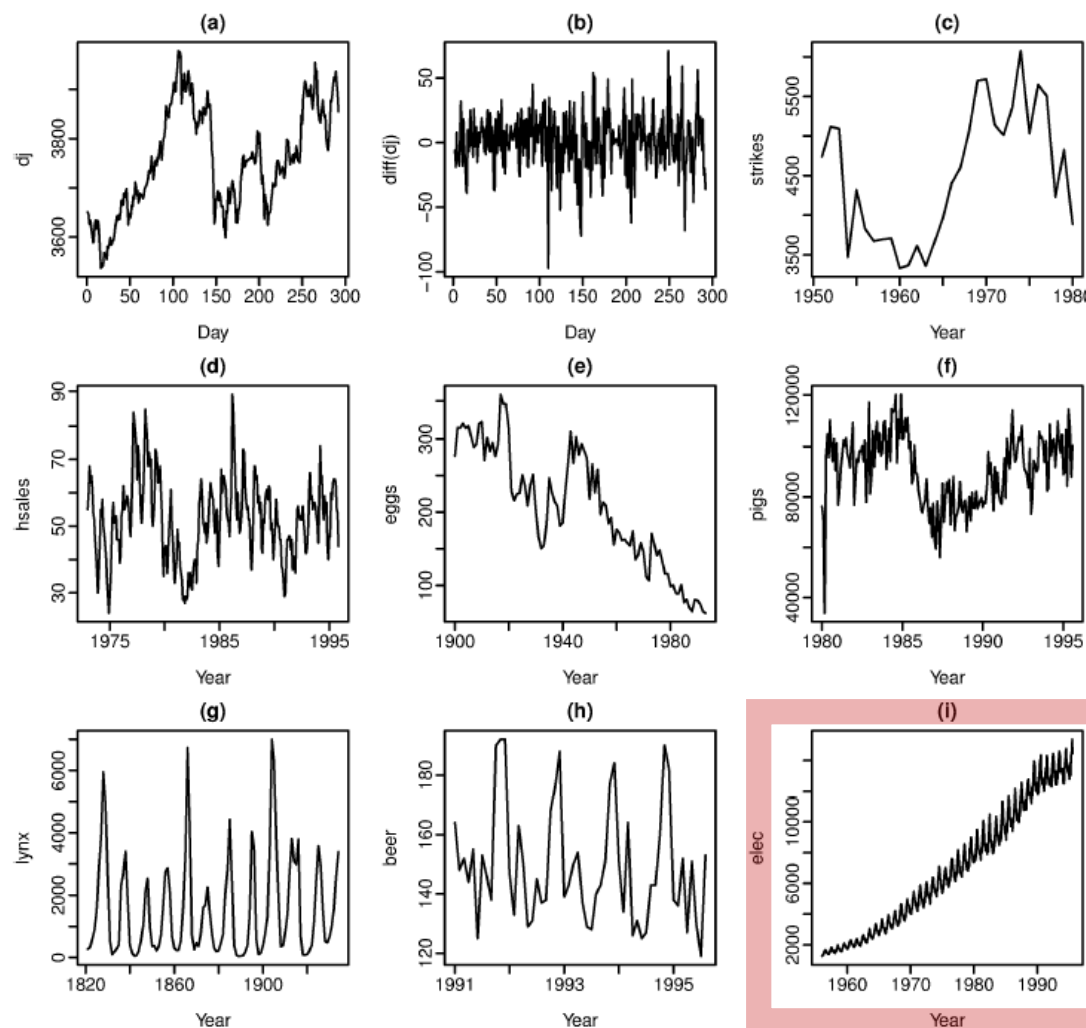
Ряды с трендом не стационарны



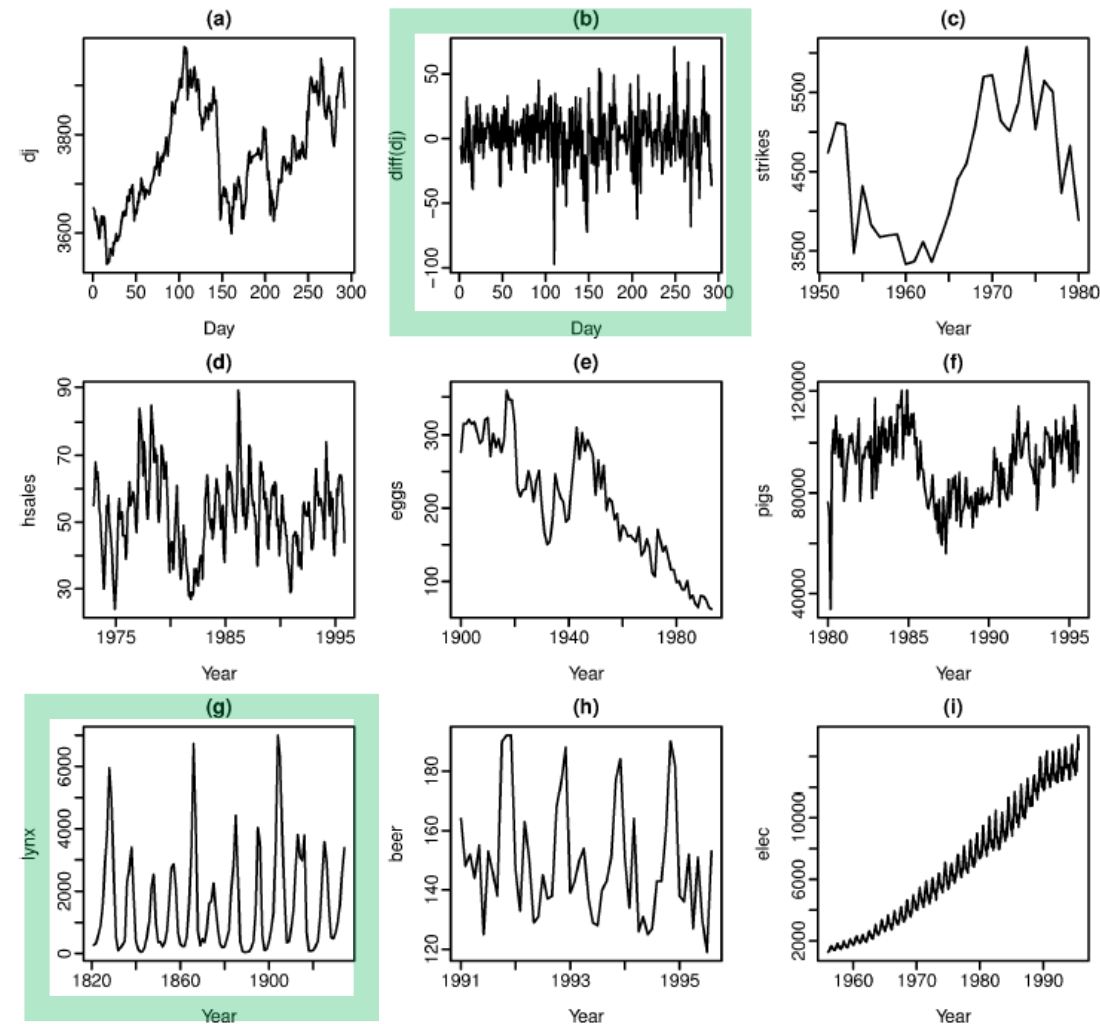
Ряды с сезонностью не стационарны



Ряды с меняющейся дисперсией не стационарны



Стационарные ряды



Проверка ряда на стационарность

- Визуальный анализ, проверка постоянства среднего и дисперсии
- Автокорреляционная функция (ACF, Autocorrelation function)
- Статистические тесты
 - Тест Дики–Фуллера (Dickey–Fuller test)
 - Тест KPSS (Kwiatkowski–Phillips–Schmidt–Shin test)

Тест Дикки–Фуллера (Dickey–Fuller test)

- Ряд y_1, \dots, y_T
- Гипотеза H_0 : ряд не стационарен
- Гипотеза H_1 : ряд стационарен
- Статистика: нет
- При H_0 $DF(y)$ имеет табличное распределение



Дэвид Дикки
(David A. Dickey)
р. 1945



Уэйн Фуллер
(Wayne A. Fuller)
р. 1931

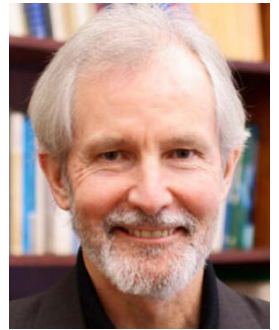
Тест KPSS (Kwiatkowski–Phillips–Schmidt–Shin test)

- Ряд y_1, \dots, y_T
- Гипотеза H_0 : ряд стационарен
- Гипотеза H_1 : ряд не стационарен и описывается моделью $y_t = \alpha y_{t-1}$
- Статистика: $KPSS(y) = \frac{1}{\lambda^2 T^2} \sum_{i=1}^T \left(\sum_{t=1}^i y_t \right)^2$
- При H_0 $KPSS(y)$ имеет табличное распределение



Денис Квятковский
(Denis Kwiatkowski)

?



Питер Филипс
(P.C.B. Philips)

р. 1948



Питер Шмидт
(Peter Schmidt)

р. 1947



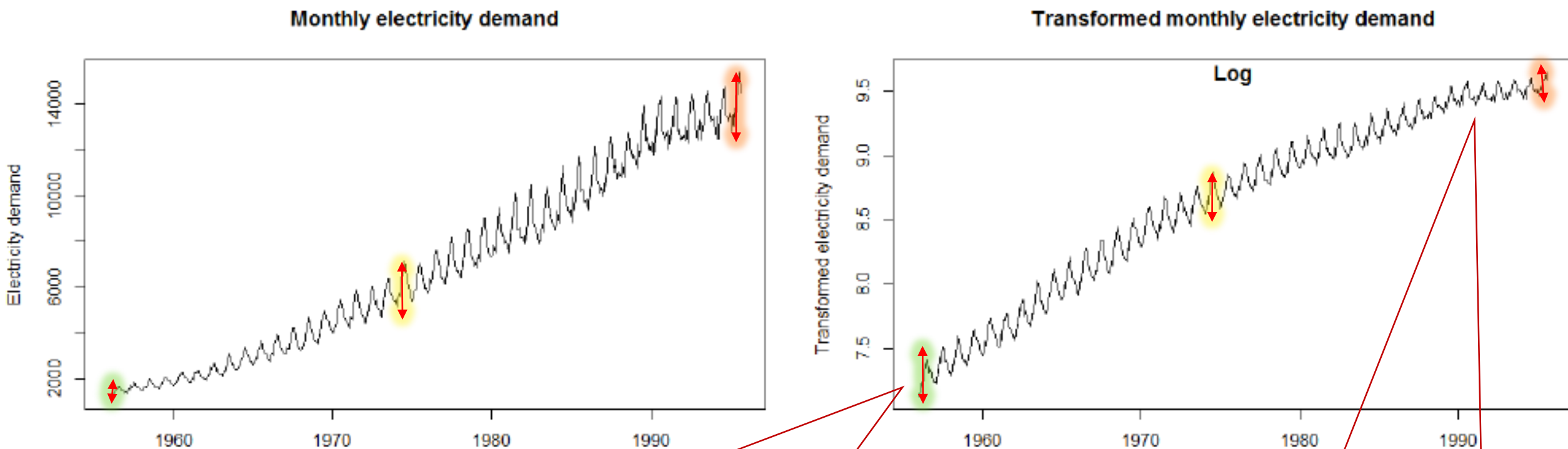
Юнчхоль Шин
(Youngcheol Shin)

р. 1960

Стабилизирующие преобразования ряда

- Стабилизация монотонно меняющейся дисперсии ряда
 - Логарифмирование
 - Преобразование Бокса–Кокса
- Стабилизация среднего значения ряда, удаление тренда и сезонности
 - Дифференцирование ряда
 - Сезонное дифференцирование ряда
 - Комбинированное дифференцирование ряда

