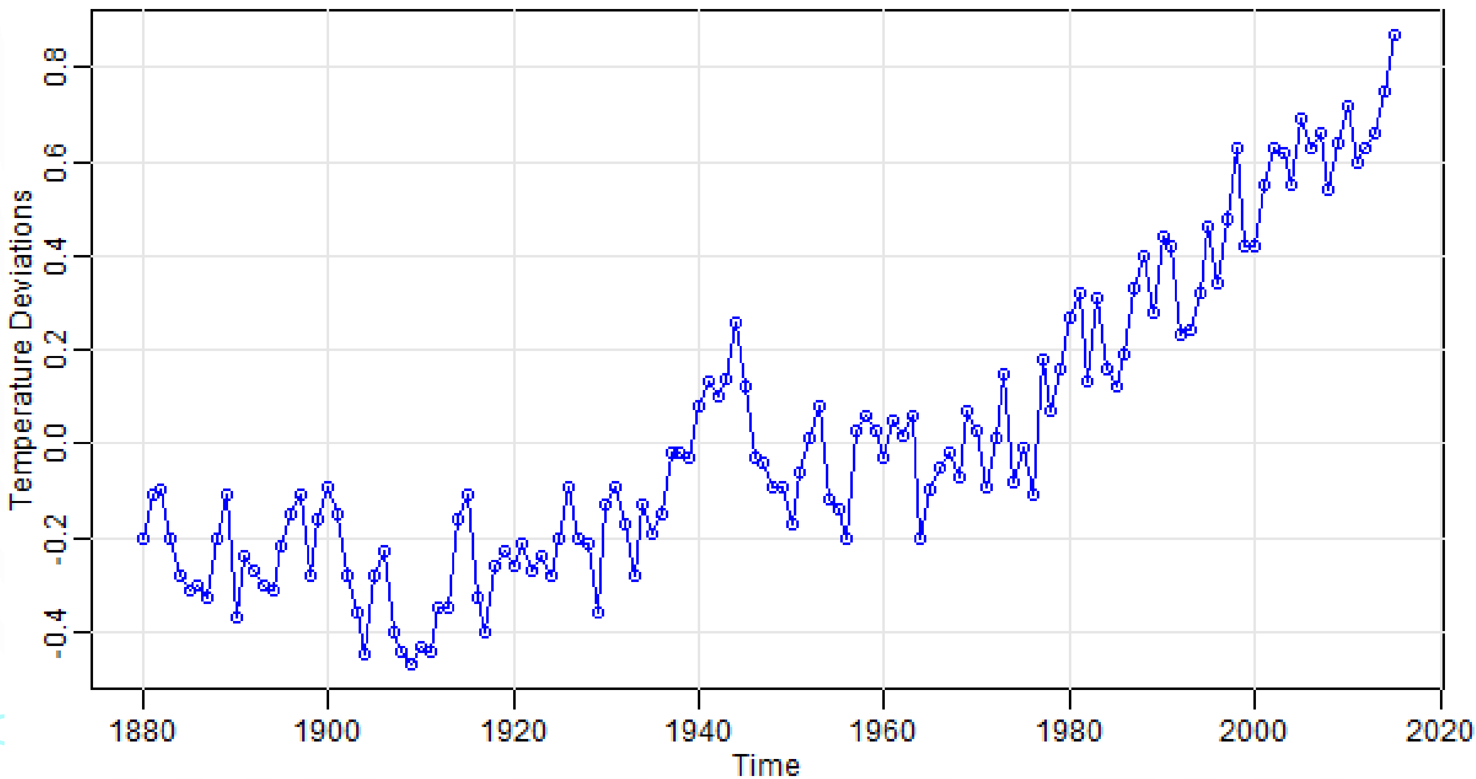
Временные ряды

Кантонистова Е.О.

ВШЭ, 2019

ВРЕМЕННОЙ РЯД

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



ЗАДАЧА ПРОГНОЗИРОВАНИЯ

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

<u>Задача:</u> построить функцию

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

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

<u>Метод наименьших квадратов:</u>

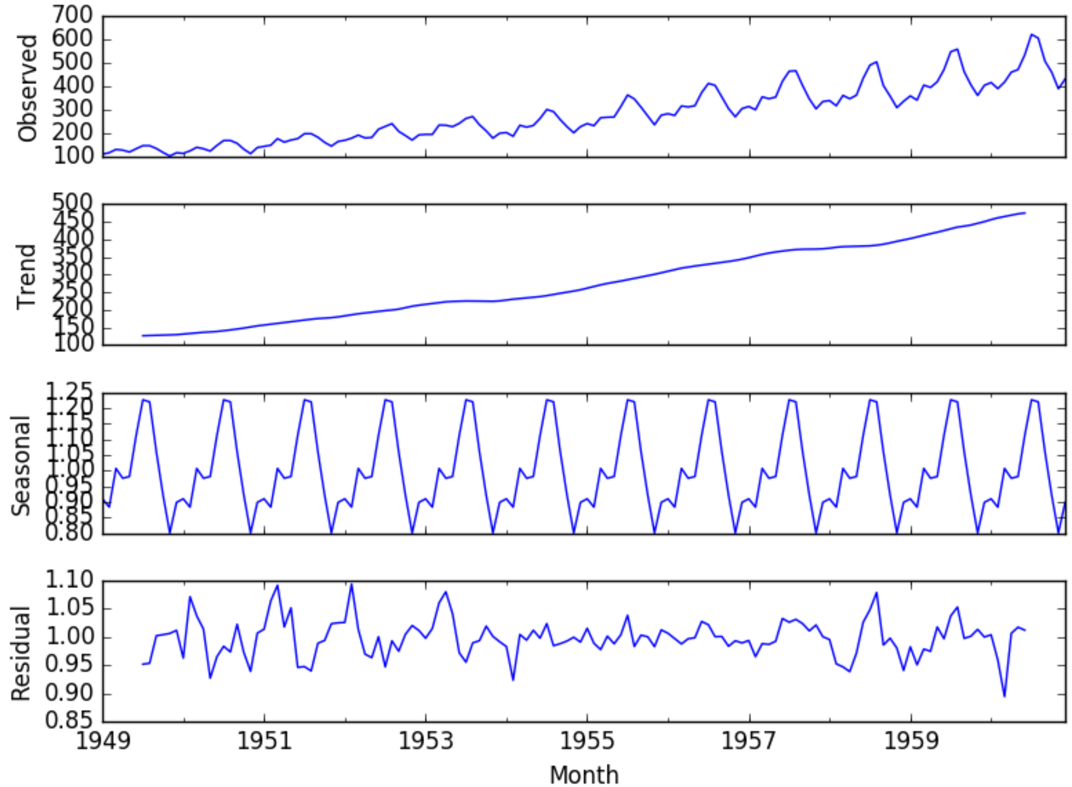
$$Q_t(w) = \sum_{i=t_0}^{t} (\hat{y}_i(w) - y_i)^2 \to \min_w$$