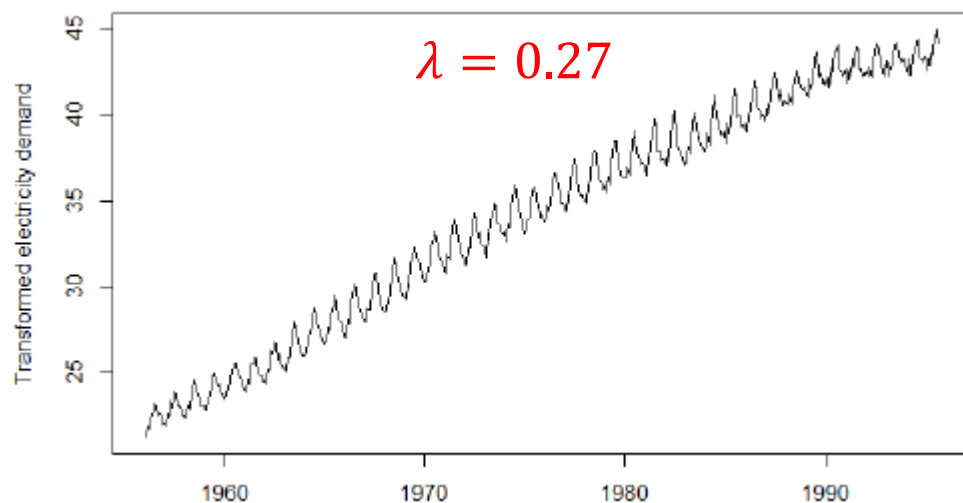
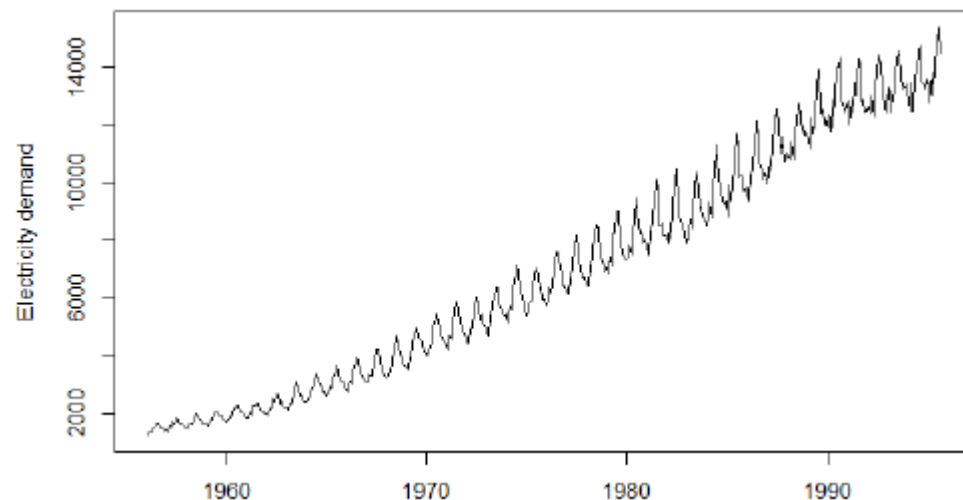
Логарифмирование ряда



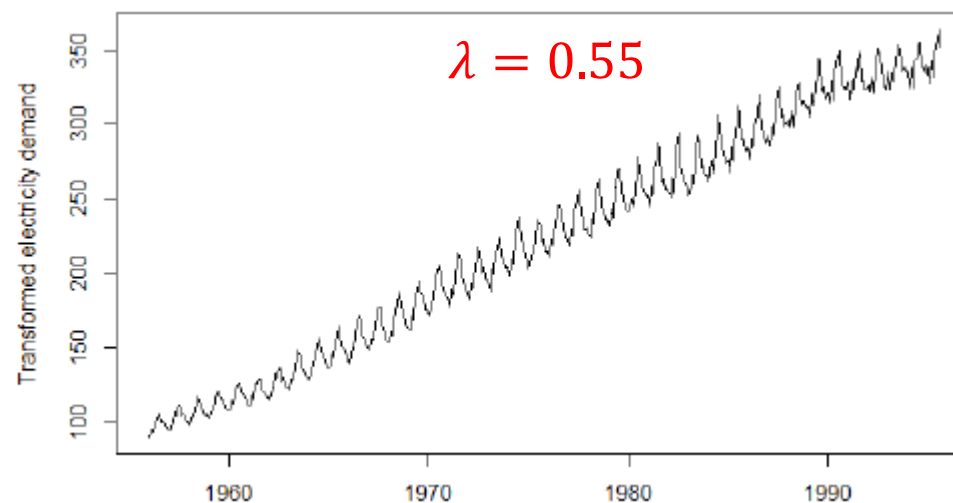
Стабилизация дисперсии
не обязательно приведет к стационарности ряда

После логарифмирования
размах колебаний стал примерно одинаков

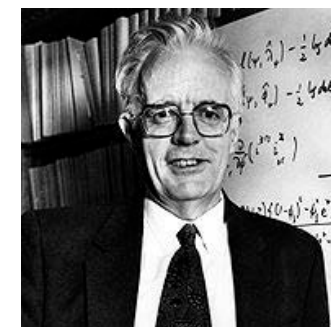
Преобразование Бокса–Кокса (Box–Cox transformation)



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



Дж.Э.П. Бокс
(G.E.P. Box)
1919-2013



Сэр Дэвид Кокс
(Sir D.R. Cox)
1924-2022

Преобразование Бокса–Кокса (Box–Cox transformation)

Преобразование
для выполнения прогноза

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

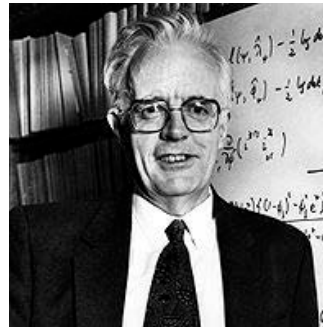
Преобразование прогноза
трансформированного ряда
в прогноз исходного ряда

$$\hat{y}_t = \begin{cases} \exp(y'_t), & \lambda = 0 \\ (\lambda \hat{y}'_t + 1)^{1/\lambda}, & \lambda \neq 0 \end{cases}$$

- Если $\exists y_t \leq 0$, то к ряду нужно добавить константу (и вычесть ее для получения прогноза исходного ряда)
- Значение λ слабо влияет на прогноз (можно округлять)
- Значение λ сильно влияет на предсказательный интервал



Дж.Э.П. Бокс
(G.E.P. Box)
1919-2013



Сэр Д.Р. Кокс
(Sir D.R. Cox)
1924-2022

Дифференцирование ряда для удаления тренда и сезонности

- Переход к попарным разностям соседних значений (дифференцирование 1-го порядка):

$$y_1, \dots, y_T \rightarrow y'_2, \dots, y'_T$$
$$y'_t = y_t - y_{t-1} = (1 - B)y_t$$

- Дифференцирование 2-го порядка

$$y_1, \dots, y_T \rightarrow y'_2, \dots, y'_T \rightarrow y''_3, \dots, y''_T$$
$$y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2} = (1 - B)^2 y_t$$

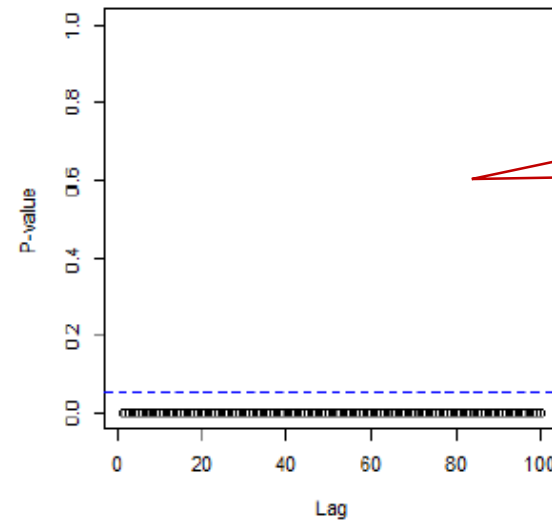
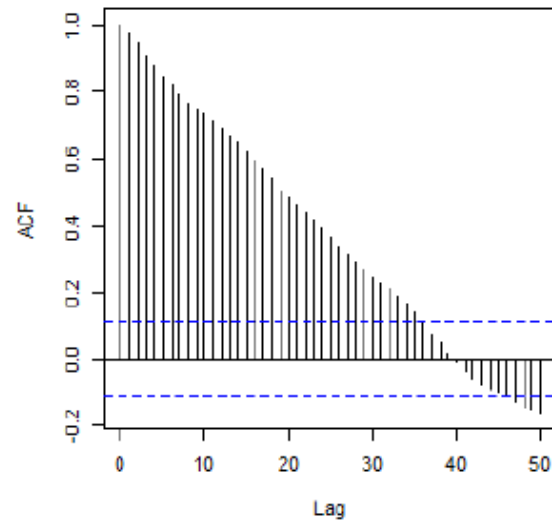
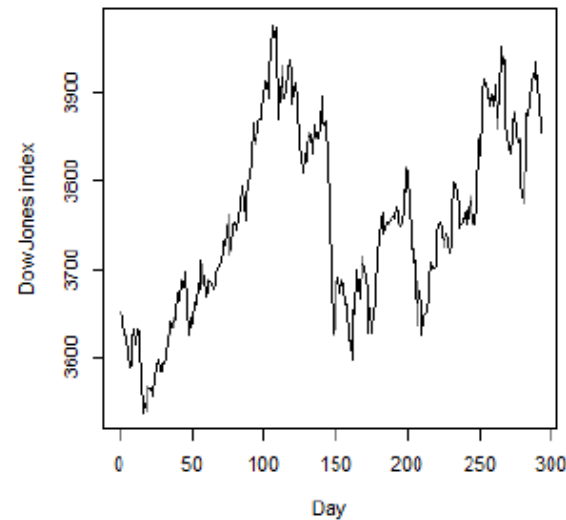
- Дифференцирование 3-го порядка

$$y_1, \dots, y_T \rightarrow y'_2, \dots, y'_T \rightarrow y''_3, \dots, y''_T \rightarrow y'''_4, \dots, y'''_T$$
$$y'''_t = y''_t - y''_{t-1} = y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3} = (1 - B)^3 y_t$$

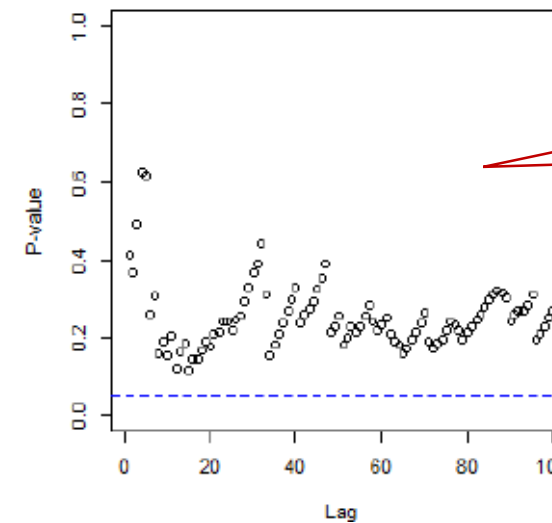
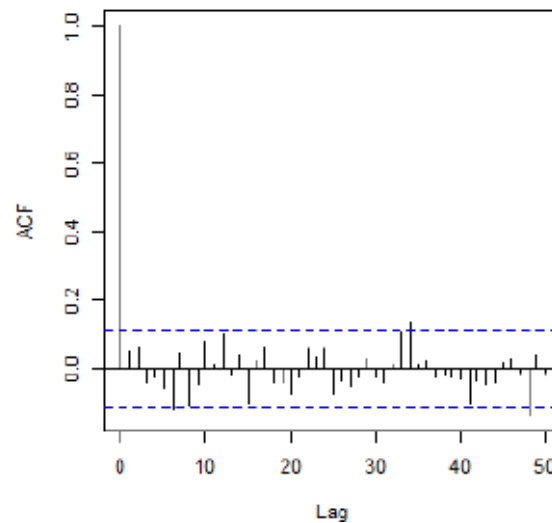
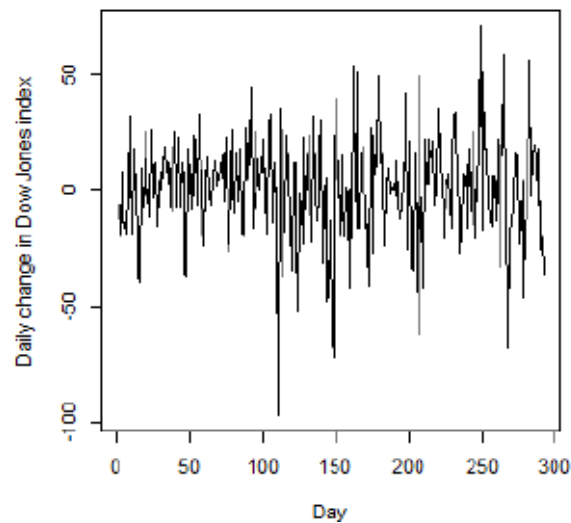
- Дифференцирование d -го порядка

$$y_1, \dots, y_T \rightarrow y'_2, \dots, y'_T \rightarrow \dots \rightarrow y^{(d-1)}_{d+1}, \dots, y^{(d-1)}_T$$
$$\nabla^d y_t = (1 - B)^d y_t$$