ОСОБЕННОСТЬ 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>*

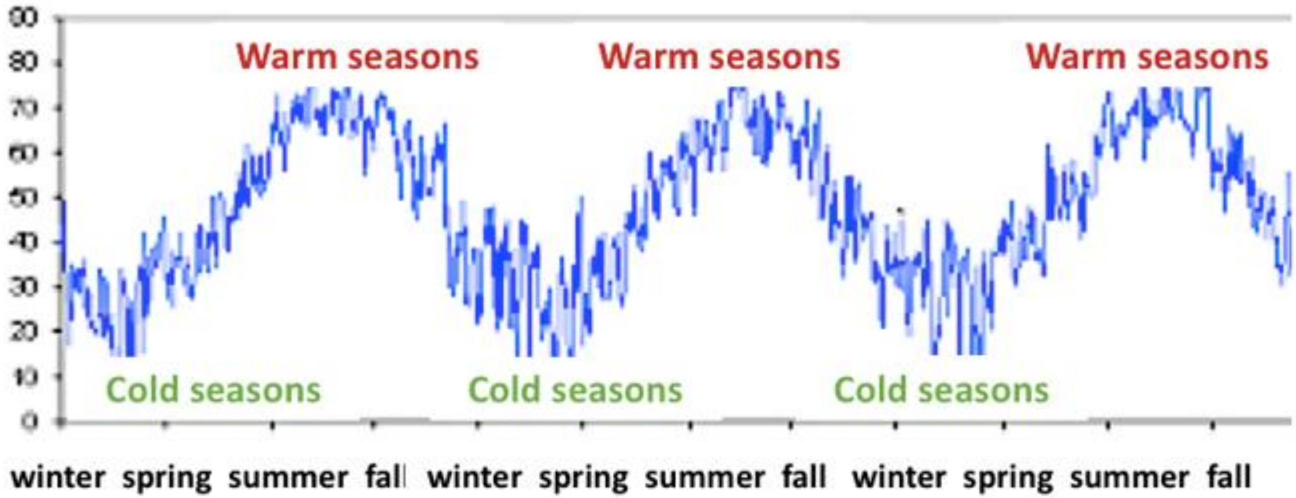
КОМПОНЕНТЫ ВРЕМЕННОГО РЯДА

- *Тренд* – 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

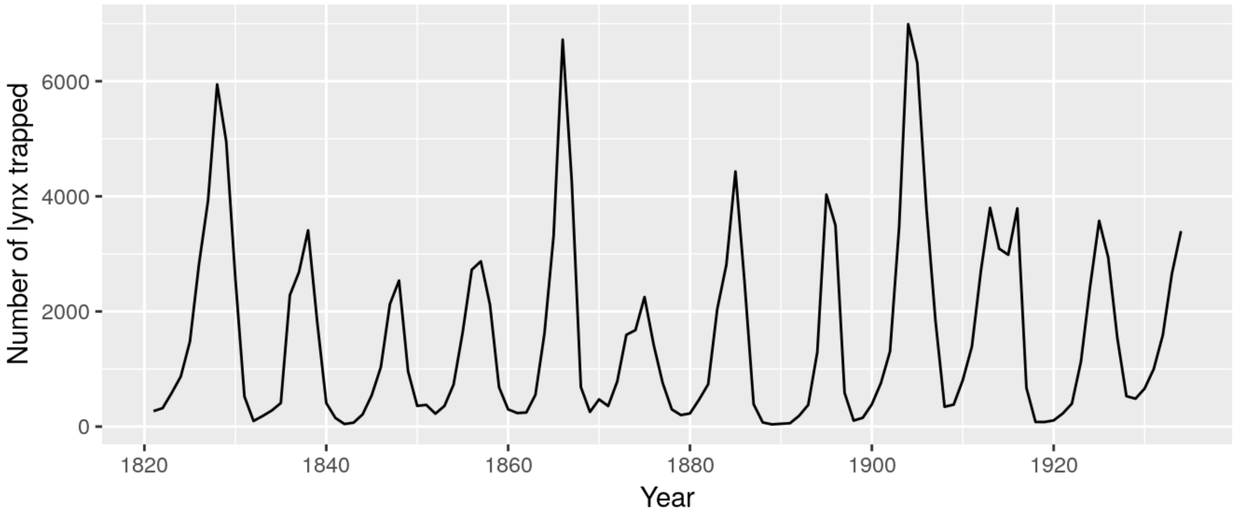


ЦИКЛЫ И СЕЗОННОСТЬ

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

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

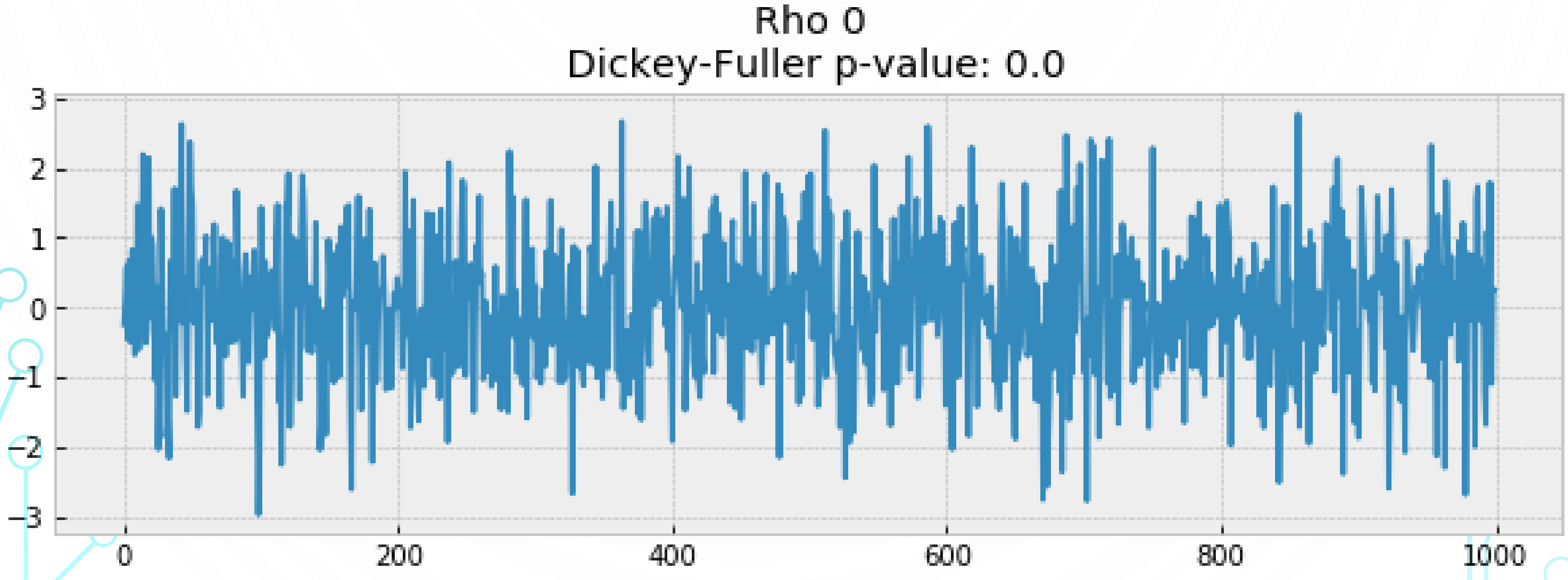
Для проверки стационарности ряда можно использовать критерий Дики-Фуллера.

ЕДИНИЧНЫЙ КОРЕНЬ

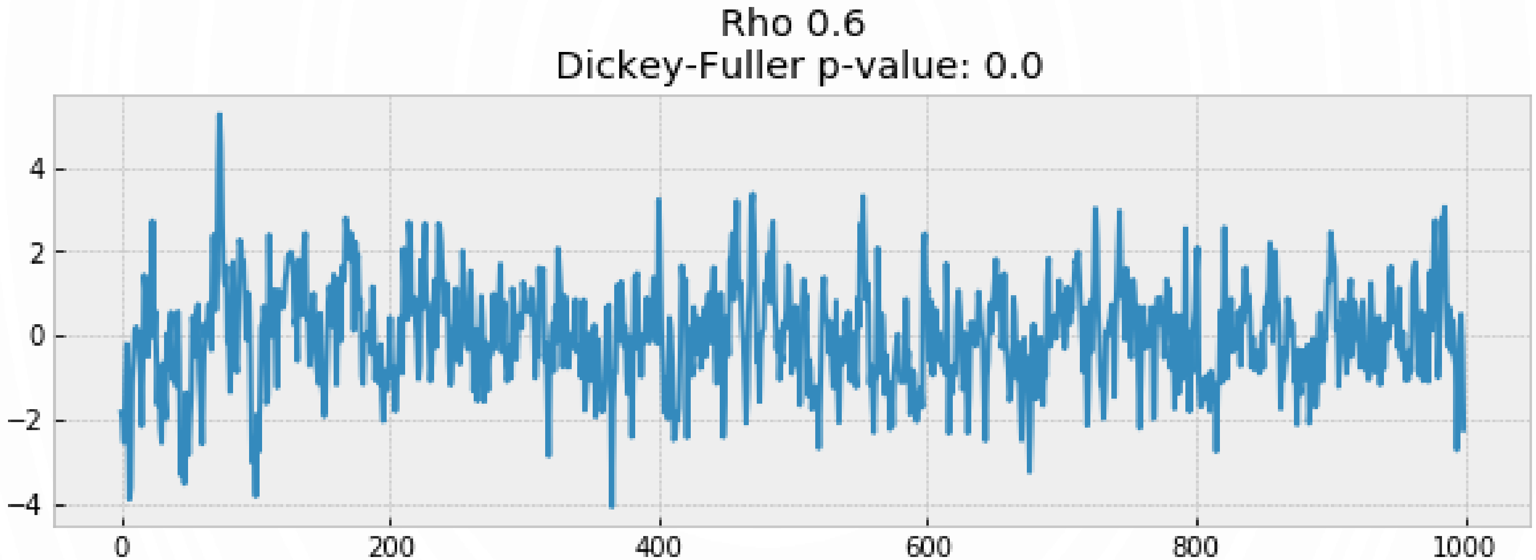
Рассмотрим модель временного ряда $\boldsymbol{X_t = \rho \cdot X_{t-1} + \varepsilon_t}$,

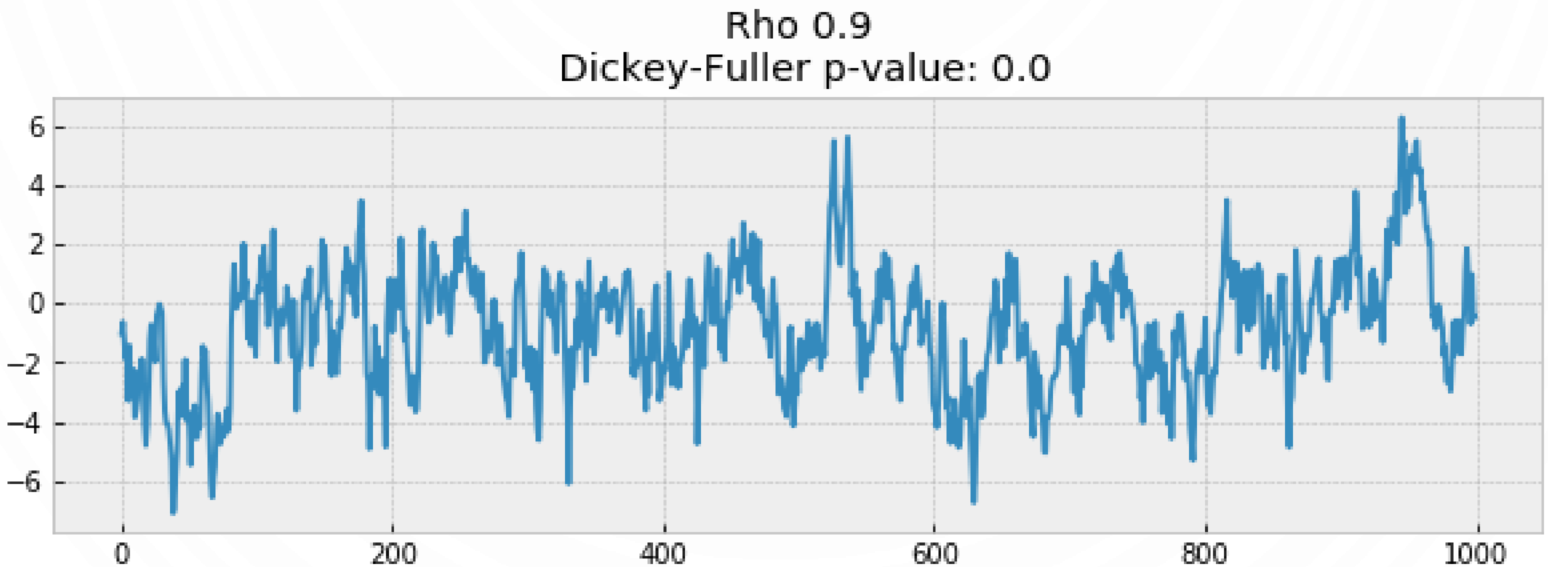
где ε_t - ошибка, не зависящая от значений временного ряда.