Дифференцирование ряда



KPSS: $p < 0.01$



KPSS: $p > 0.1$

Сезонное дифференцирование для удаления сезонности

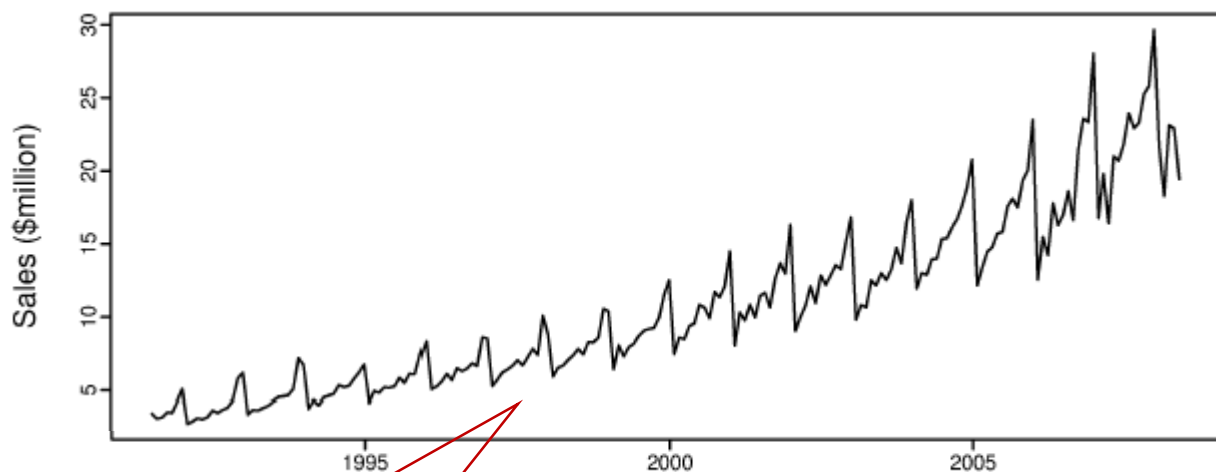
- Переход к попарным разностям значений в соседних сезонах:

$$y_1, \dots, y_T \rightarrow y'_{s+1}, \dots, y'_T$$
$$y'_t = y_t - y_{t-s} = (1 - B^s)y_t$$

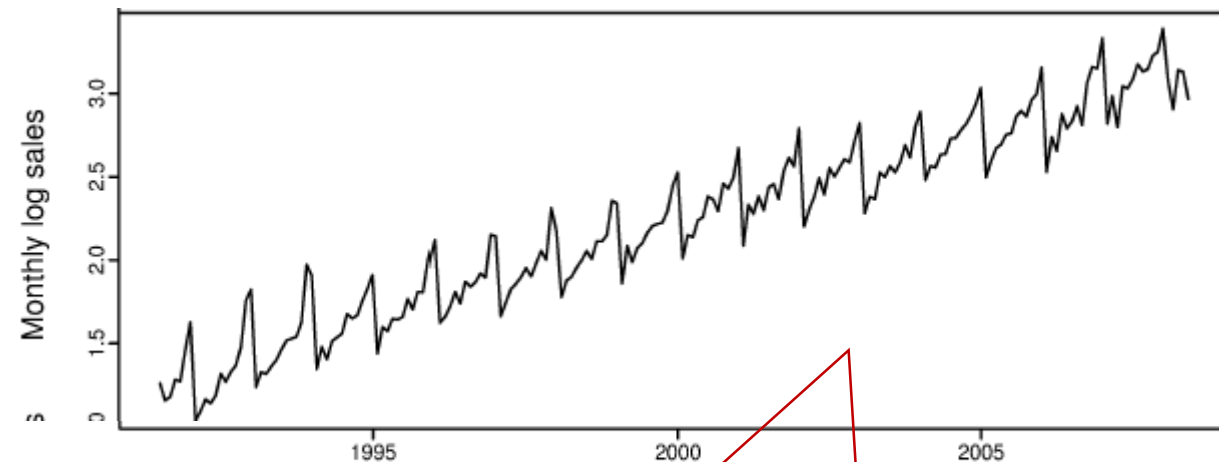
- Может обеспечить стационарность
- Может комбинироваться с обычным дифференцированием:
 $(1 - B)(1 - B^s)y_t$ или $(1 - B^s)(1 - B)y_t$.

При ярко выраженной сезонности лучше начинать с сезонного дифференцирования

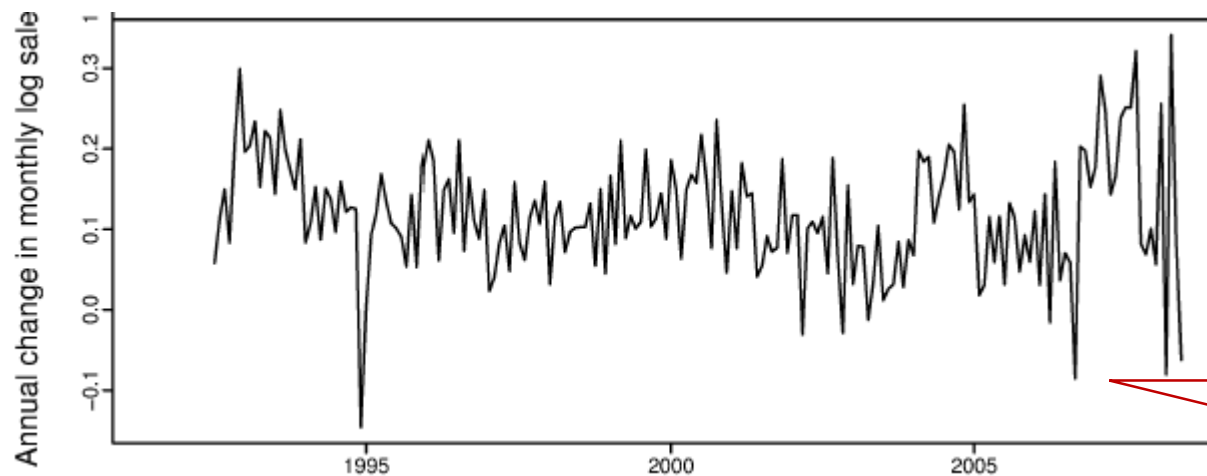
Сезонное дифференцирование



Исходный ряд
KPSS: $p < 0.01$



Логарифмированный ряд
KPSS: $p < 0.01$



Ряд после сезонного
дифференцирования
KPSS: $p > 0.1$

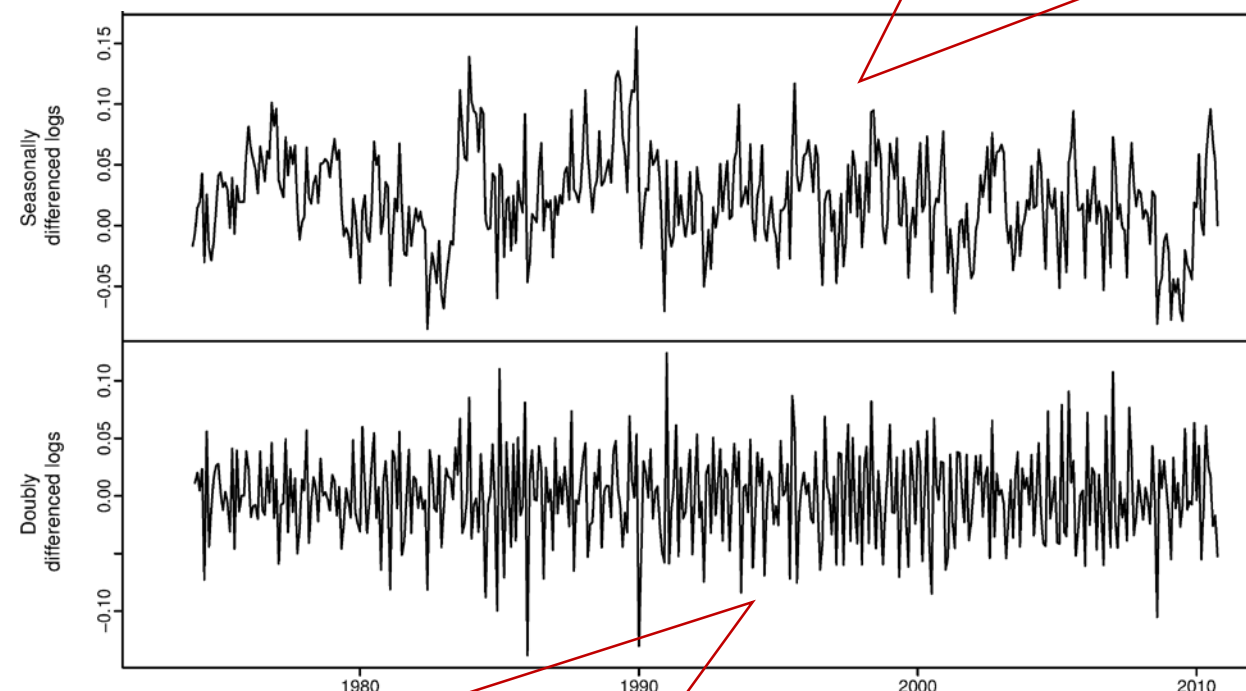
Комбинированное дифференцирование

Исходный ряд,
KPSS: $p < 0.01$



Логарифмированный ряд,
KPSS: $p < 0.01$

Сез. дифф-й логарифмированный
ряд, KPSS: $p = 0.0355$



Дважды диф-й логарифмированный ряд,
KPSS: $p > 0.1$

Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- **Модель AR (авторегрессия ряда)**
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Анализ остатков
- Построение прогноза

Авторегрессия $AR(p)$

- $AR(p)$ -процесс, авторегрессионный процесс порядка p :
$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$
 - y_t – стационарный ряд со средним μ
 - ϕ_1, \dots, ϕ_p – коэффициенты, параметры модели ($\phi_p \neq 0$)
 - ε_t – гауссов белый шум с нулевым средним и постоянной дисперсией σ_ε^2
 - $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$
- Запись посредством оператора обратного сдвига:
$$\phi(B)y_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)y_t = \varepsilon_t, \text{ где } By_t = y_{t-1}$$
- Линейная комбинация p подряд идущих элементов ряда дает белый шум

Авторегрессия AR(1) и AR(2)

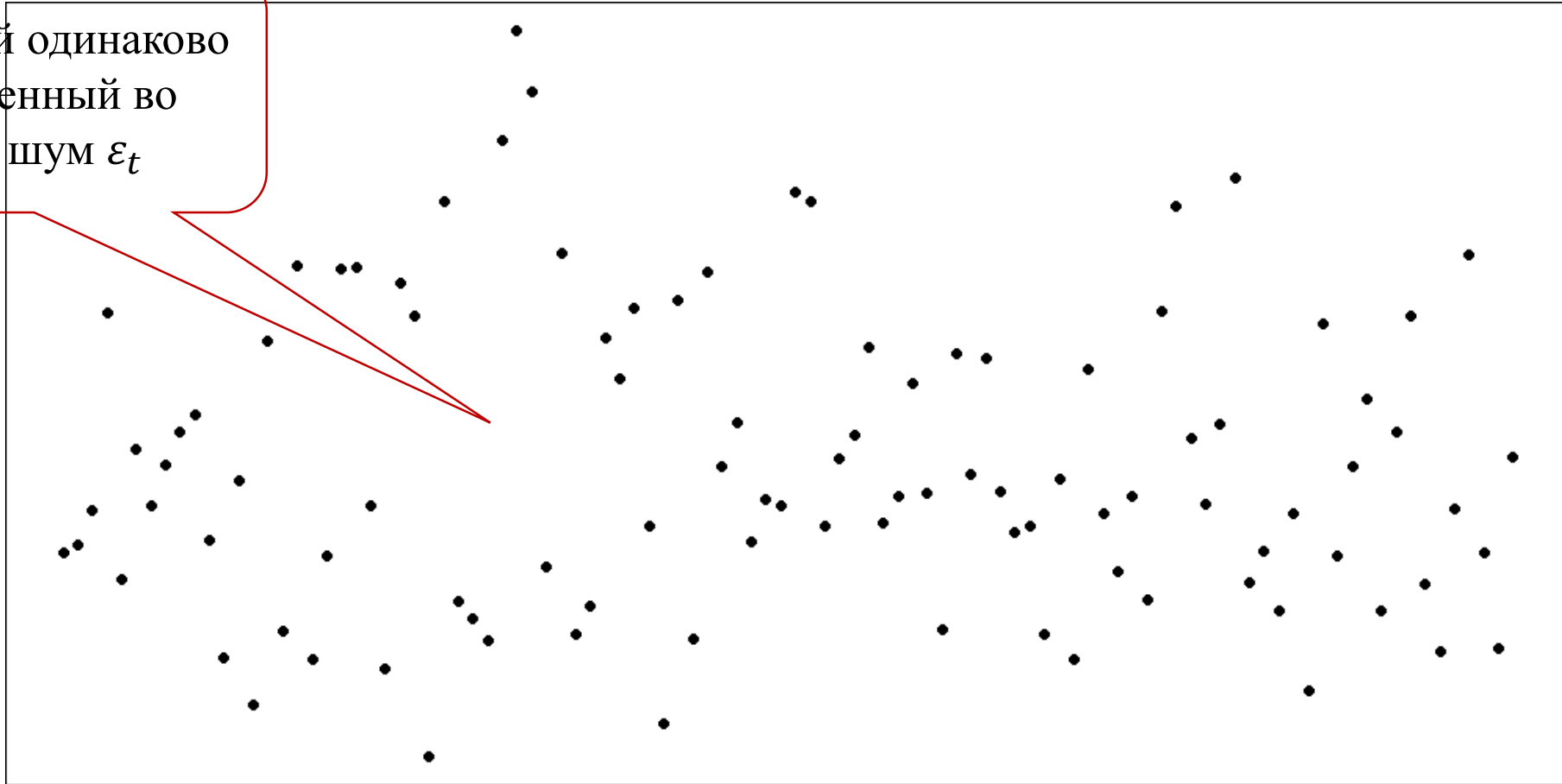
- Ряд $AR(p)$ стационарный, если выполняются ограничения на коэффициенты:
 - $AR(1): -1 < \phi_1 < 1$
 - $AR(2): -1 < \phi_2 < 1, \phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1$
 - $AR(\dots)$: более сложные ограничения

Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- **Модель MA (скользящее среднее ряда)**
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Анализ остатков
- Построение прогноза

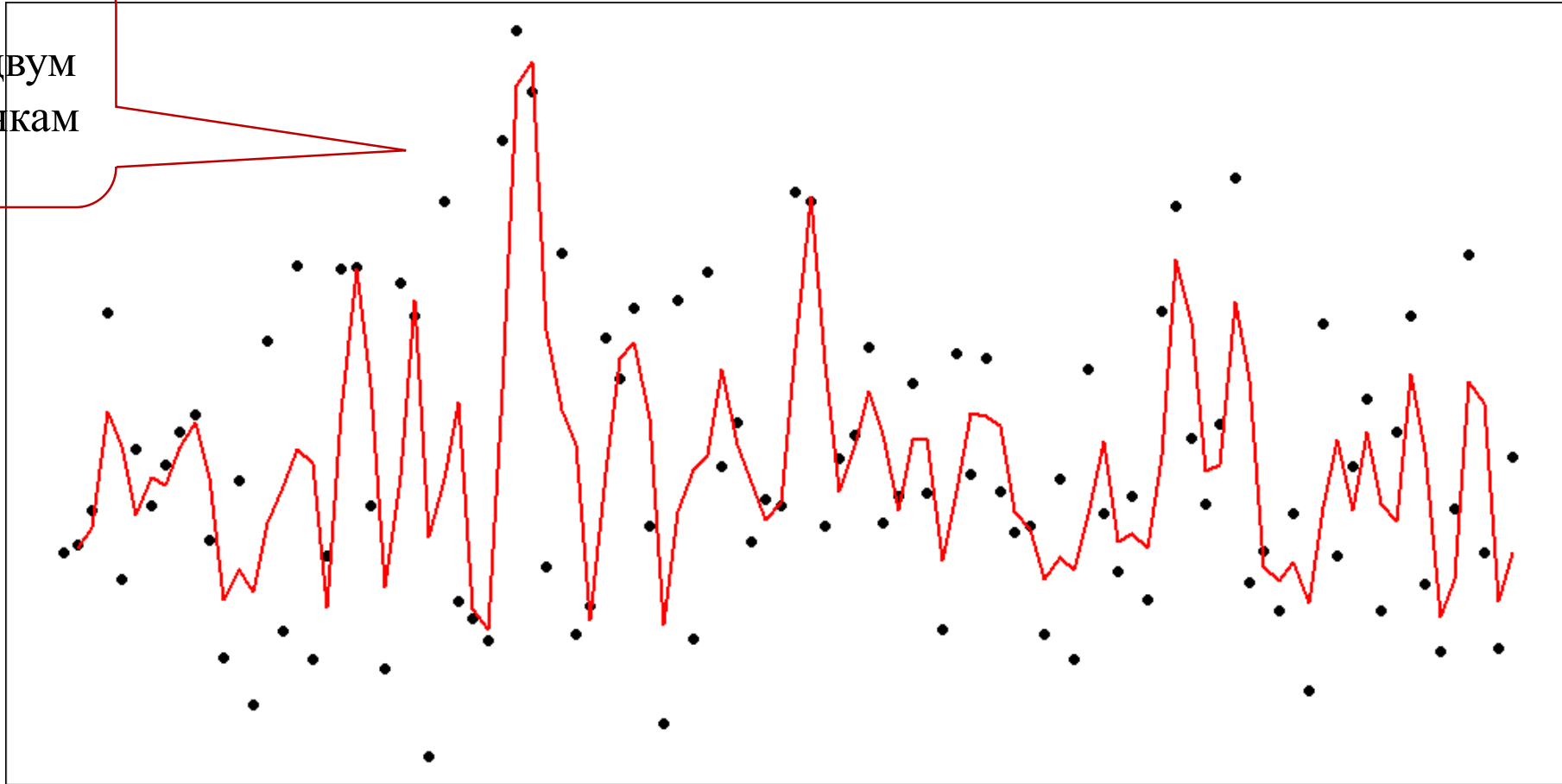
Скользящее среднее

Независимый одинаково
распределенный во
времени шум ε_t



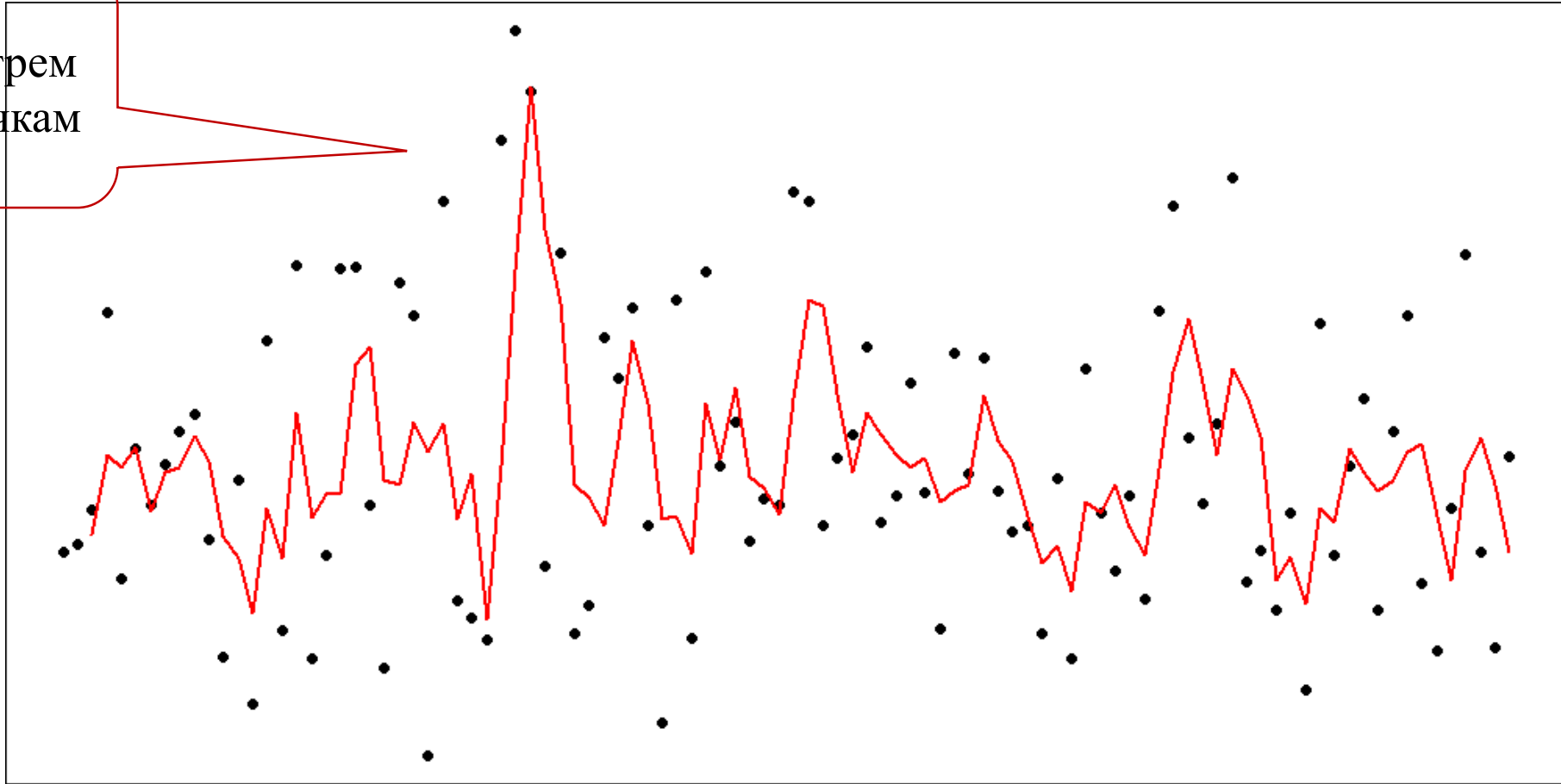
Скользящее среднее

Среднее по двум
соседним точкам



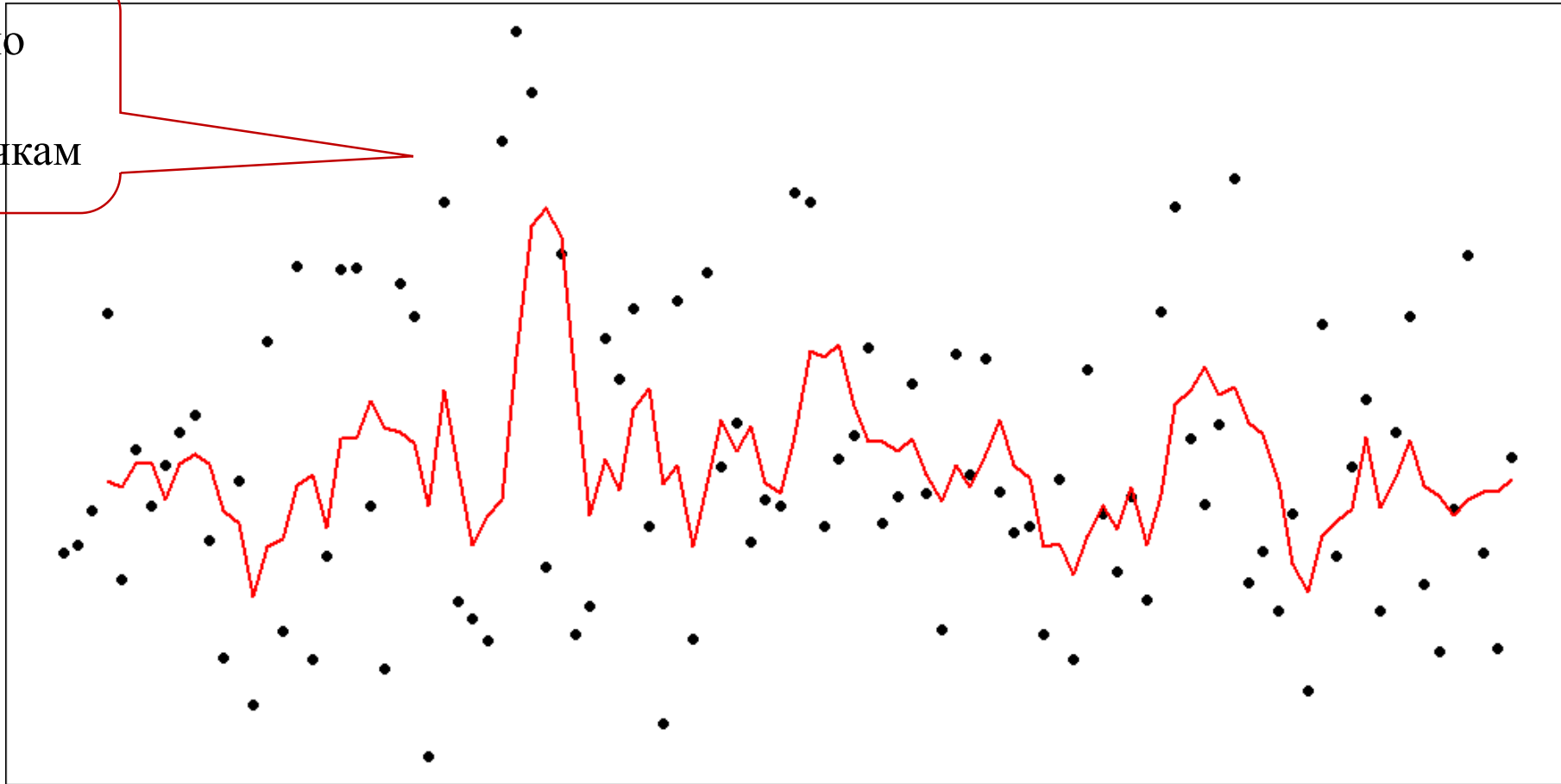
Скользящее среднее

Среднее по трем соседним точкам

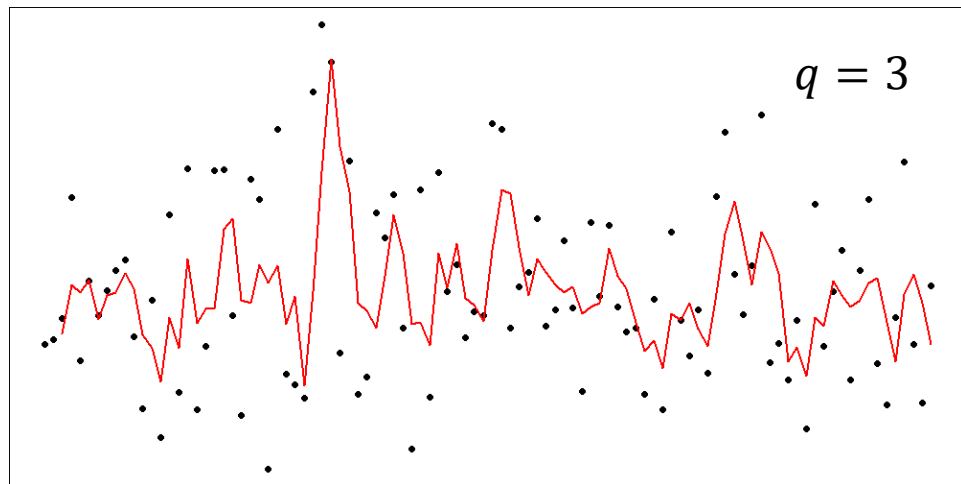
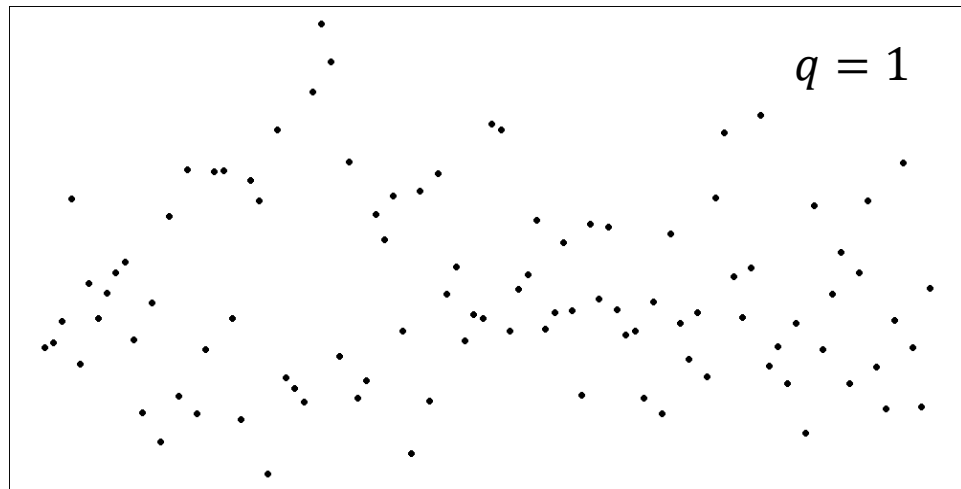


Скользящее среднее

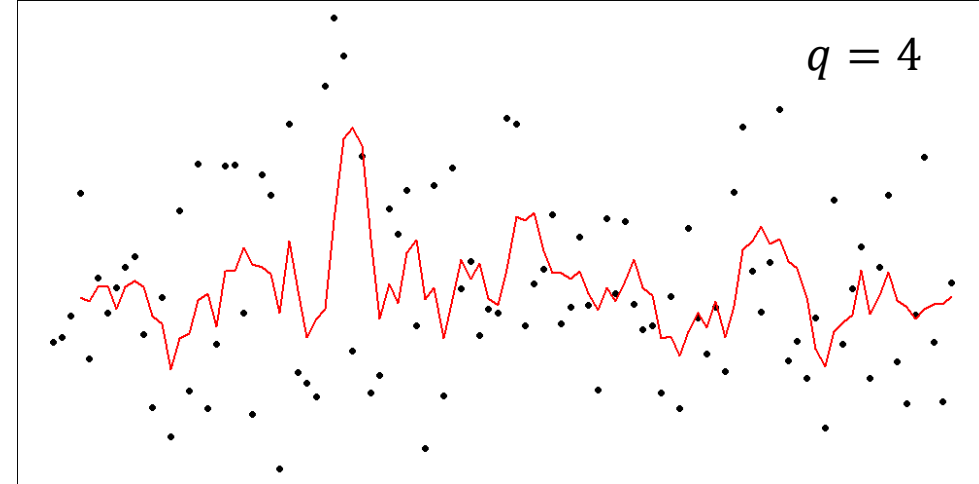
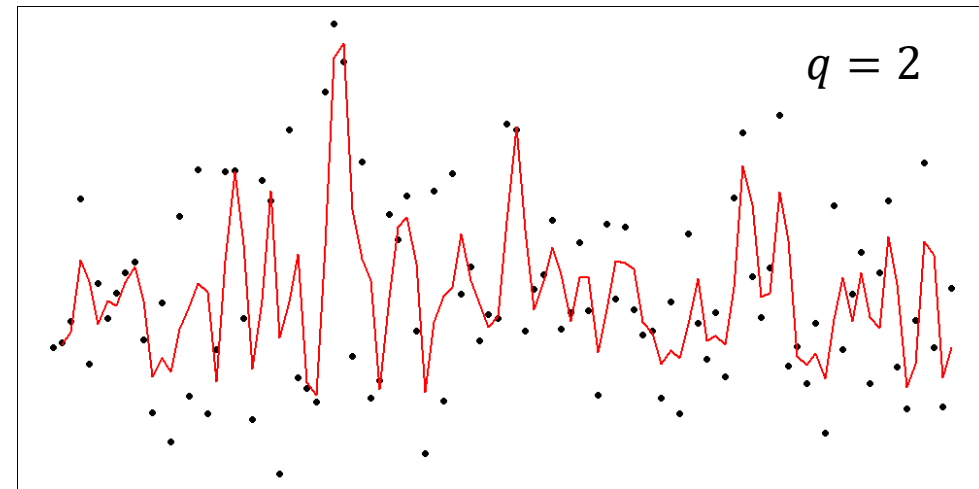
Среднее по
четырем
соседним точкам



Скользящее среднее



Скользящее
усреднение
стабилизирует
шум



Скользящее среднее $MA(q)$

- $MA(\textcolor{red}{q})$:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_{\textcolor{red}{q}} \varepsilon_{t-\textcolor{red}{q}}$$

- y_t – стационарный ряд с нулевым средним