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

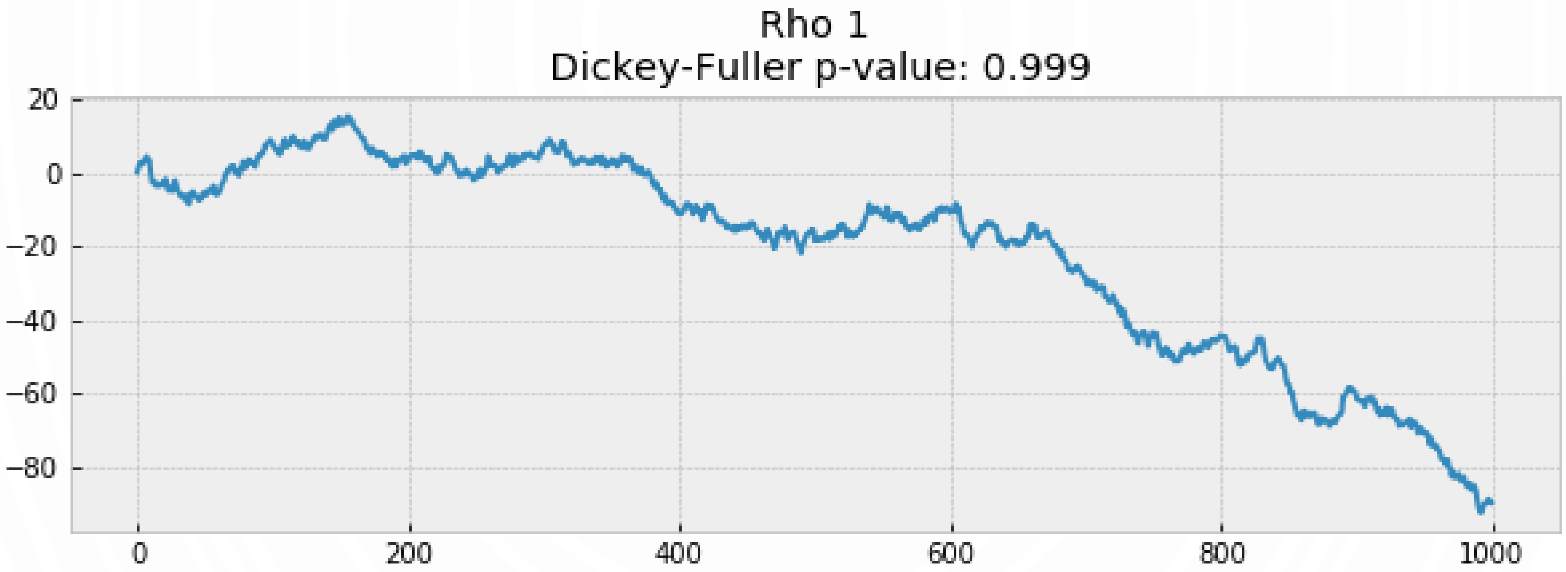


ЕДИНИЧНЫЙ КОРЕНЬ





ЕДИНИЧНЫЙ КОРЕНЬ



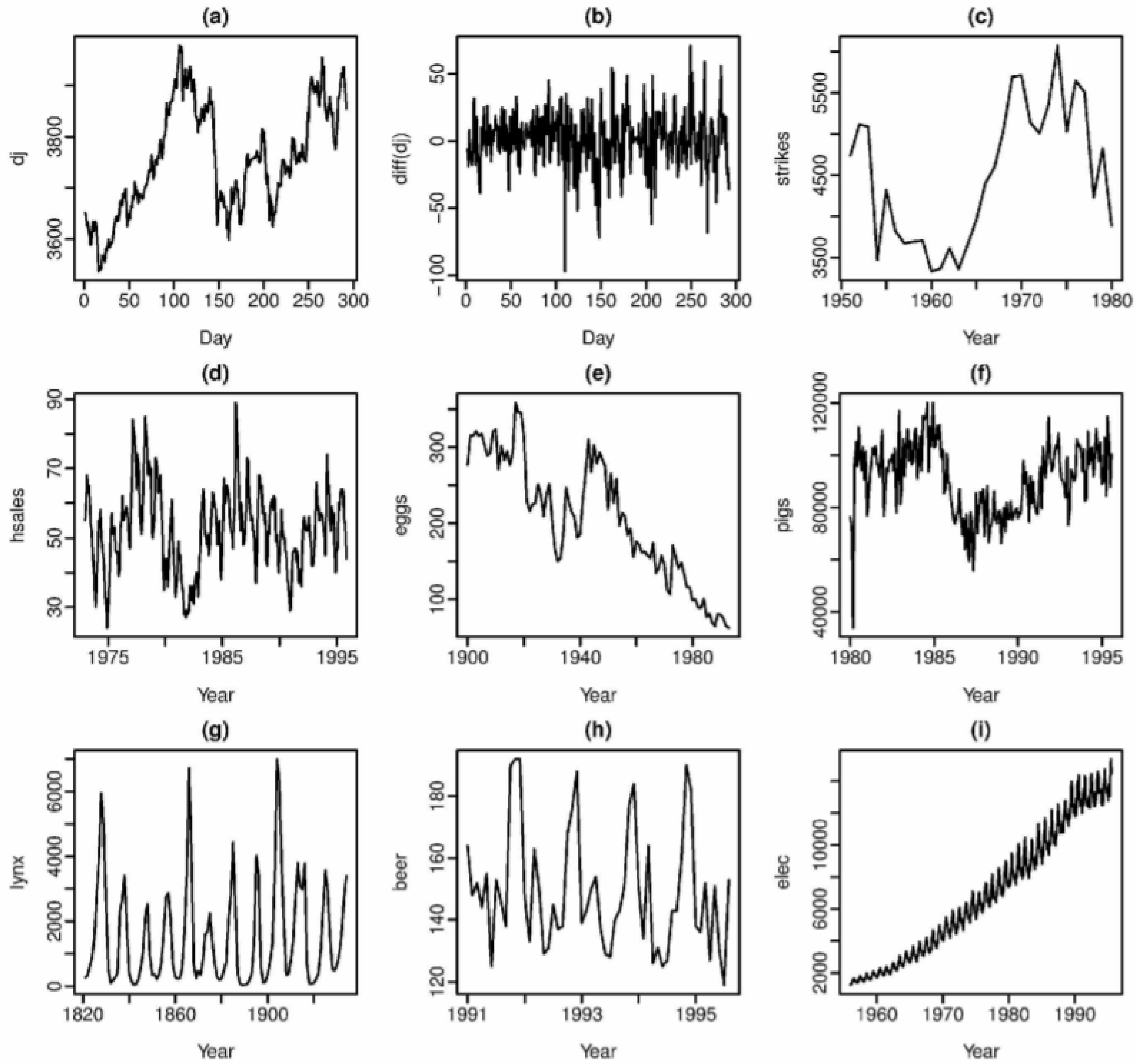
Видно, что при $\rho = 1$ процесс не возвращается к своему среднему, а значит, не является стационарным.

Проверка стационарности ряда

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

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

ПРИМЕРЫ ВРЕМЕННЫХ РЯДОВ



АВТОКОРРЕЛЯЦИЯ

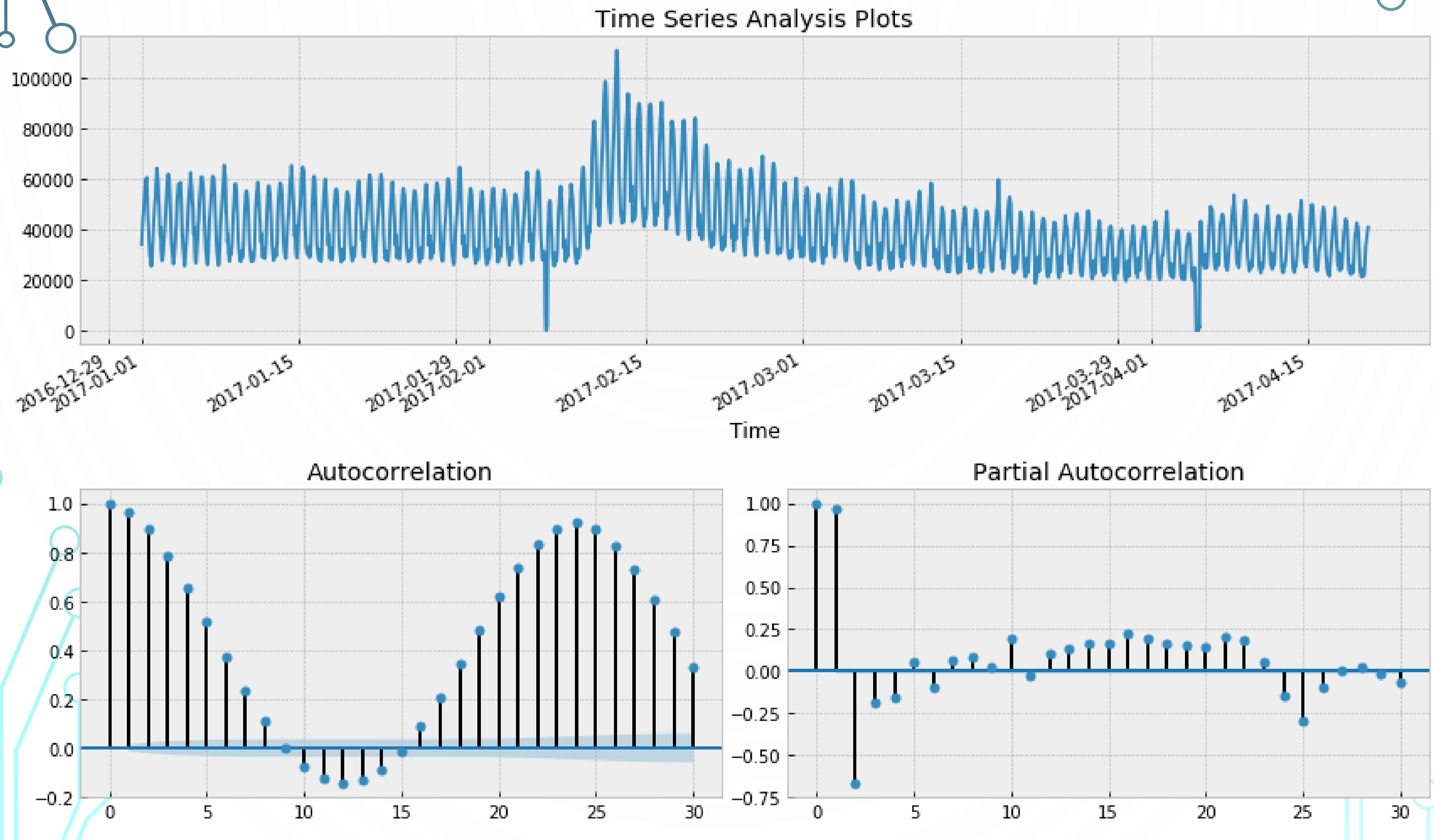
При наличии во временном ряде тренда и сезонных колебаний значения любого последующего элемента ряда зависят от предыдущих.

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)***.