- $\theta_1, \dots, \theta_{\textcolor{red}{q}}$ – коэффициенты, параметры модели ($\theta_{\textcolor{red}{q}} \neq 0$)
- ε_t – гауссов белый шум с нулевым средним и постоянной дисперсией σ_ε^2
- Запись посредством оператора обратного сдвига:
 $y_t = \theta(B)\varepsilon_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_{\textcolor{red}{q}} B^{\textcolor{red}{q}})\varepsilon_t$, где $B y_t = y_{t-1}$
- Линейная комбинация q подряд идущих компонент белого шума дает элемент ряда

Скользящее среднее $MA(1)$ и $MA(2)$

- Модель $MA(p)$ обратима (допускать настройку под данные), если выполняются ограничения на коэффициенты:
 - $MA(1): -1 < \theta_1 < 1$
 - $MA(2): -1 < \theta_2 < 1, \theta_1 + \theta_2 > -1, \phi_2 - \phi_1 < 1$
 - $MA(\dots)$: более сложные ограничения

Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- **Модели ARMA, ARIMA, SARMA, SARIMA**
- Подбор параметров моделей
- Анализ остатков
- Построение прогноза

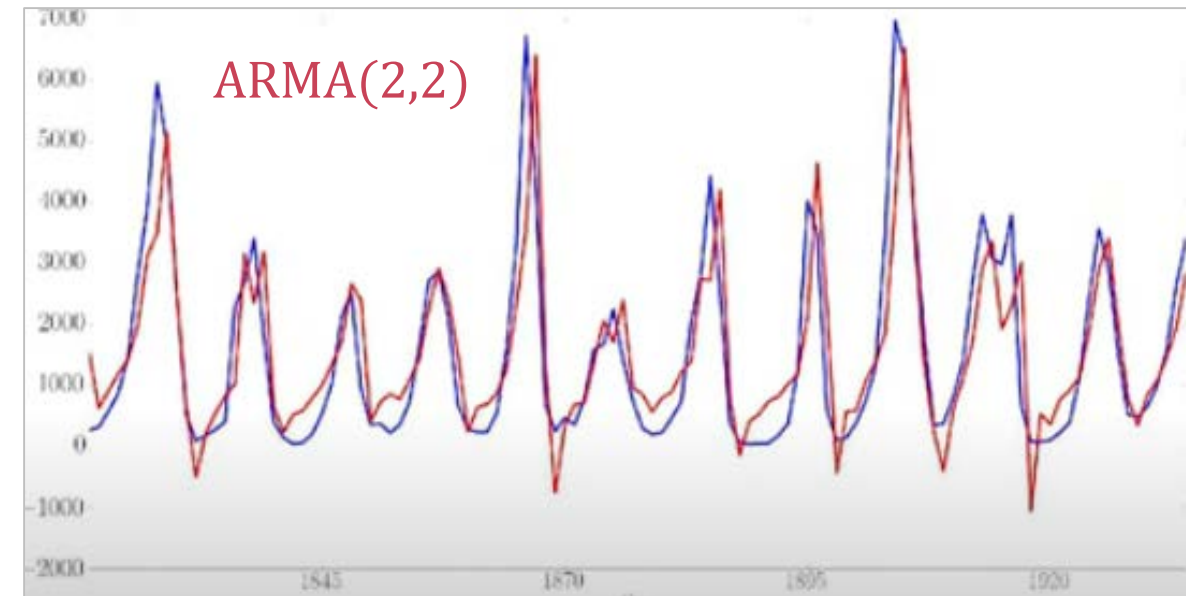
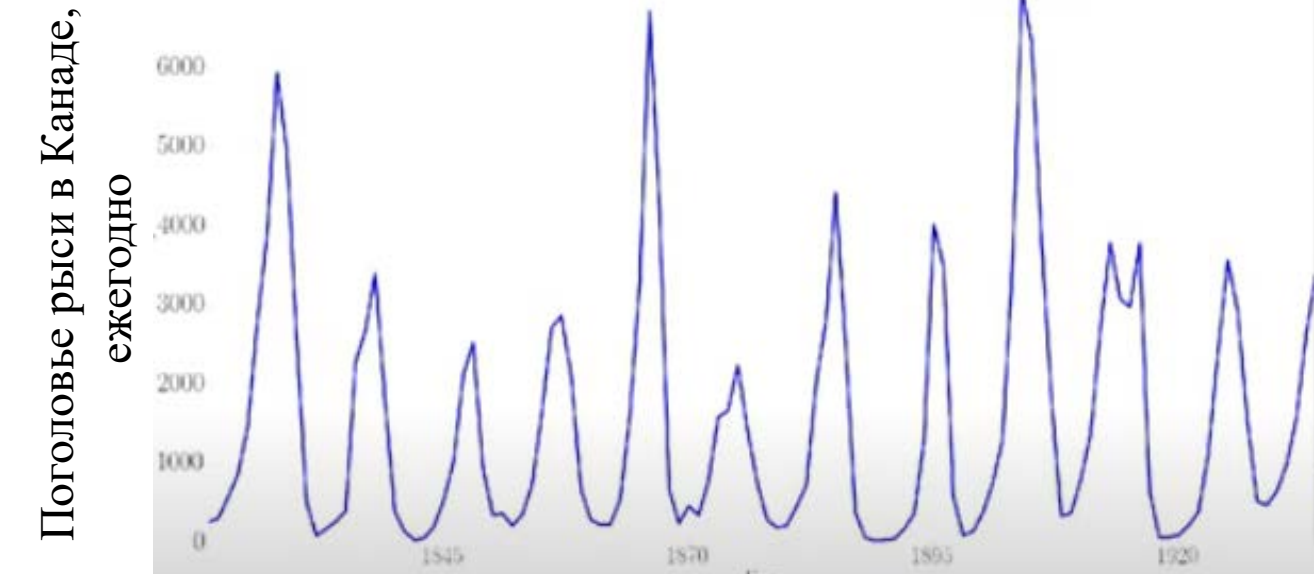
Модель $\text{ARMA}(p, q) = \text{AR}(p) + \text{MA}(q)$ (AutoRegressive Moving Average)

- $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$
 - y_t – стационарный ряд с нулевым средним
 - ϕ_1, \dots, ϕ_p и $\theta_1, \dots, \theta_q$ – параметры модели ($\theta_q \neq 0, \phi_p \neq 0$)
 - ε_t – гауссов белый шум с нулевым средним и постоянной дисперсией σ_ε^2
- Вид модели, если среднее равно μ :
$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$
$$\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$$
- Запись посредством оператора обратного сдвига: $\phi(B)y_t = \theta(B)\varepsilon_t$
- Теорема Волда: любой стационарный ряд может быть аппроксимирован моделью $\text{ARMA}(p, q)$ с любой точностью



Х.О.А. Волд
(H.O.A. Wold)
1908-1992

Модель $ARMA(p, q)$: пример



Модель ARIMA (AutoRegressive Integrated Moving Average)

- Ряд описывается моделью $ARIMA(p, d, q)$, если его дифференцирование d -го порядка

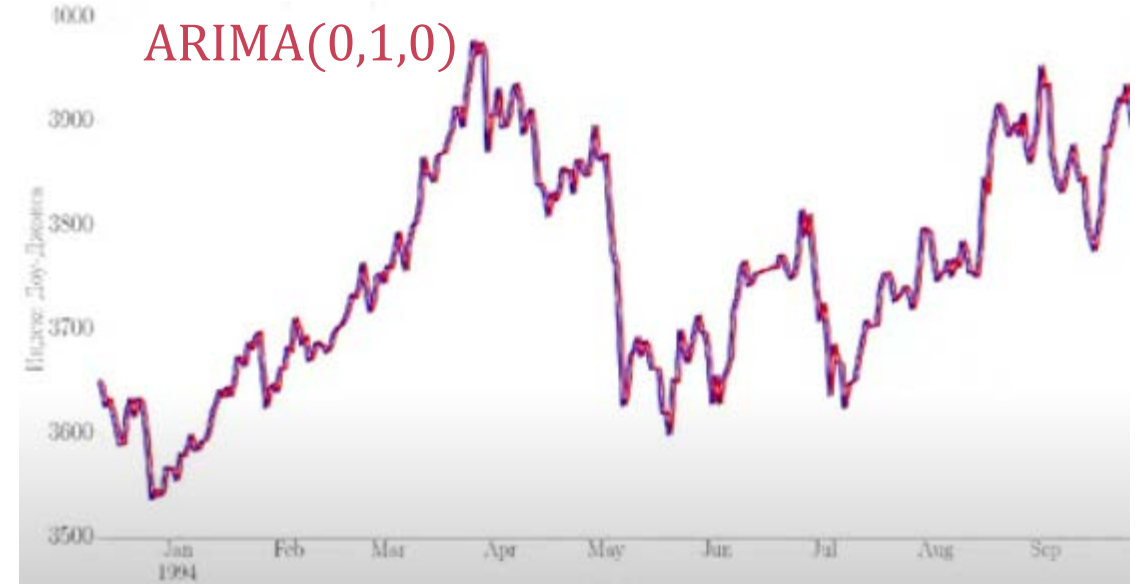
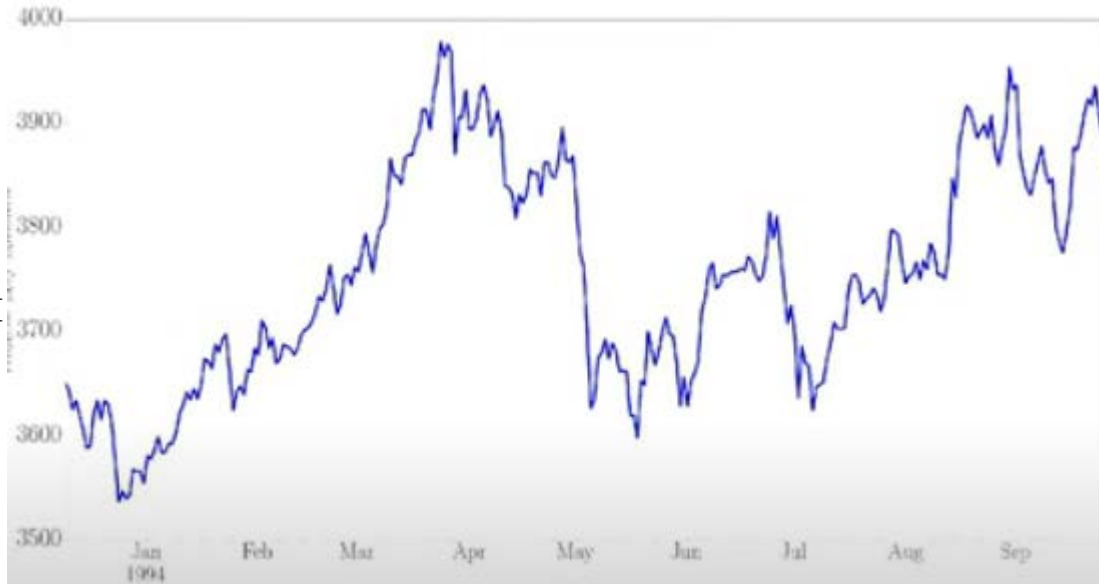
$$\nabla^d y_t = (1 - B)^d y_t$$

описывается моделью $ARMA(p, q)$

$$\phi(B)\nabla^d y_t = \theta(B)\varepsilon_t$$

Модель $ARIMA(p, q, d)$: пример

Индекс Доу Джонса,
ежедневно



Модель Seasonal ARMA: $\text{SARMA}(p, q) \times (P, Q)_S = \text{ARMA}(p, q) \times (P, Q)_S$

- В ряд, имеющий сезонный период длины S , описываемый моделью $\text{ARMA}(p, q)$

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$
 добавлены P компонент авторегрессии

$$+ \phi_S y_{t-S} + \phi_{2S} y_{t-2S} + \dots + \phi_{PS} y_{t-PS}$$
 и Q компонент скользящего среднего

$$+ \theta_S y_{t-S} + \theta_{2S} y_{t-2S} + \dots + \theta_{QS} y_{t-QS}$$
- $\text{ARMA}(p, q) \times (P, Q)_S$: $\Phi_P(B^S)\phi(B)y_t = \alpha + \Theta_Q(B^S)\theta(B)y_t$, где

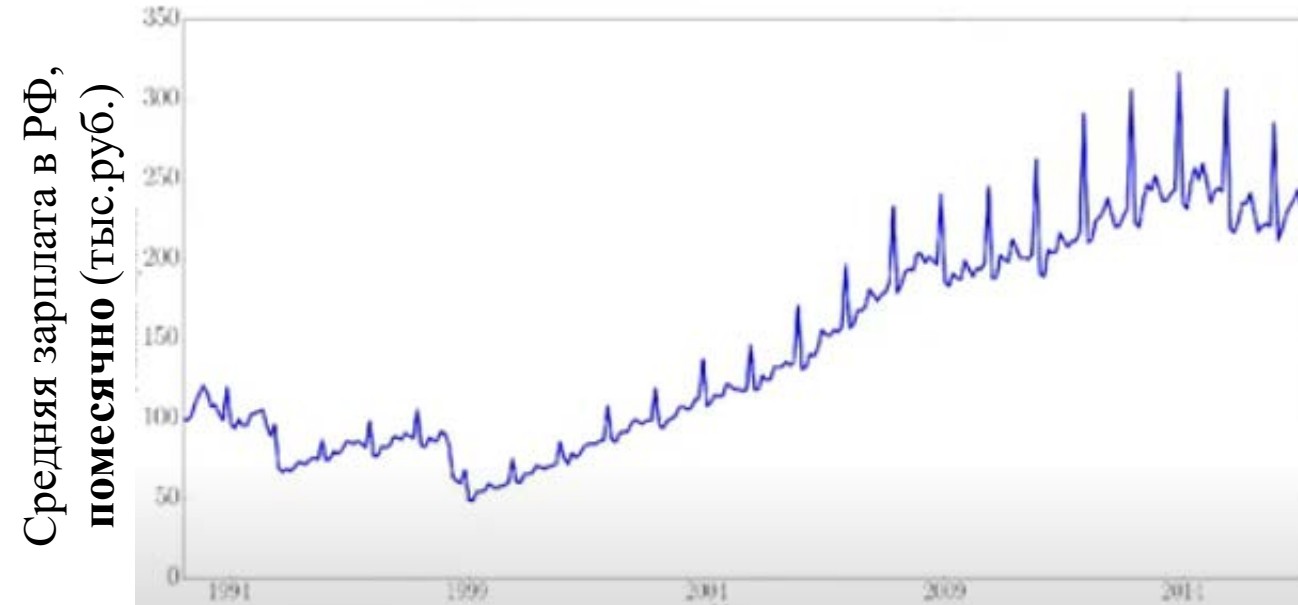
$$\Phi_P(B^S) = 1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_P B^{PS}$$

$$\Theta_Q(B^S) = 1 + \theta_1(B^S) + \theta_2(B^{2S}) + \dots + \theta_Q(B^{QS})$$

Модель Seasonal Integrated ARMA: SARIMA(p, d, q) \times (P, D, Q)

- SARIMA(p, d, q) \times (P, D, Q) – это модель SARMA(p, q) \times (P, Q)_S для ряда, к которому применили дифференцирование d -го порядка и сезонное дифференцирование D -го порядка
- $\Phi_P(B^S)\phi(B)\nabla_S^D\nabla^d y_t = \alpha + \Theta_Q(B^S)\theta(B)y_t$, где
$$\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}$$
$$\Theta_Q(B^S) = 1 + \Theta_1(B^S) + \Theta_2(B^{2S}) + \dots + \Theta_Q(B^{QS})$$

Модель $SARIMA(p, d, q) \times (P, D, Q)$: пример



Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- **Подбор параметров моделей**
- Анализ остатков
- Построение прогноза

Параметры моделей

- α, ϕ, θ
- d, D
- q, Q
- p, P

Подбор α , ϕ , θ

- Если все остальные параметры фиксированы (p, d, q, P, D, Q), то коэффициенты регрессии ϕ_1, \dots, ϕ_p подбираются методом наименьших квадратов (МНК)
- Чтобы найти коэффициенты скользящего среднего $\theta_1, \dots, \theta_q$, шумовая компонента предварительно оценивается с помощью остатков авторегрессии
- Если шум белый (независимый одинаково распределенный гауссовский), то МНК дает оценки максимального правдоподобия

Подбор d, D

- Подбор порядков дифференцирования выполняется так, чтобы ряд стал стационарным
- Следует начинать с сезонного дифференцирования
- Меньший порядок дает меньшую дисперсию итогового прогноза

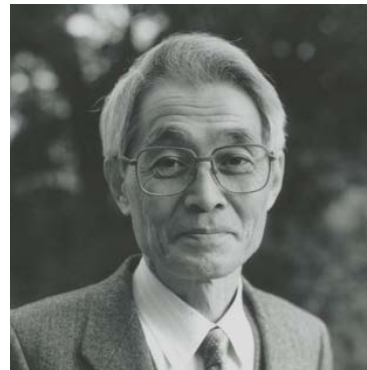
Подбор q, Q, p, P

- Начальные приближения выбираются на основе автокорреляций
- Для сравнения моделей с разными параметрами используется информационный критерий Акаике

$$AIC = 2k - 2\ln(L),$$

k – количество параметров модели ($k = p + P + q + Q + 1$),

L – максимум функции правдоподобия модели



Хироцугу Акаике
(Hirotugu Akaike)