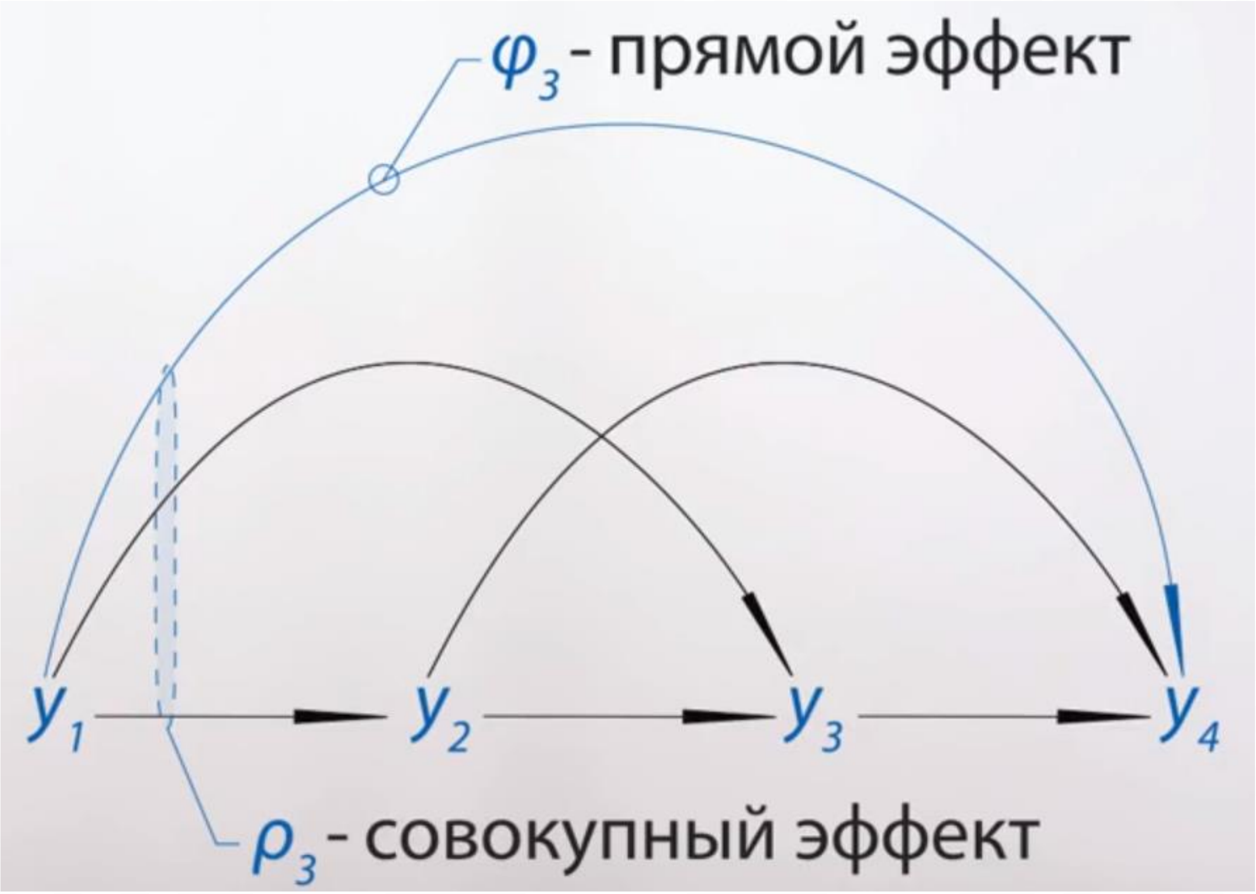


АВТОКОРРЕЛЯЦИОННАЯ ФУНКЦИЯ

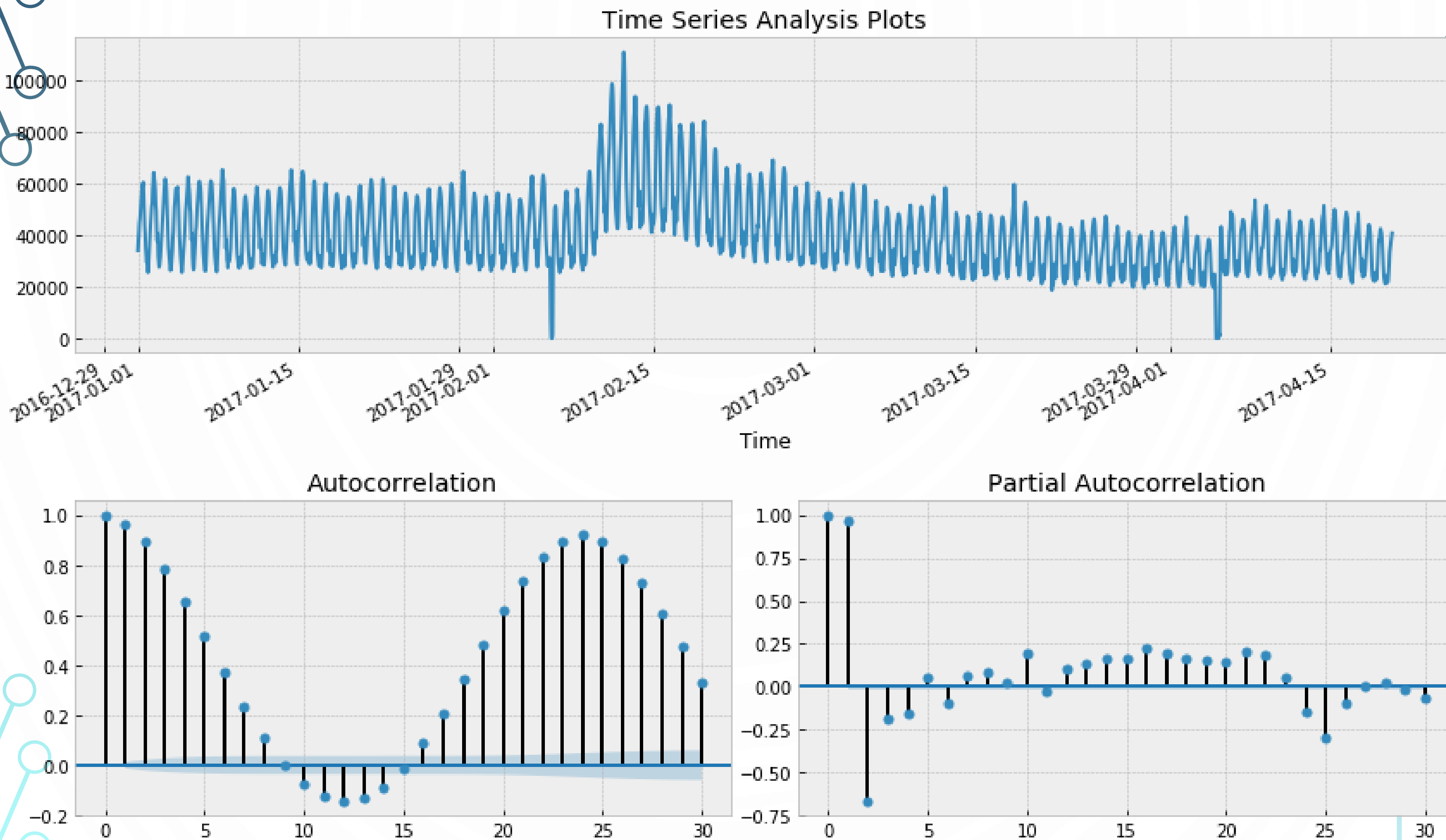
Функция автокорреляции имеет следующий смысл. Изначально два ряда полностью выровнены (смещение 0) и корреляция равна 1. Далее, **когда мы начинаем сдвигать ряды, они постепенно «рассинхронизируются», значение корреляции падает**. Скорость падения говорит нам о том, насколько хорошо ряд «помнит» свои предыдущие значения**. Если корреляция падает быстро с ростом смещения, ряд быстро «забывает» предыдущие значения. Если корреляция падает медленно, мы имеем дело с относительно устойчивым процессом во времени.** Возможны ситуации, когда функция быстро спадает, а затем снова растёт, образуя один или несколько пиков. Это означает, что ряды вновь начинают совпадать, если их сместить друг относительно друга на достаточное смещение по времени. То есть исходный ряд демонстрирует периодичность (сезонность). Количество шагов по времени, которому соответствует пик функции автокорреляции, соответствует периоду.

ЧАСТНАЯ АВТОКОРРЕЛЯЦИЯ

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



ЧАСТНАЯ АВТОКОРРЕЛЯЦИЯ



МЕТОДЫ ИЗБАВЛЕНИЯ ОТ НЕСТАЦИОНАРНОСТИ

1. Стабилизация дисперсии

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

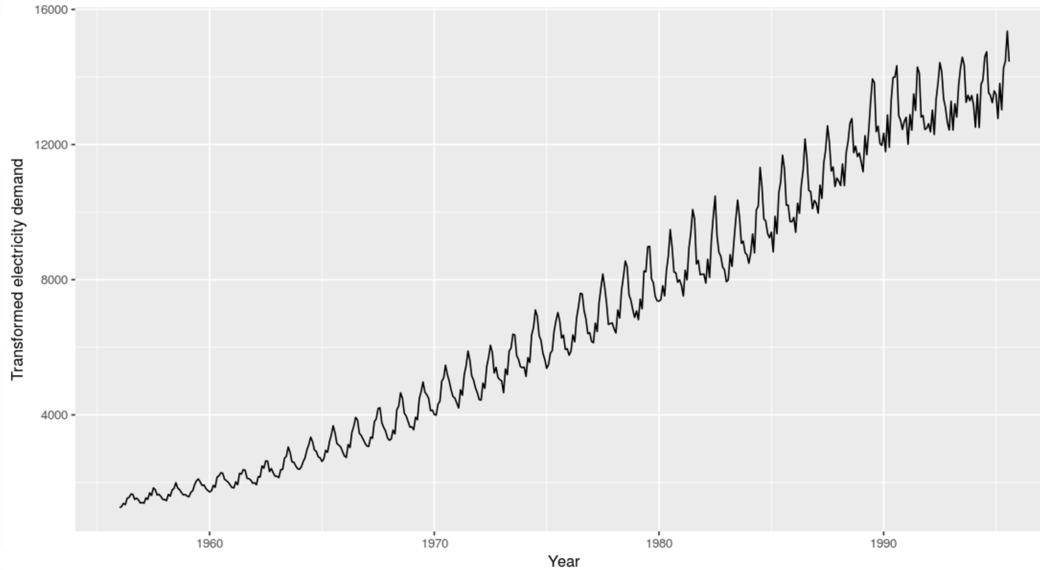
$$y_t' = \begin{cases} \ln y_t, \lambda = 0 \\ \dfrac{y_t^\lambda - 1}{\lambda}, \lambda \neq 0 \end{cases}$$

- логарифмирование – частный случай

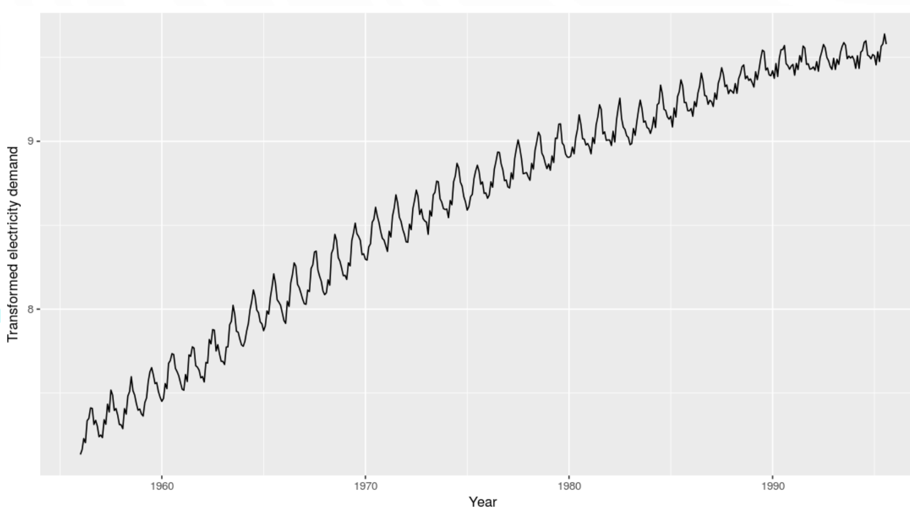
Параметр λ подбирается так, чтобы сделать дисперсию как можно более однородной.

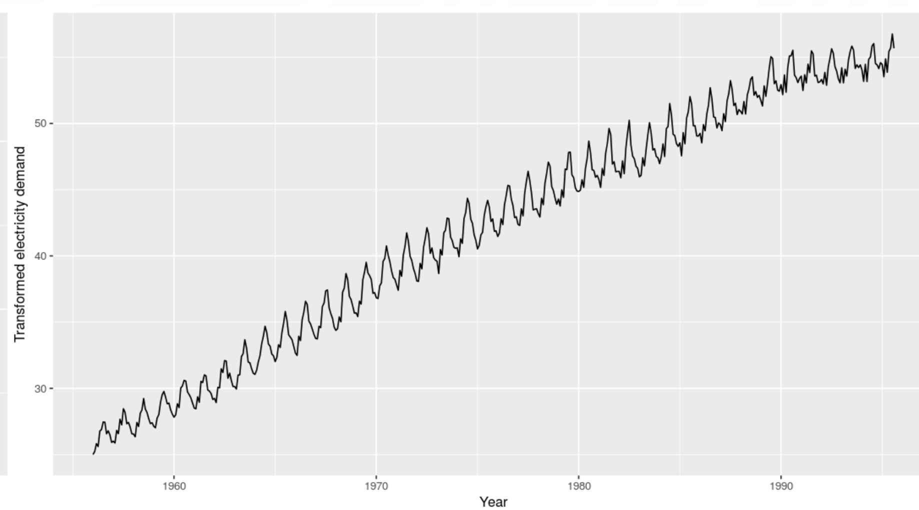
СТАБИЛИЗАЦИЯ ДИСПЕРСИИ (ПРИМЕР)

Исходный ряд:



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



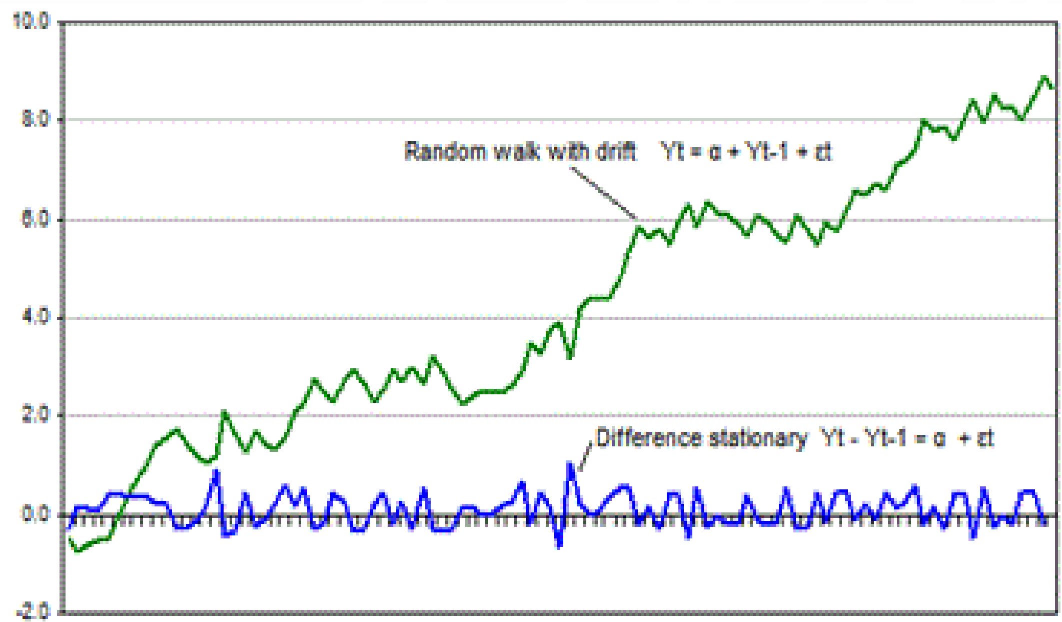


МЕТОДЫ ИЗБАВЛЕНИЯ ОТ НЕСТАЦИОНАРНОСТИ

2. <u>Дифференцирование</u> – 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

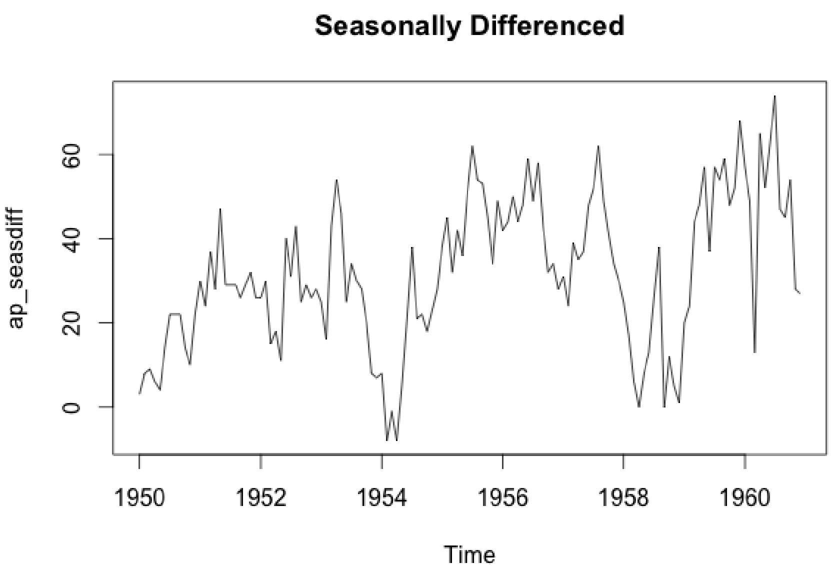


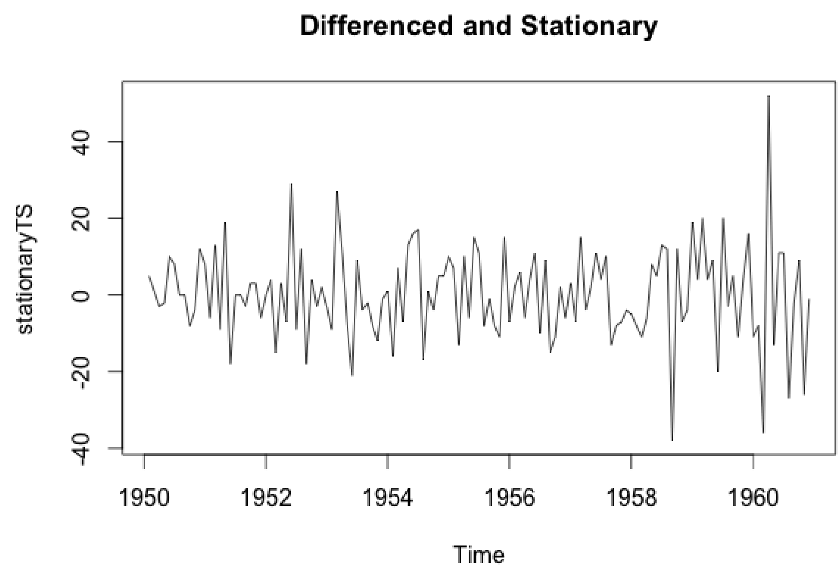
МЕТОДЫ ИЗБАВЛЕНИЯ ОТ НЕСТАЦИОНАРНОСТИ

3. <u>Сезонное дифференцирование</u> – переход к попарным разностям значений в соседних сезонах

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

- убирает сезонность





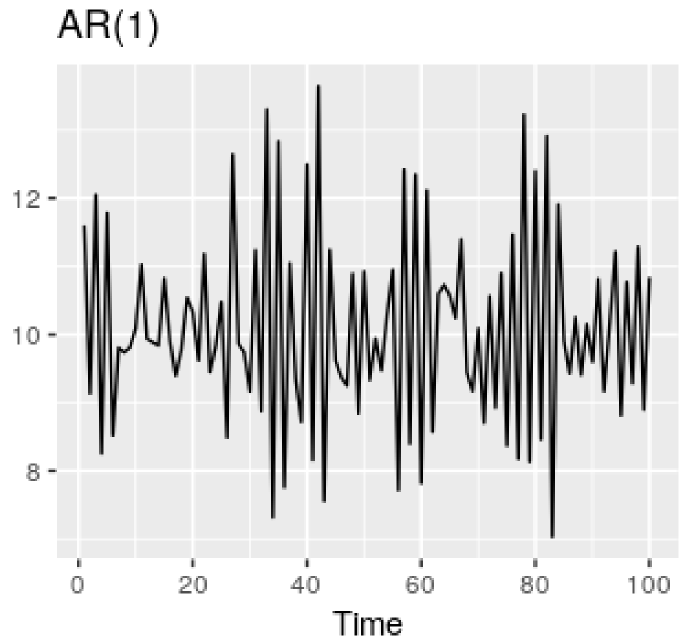
Сезонное дифференцирование лучше применять в начале – возможно, после него ряд уже станет стационарным.