1927-2009

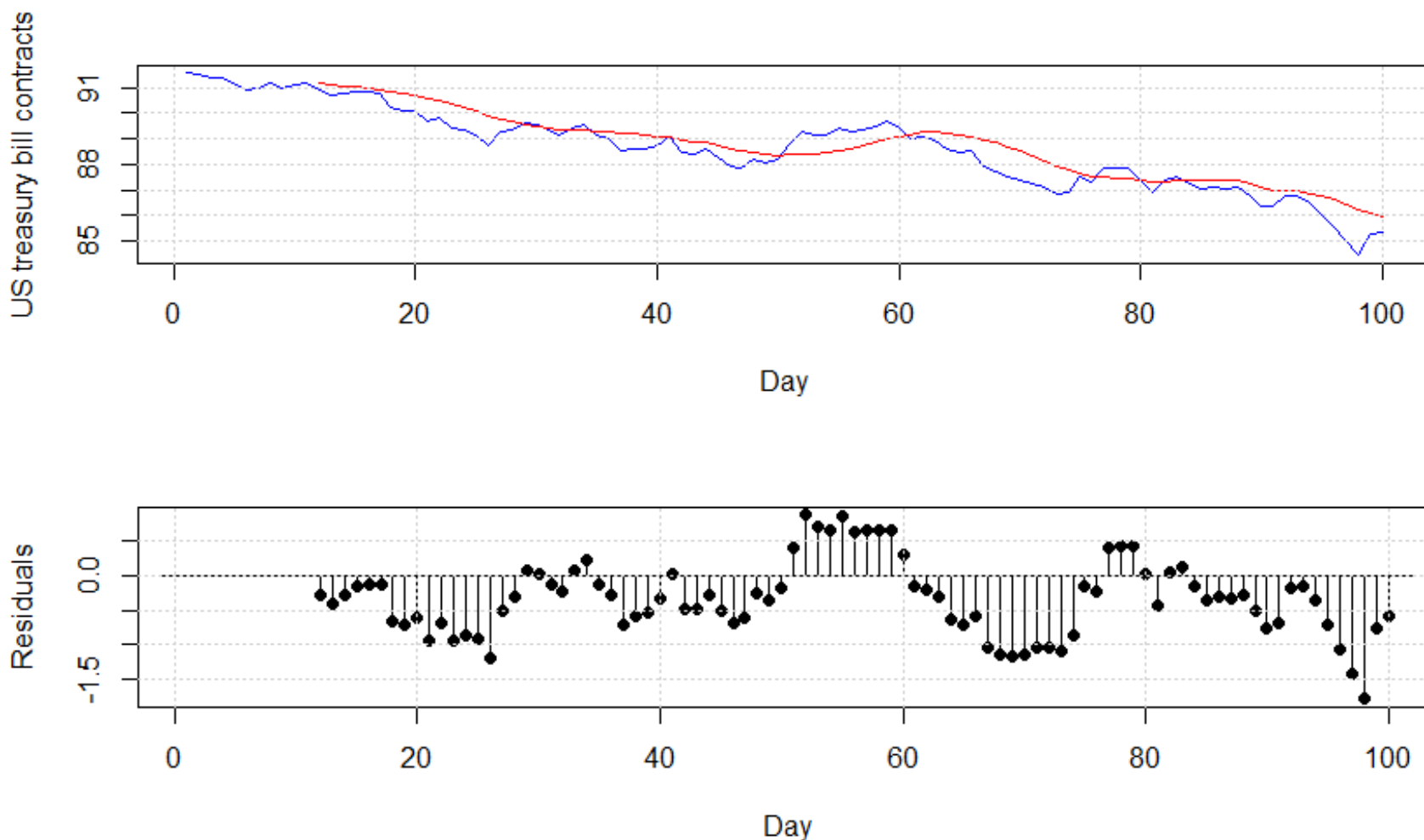
Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- **Анализ остатков**
- Построение прогноза

Остатки (residuals): $\hat{\varepsilon}_t = y_t - \hat{y}_t$, разность между фактом и прогнозом

- **Необходимые свойства остатков**
 - несмещенность
 - стационарность
 - неавтокоррелированность
- **Желательные свойства остатков**
 - нормальность распределения

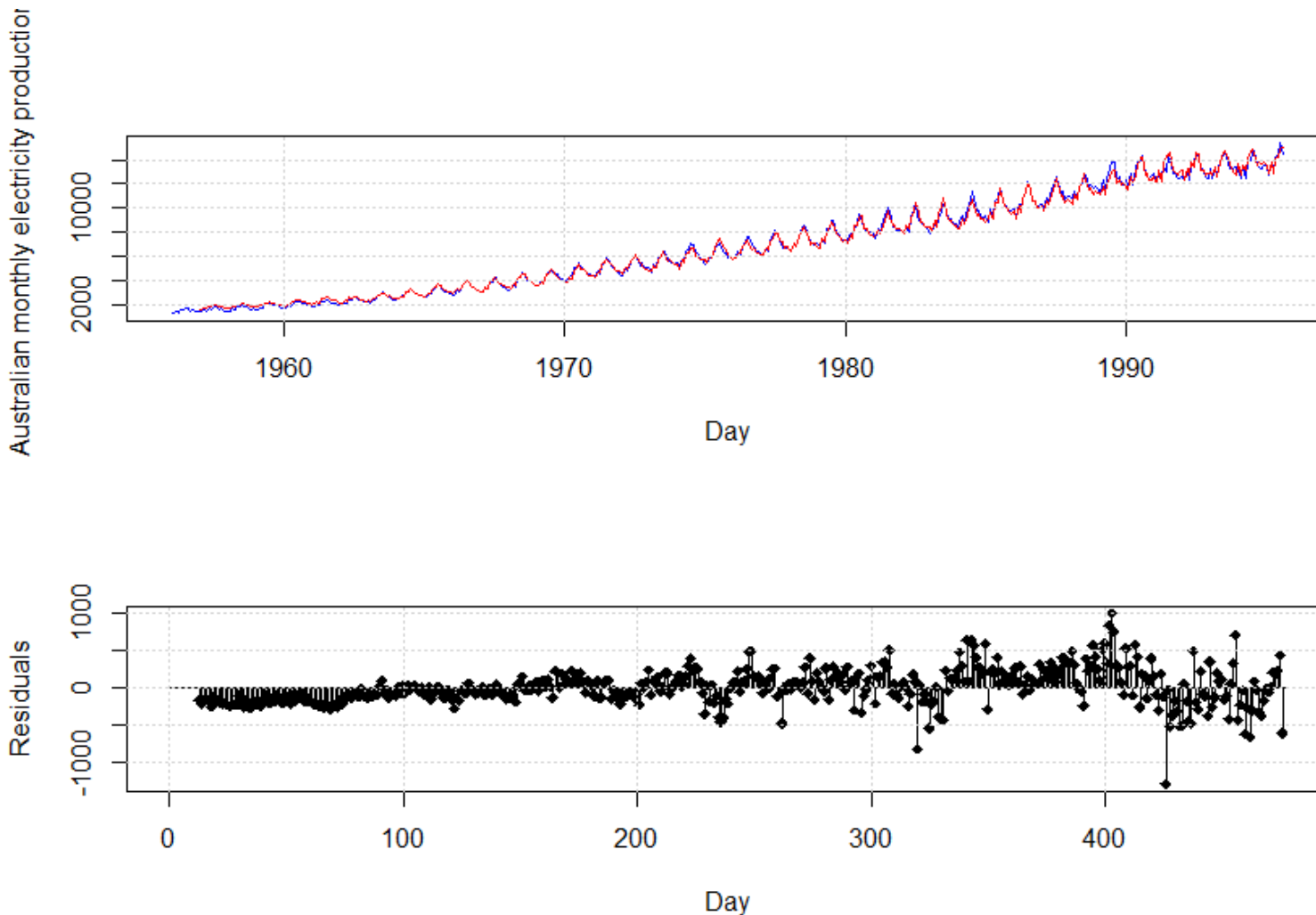
Несмещенность остатков (равенство нулю среднего значения)



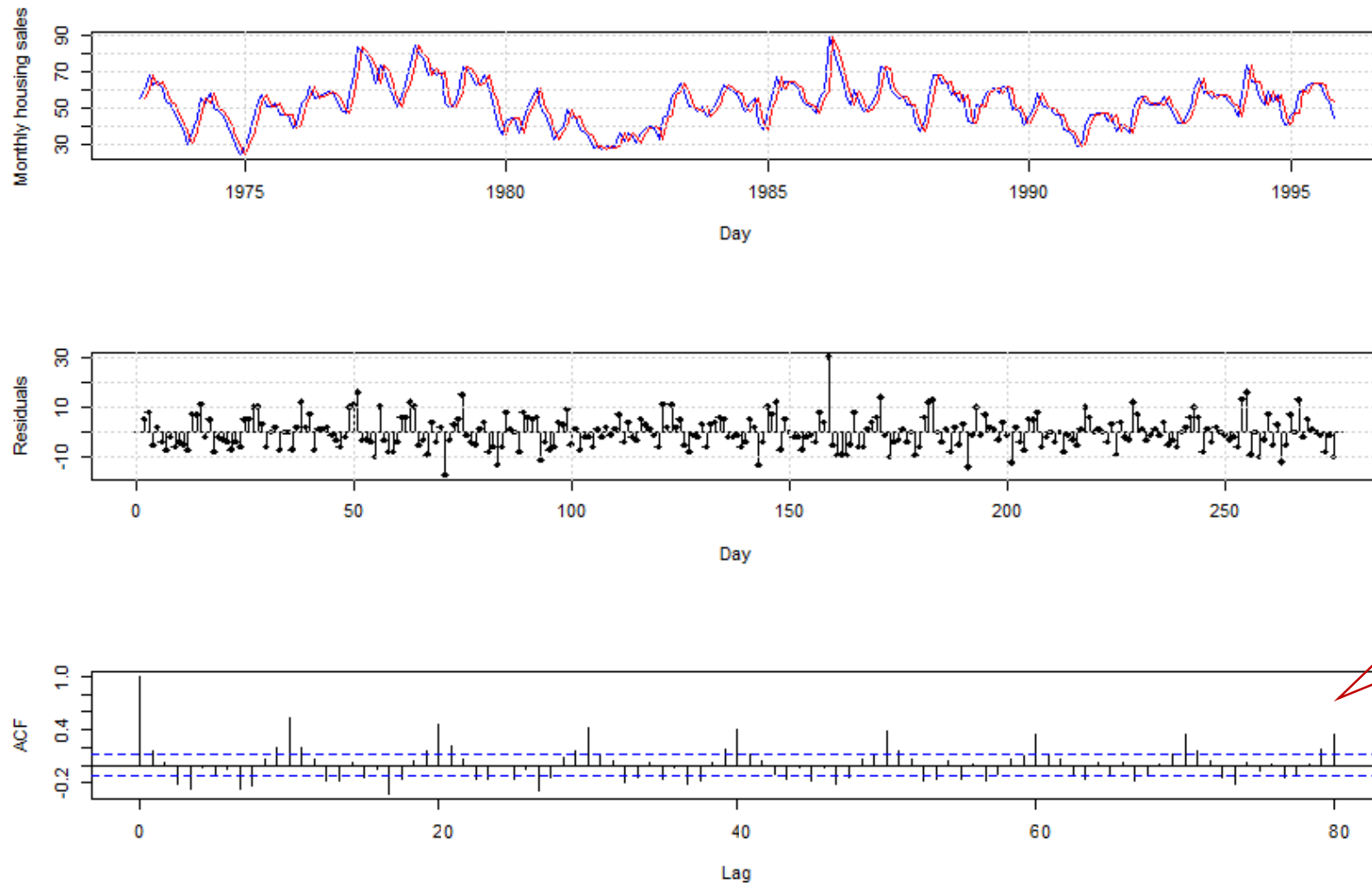
- Проверка критериями Стьюдента или Уилкоксона

Стационарность остатков

- Визуальная проверка, критерий KPSS



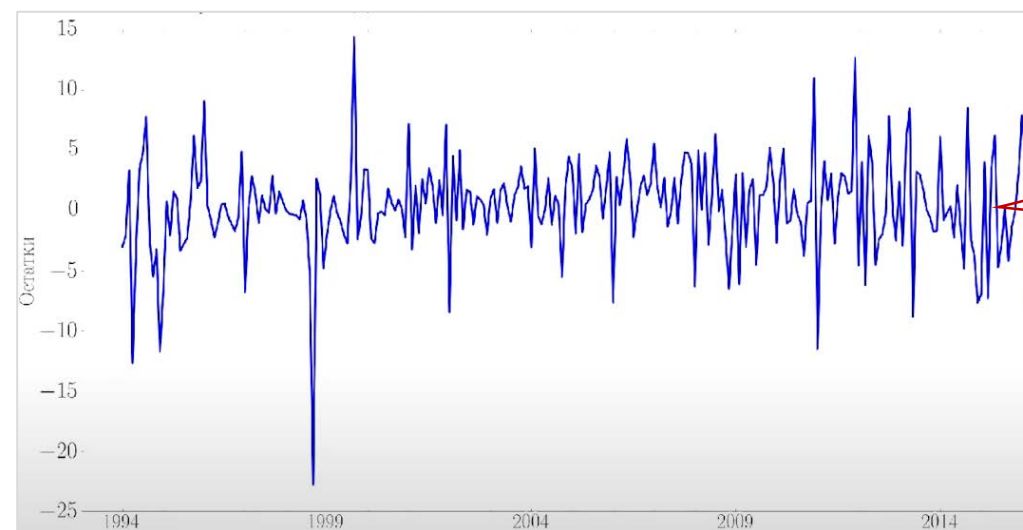
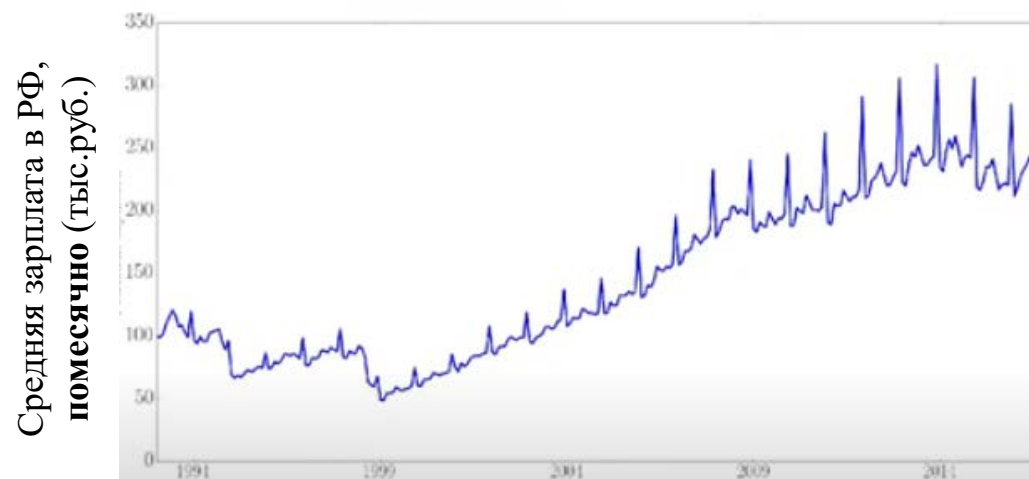
Неавтокоррелированность остатков