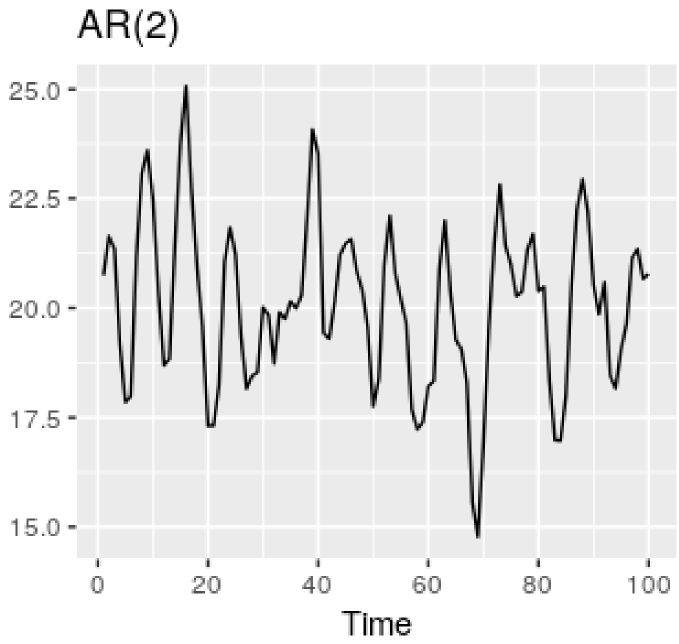
МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

- Модель авторегрессии $AR(p)$:

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

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



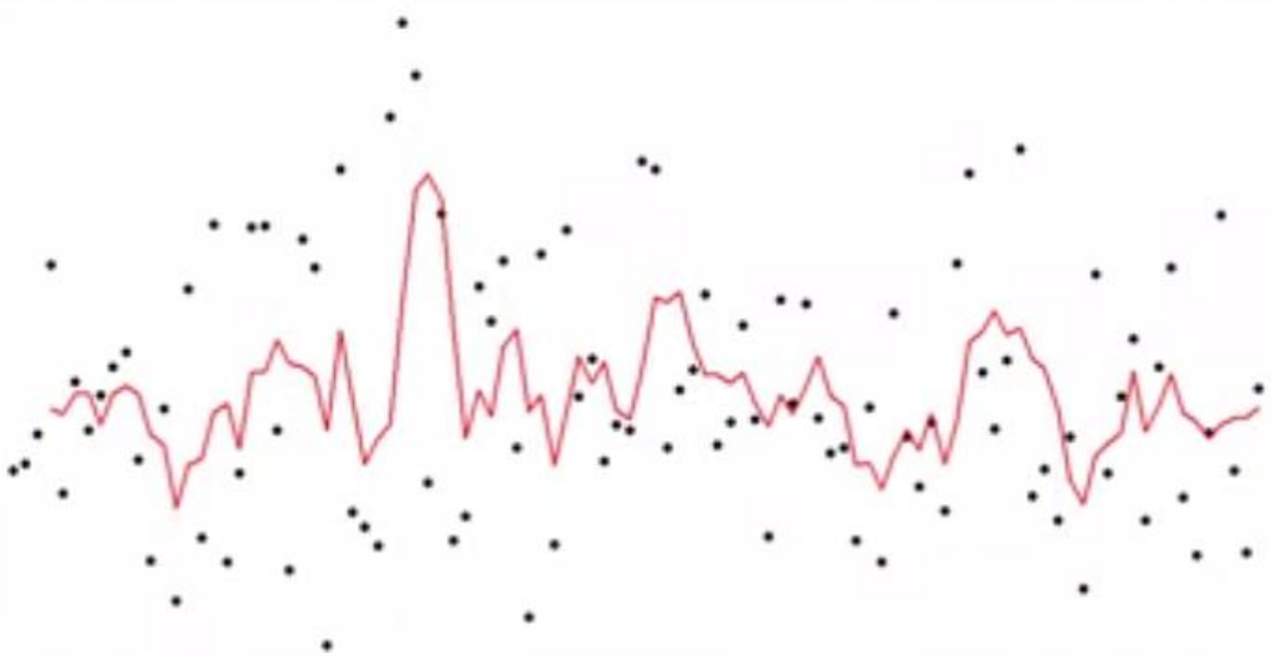


МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

- Процесс скользящего среднего порядка q $\left(MA(q)\right)$:

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

$b_q \neq 0, \varepsilon_t$ - 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.



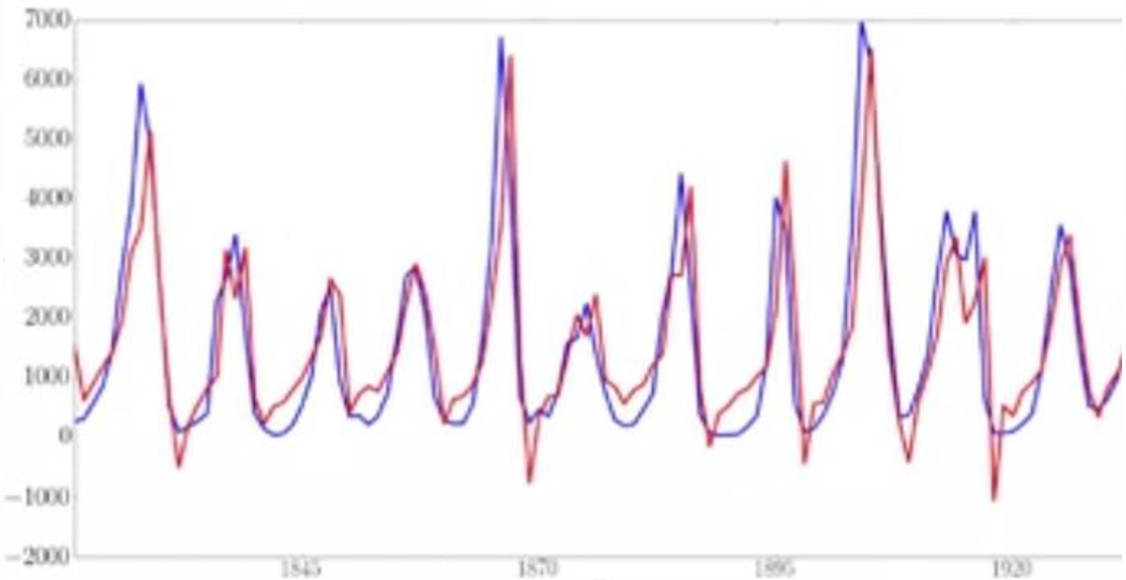
МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

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

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

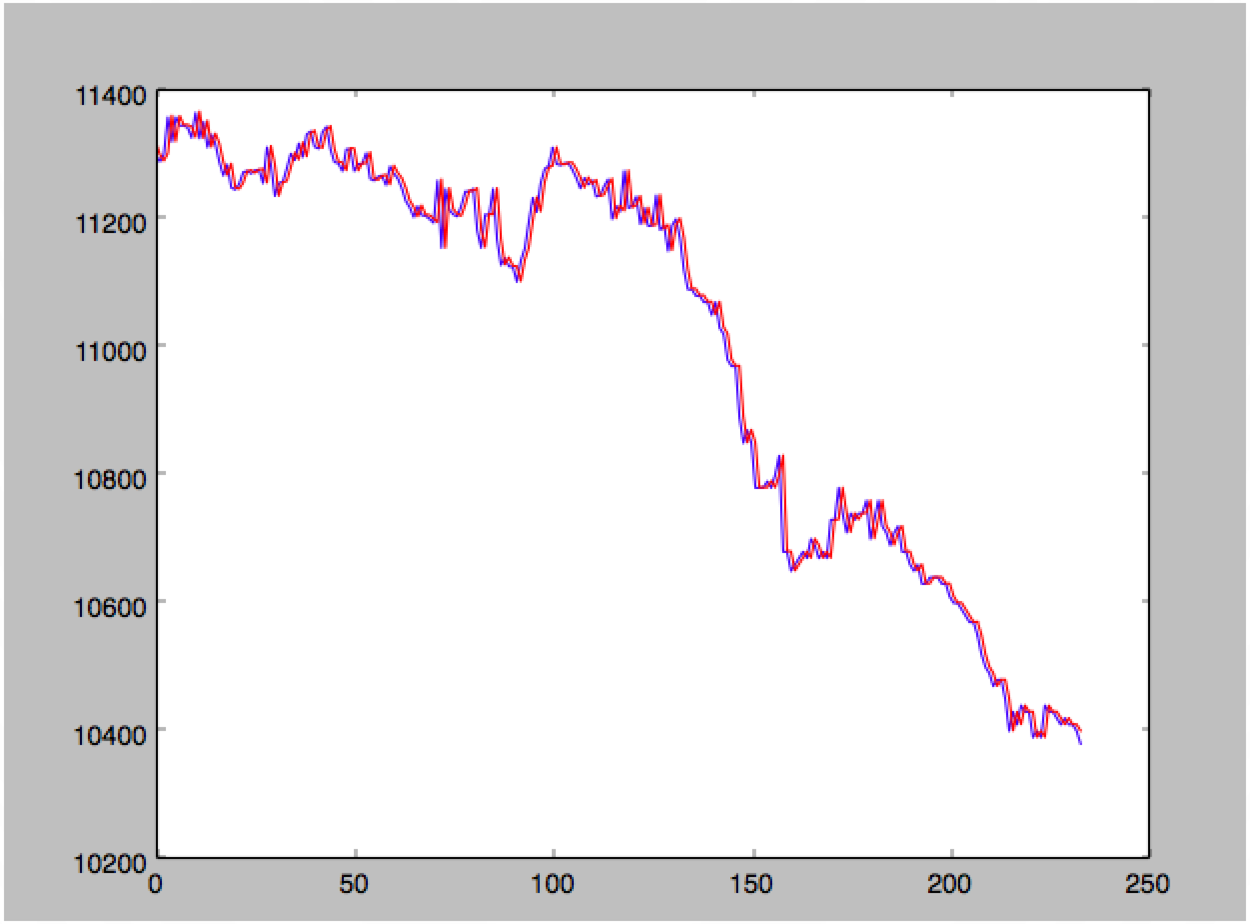
$a_p, b_q \neq 0.$

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



МОДЕЛИ ВРЕМЕННЫХ РЯДОВ – ЭКОНОМЕТРИЧЕСКИЙ ПОДХОД

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



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.

ЛИНЕЙНАЯ МОДЕЛЬ АВТОРЕГРЕССИИ

- признаки - n предыдущих наблюдений ряда:

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

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

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

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

АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

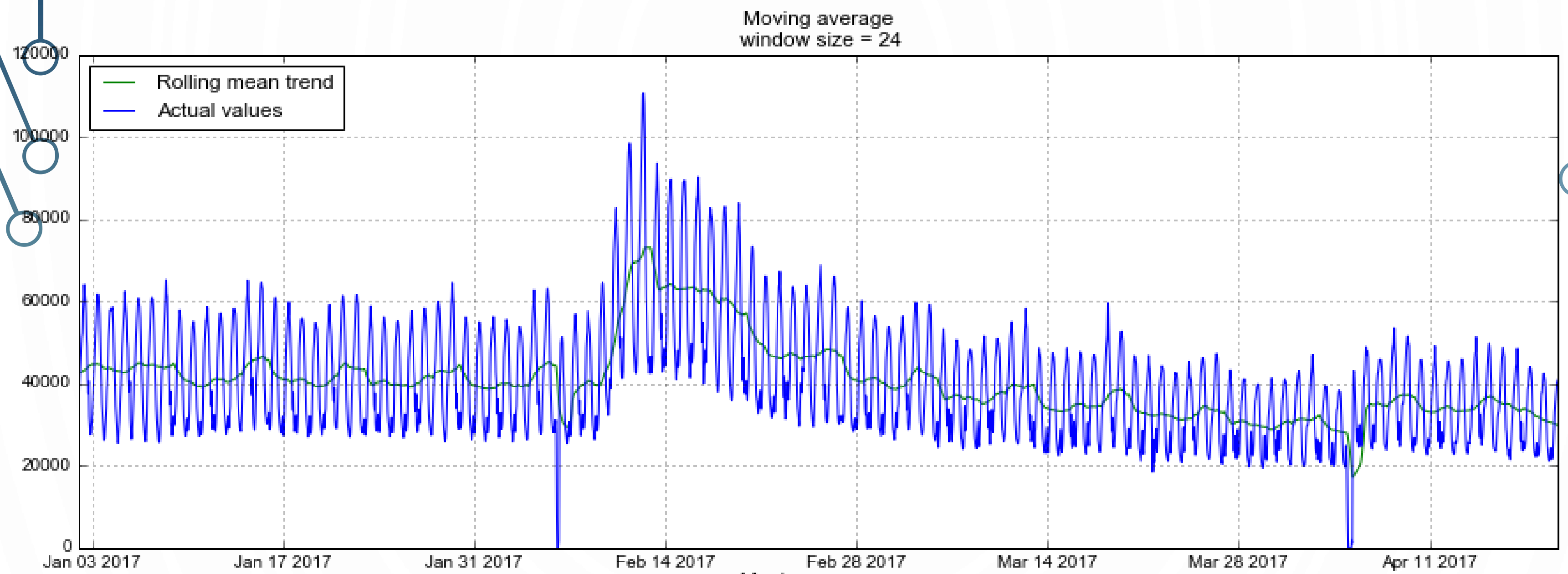
СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

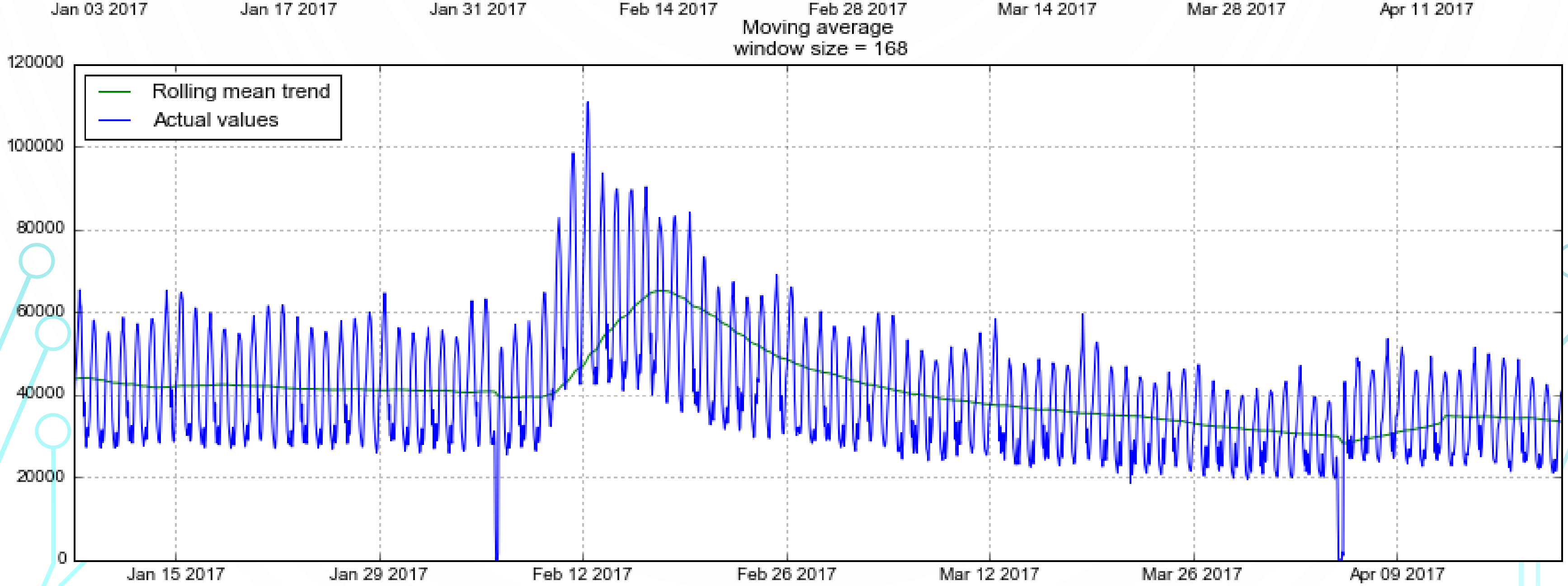
$$\hat{y}_{t+1} = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-i}$$

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

+ *сглаживает данные*

СКОЛЬЗЯЩЕЕ СРЕДНЕЕ





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:

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

Экспоненциальное скользящее среднее

- Простейшая регрессионная модель – константа:

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

Минимизируем квадратичную ошибку:

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

<u>Аналитическое решение (формула Надарая-Ватсона):</u>

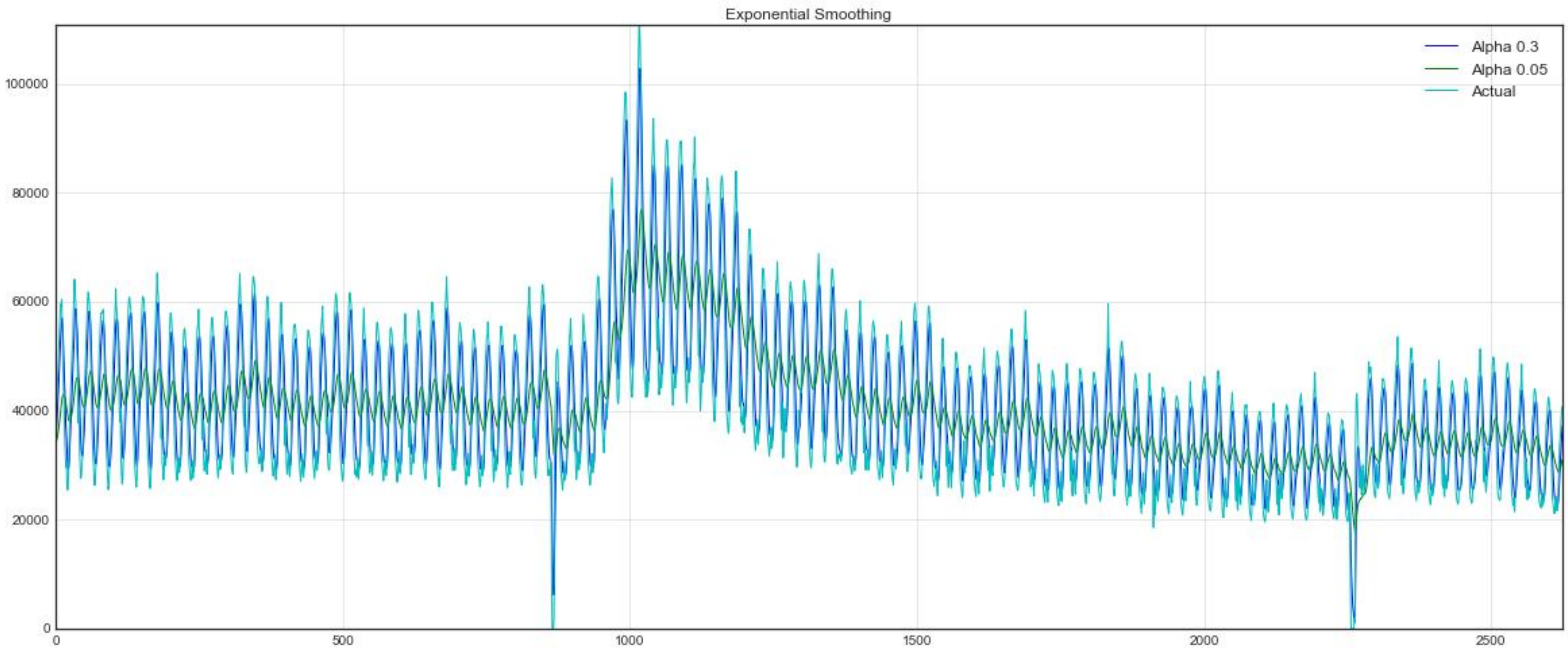
$$\hat{y}_{t+1} = \frac{\sum_{i=0}^{t} \beta^i y_{t-i}}{\sum_{i=0}^{t} \beta^i}$$

ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ (ЭСС)

Утверждение. Модель ЭСС можно записать в виде

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

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



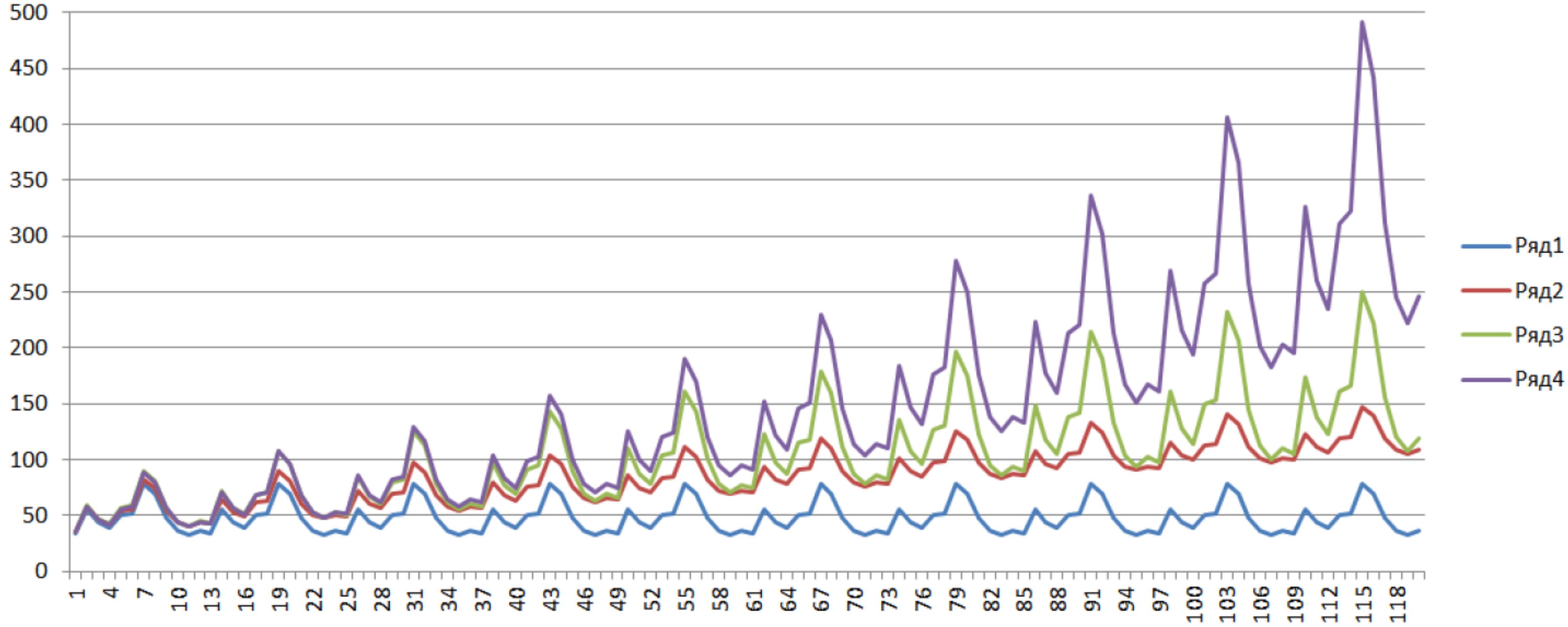
Экспоненциальное скользящее среднее (ЭСС)