- Визуальная проверка АСФ, Q-критерий Льюнга–Бокса

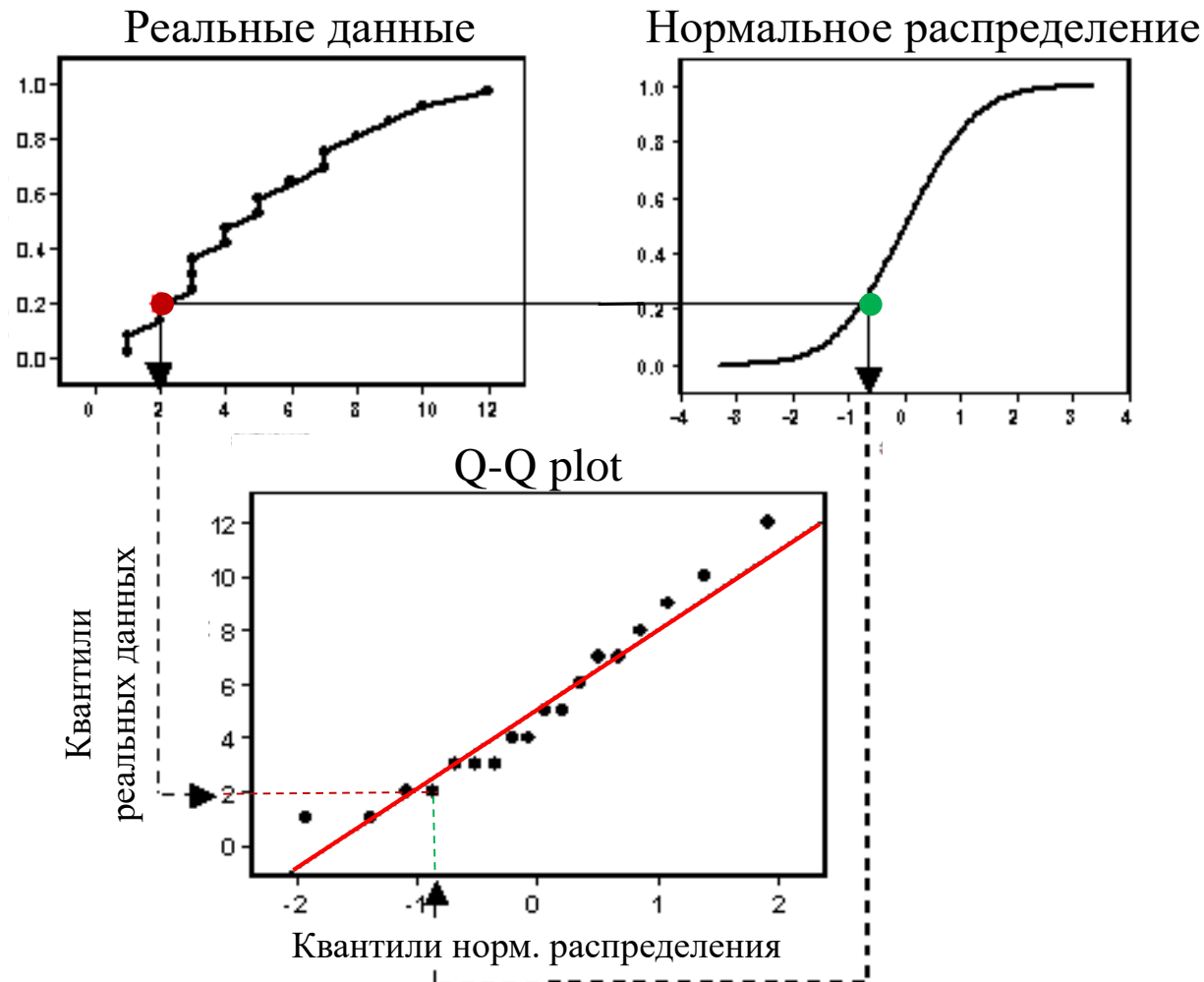
Отсутствие автокорреляции (белый шум) в остатках показывает, что при прогнозе учтены все зависимости текущего значения ряда от предыдущих

Пример: остатки модели $SARIMA(p, d, q) \times (P, D, Q)$



Остатки модели
выглядят
как белый шум

Нормальность распределения остатков



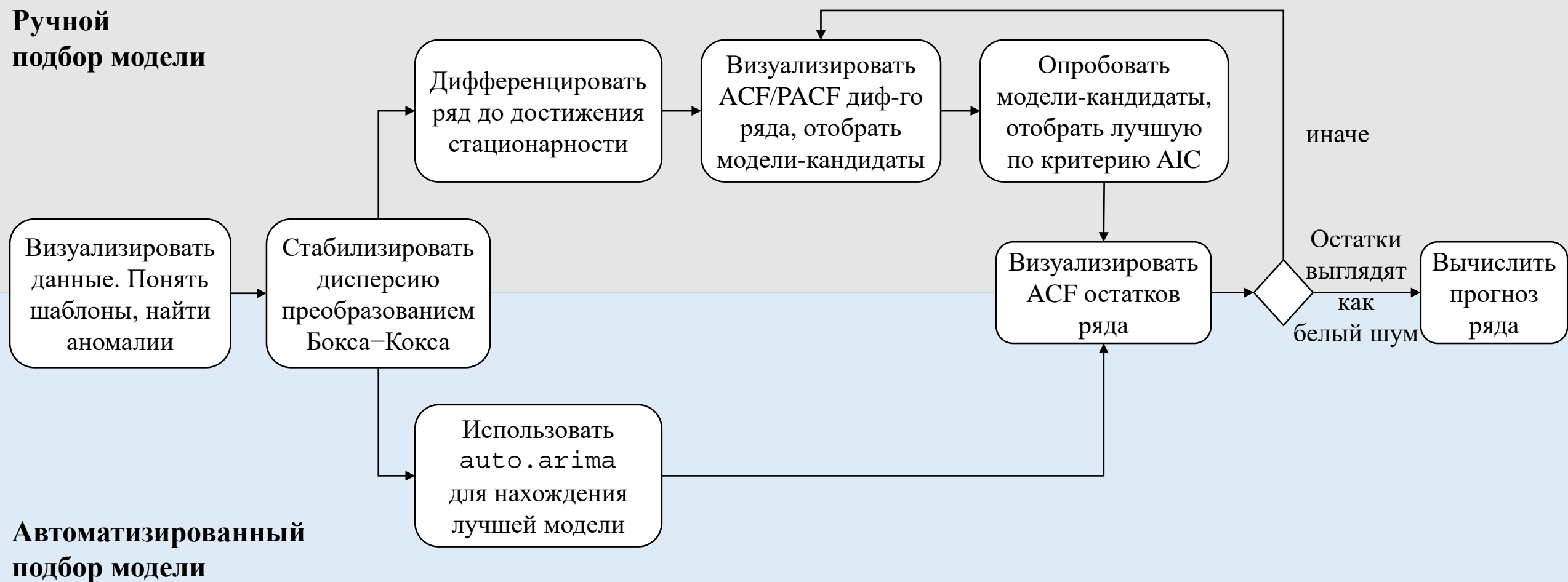
- Визуализация Q-Q plot, один из 20+ критериев нормальности (№ 1 – критерий Шапиро–Уилка)

Содержание

- Постановка задачи прогнозирования
- Компоненты временного ряда
- Автокорреляция временного ряда
- Стационарность и стабилизирующие преобразования ряда
- Модель AR (авторегрессия ряда)
- Модель MA (скользящее среднее ряда)
- Модели ARMA, ARIMA, SARMA, SARIMA
- Подбор параметров моделей
- Анализ остатков
- **Построение прогноза**

Общая схема прогноза

Ручной подбор модели



Вычисление прогноза

- Модель

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \dots + \hat{\phi}_p y_{t-p} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q}$$

Вычисление прогноза

- Модель

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \dots + \hat{\phi}_p y_{t-p} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q}$$

- Замена t на $T + 1$

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \varepsilon_{T+1} + \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$

Вычисление прогноза

- Модель

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \dots + \hat{\phi}_p y_{t-p} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q}$$

- Замена t на $T + 1$

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \boldsymbol{\varepsilon}_{T+1} + \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$

- Замена будущих ошибок на нули

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \mathbf{0} + \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$



Вычисление прогноза

- Модель

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \dots + \hat{\phi}_p y_{t-p} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q}$$

- Замена t на $T + 1$

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \varepsilon_{T+1} + \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$

- Замена будущих ошибок на нули

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \hat{\theta}_1 \mathbf{\varepsilon}_T + \dots + \hat{\theta}_q \mathbf{\varepsilon}_{T+1-q}$$

- Замена прошлых ошибок на остатки

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \hat{\theta}_1 \hat{\varepsilon}_T + \dots + \hat{\theta}_q \hat{\varepsilon}_{T+1-q}$$



Вычисление прогноза

- Модель

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \dots + \hat{\phi}_p y_{t-p} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q}$$

- Замена t на $T + 1$

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \varepsilon_{T+1} + \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$

- Замена будущих ошибок на нули

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \hat{\theta}_1 \varepsilon_T + \dots + \hat{\theta}_q \varepsilon_{T+1-q}$$

- Замена прошлых ошибок на остатки

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1 y_T + \dots + \hat{\phi}_p y_{T+1-p} + \hat{\theta}_1 \hat{\varepsilon}_T + \dots + \hat{\theta}_q \hat{\varepsilon}_{T+1-q}$$

- Применение прогнозного значения в будущих прогнозах

$$\hat{y}_{T+2|T} = \hat{\alpha} + \hat{\phi}_1 \mathbf{y}_{T+1} + \dots + \hat{\phi}_p y_{T+2-p} + \hat{\theta}_1 \hat{\varepsilon}_{T+1} + \dots + \hat{\theta}_q \hat{\varepsilon}_{T+2-q}$$

$$\hat{y}_{T+2|T} = \hat{\alpha} + \hat{\phi}_1 \hat{y}_{T+1|T} + \dots + \hat{\phi}_p y_{T+2-p} + \hat{\theta}_1 \hat{\varepsilon}_{T+1} + \dots + \hat{\theta}_q \hat{\varepsilon}_{T+2-q}$$

Реализация в R, пакет forecast

Подбор оптимальных параметров ARIMA

```
auto.arima(x, d=NA, D=NA, max.p=5, max.q=5,  
  max.P=2, max.Q=2, max.order=5, max.d=2,  
  max.D=1, start.p=2, start.q=2, start.P=1, start.Q=1,  
  stationary=FALSE,  
  seasonal=TRUE, ic=c("aicc", "aic", "bic"),  
  stepwise=TRUE, trace=FALSE,  
  approximation=(length(x)>100 | frequency(x)>12),  
  truncate=NULL, xreg=NULL,  
  test=c("kpss", "adf", "pp"),  
  seasonal.test=c("ocsb", "ch"), allowdrift=TRUE,  
  allowmean=TRUE, lambda=NULL,  
  parallel=FALSE, num.cores=2, ...)
```

Прогноз по подобранной модели

```
forecast(object,  
  h=ifelse(frequency(object)>1,  
    2*frequency(object),10),  
  level=c(80,95), fan=FALSE, robust=FALSE,  
  lambda=NULL, find.frequency=FALSE,  
  allow.multiplicative.trend=FALSE, ...)
```

Параметр λ
преобразования Бокса–Кокса
подбирается вручную

Литература

1. Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice (2nd edition). OTexts: Melbourne, Australia, 2018.
<https://otexts.com/fpp2/>
2. Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice (3rd edition). OTexts: Melbourne, Australia, 2021.
<https://otexts.com/fpp3/>
3. Сильвер Н. Сигнал и Шум. Почему одни прогнозы сбываются, а другие – нет. М.: Колибри, 2015. 608 с.

Сезонность

Data	Minute	Hour	Day	Week	Year
Quarters					4
Months					12
Weeks					52
Days				7	365.25
Hours			24	168	8 766
Minutes		60	1 440	10 080	525 960
Seconds	60	3 600	86 400	604 800	31 557 600

В году
52.18
недель