Оптимальное значение α подбираем по скользящему контролю:

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

- 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

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

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

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

$a_t + b_t d$ – 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

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

$\theta_0, \dots, \theta_{s-1}$ - 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

<u>Модель Уинтерса с линейным трендом</u>

- модель мультипликативной сезонности периода s с линейным трендом

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

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

$\theta_0, \ldots, \theta_{s-1}$ - 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

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

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

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

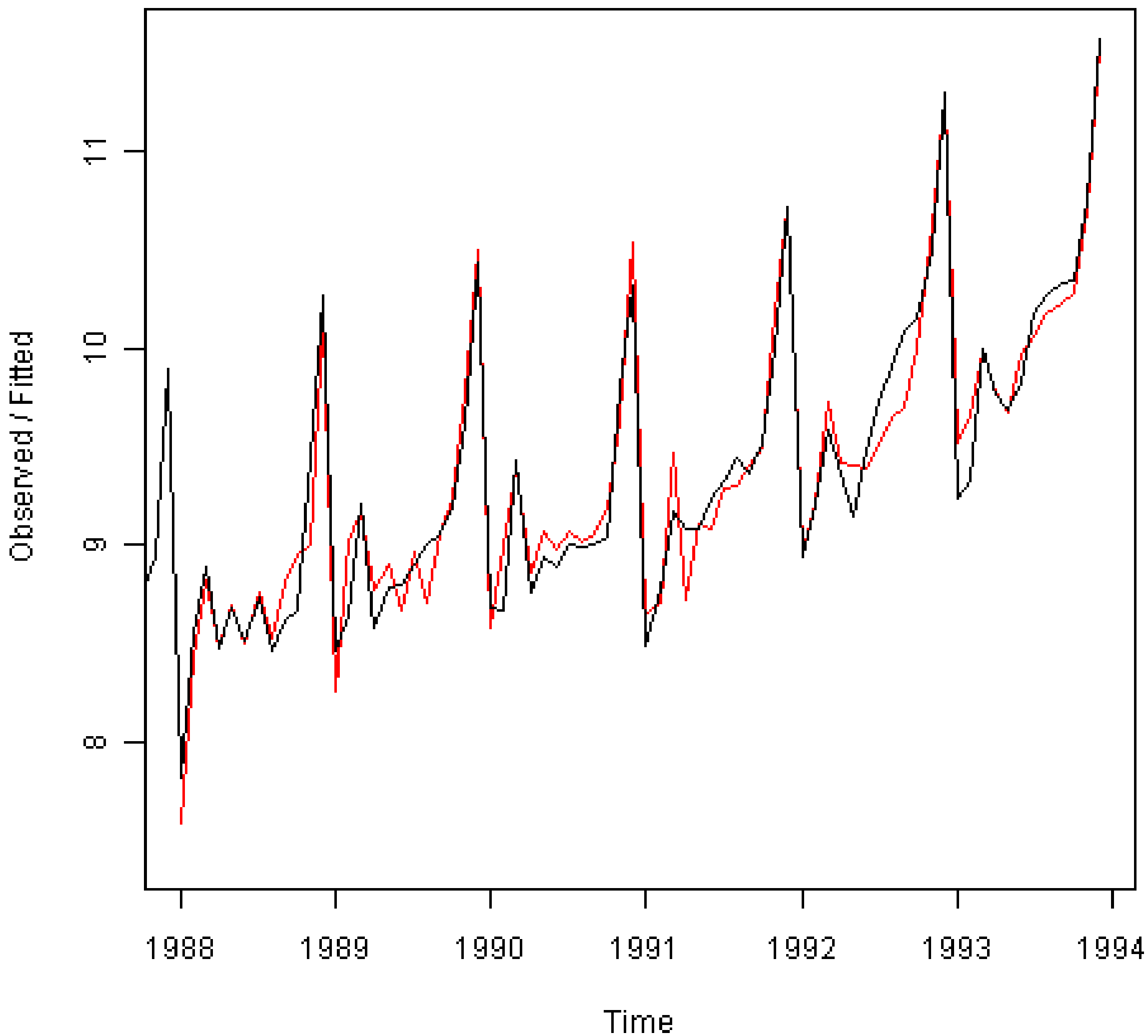
формулы для a_t, b_t, θ_t:

$$a_t = \alpha_1(y_t/\theta_{t-s}) + (1-\alpha_1)a_{t-1}r_{t-1}$$
$$r_t = \alpha_2(a_t/a_{t-1}) + (1-\alpha_2)r_{t-1}$$
$$\theta_t = \alpha_3(y_t/a_t) + (1-\alpha_3)\theta_{t-s},$$

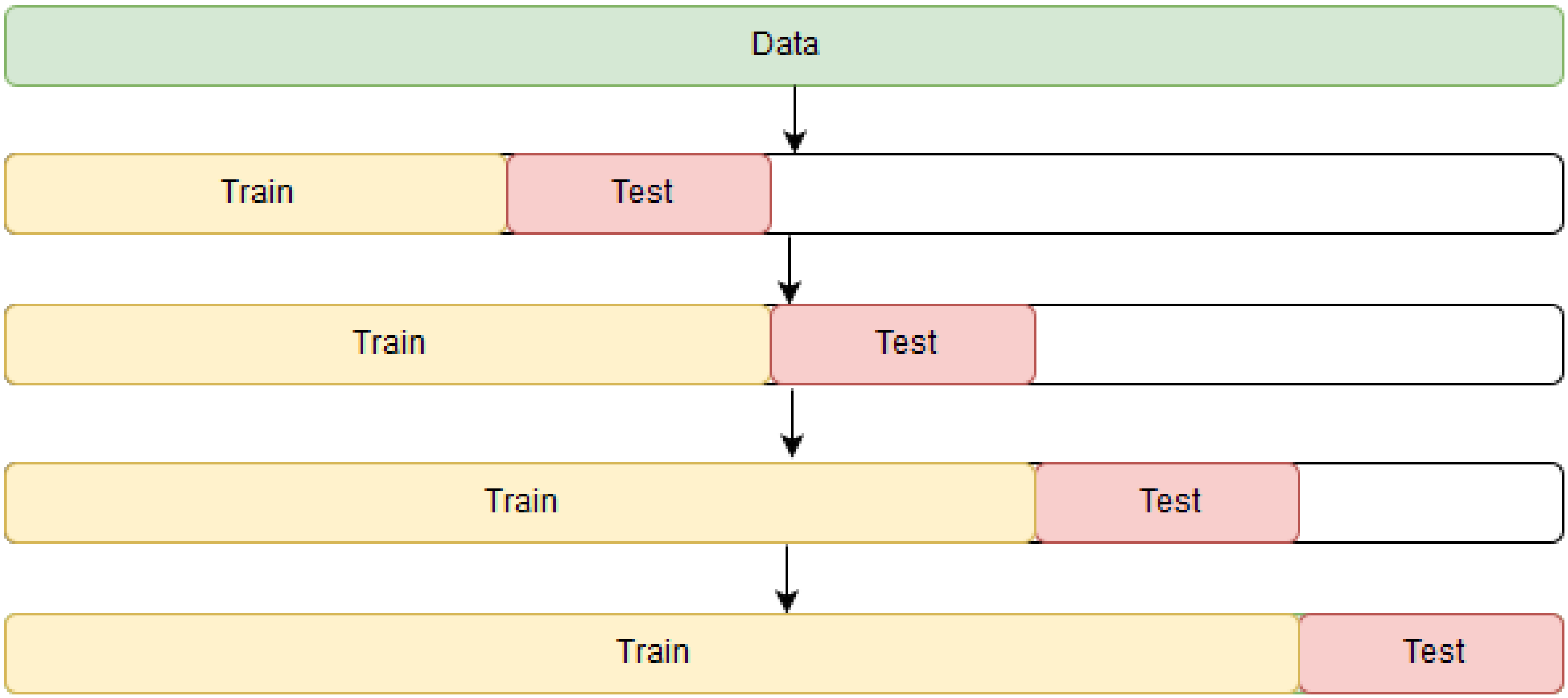
где $\alpha_1, \alpha_2, \alpha_3$ - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса с линейным трендом (модель Хольта-Уинтерса)



КРОСС-ВАЛИДАЦИЯ НА ВРЕМЕННЫХ РЯДАХ



ВРЕМЕННЫЕ РЯДЫ, ЧАСТЬ 2

АДАПТИВНАЯ МОДЕЛЬ АВТОРЕГРЕССИИ

Линейная модель авторегрессии (линейный фильтр):

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

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

Метод наименьших квадратов: $\varepsilon_t^2 \to \min\limits_{w}$

Один шаг градиентного спуска в момент времени t:

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

Градиентный шаг:

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

α – аналог параметра сглаживания.

ОЦЕНКА АДЕКВАДТНОСТИ АДАПТИВНОЙ МОДЕЛИ

Следящий скользящий контроль:

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

$$\boldsymbol{K_t = \frac{\varepsilon'_t}{\varepsilon^*_t}},$$

- $\varepsilon'_t = \gamma\varepsilon_t + (1-\gamma)\varepsilon'_{t-1}$ - ЭСС ошибки
- $\varepsilon^*_t = \gamma|\varepsilon_t| + (1-\gamma)\varepsilon^*_{t-1}$ - ЭСС модуля ошибки

Обычно берут $\gamma \in (0.05, 0.1)$

ОЦЕНКА АДЕКВАТНОСТИ АДАПТИВНОЙ МОДЕЛИ

Следящий скользящий контроль:

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

$$\boldsymbol{K_t = \frac{\varepsilon'_t}{\varepsilon^*_t}},$$

Статистический тест адекватности:

гипотеза $H_0\colon \mathrm{E}\varepsilon_t = 0, E\varepsilon_t\varepsilon_{t+d} = 0$

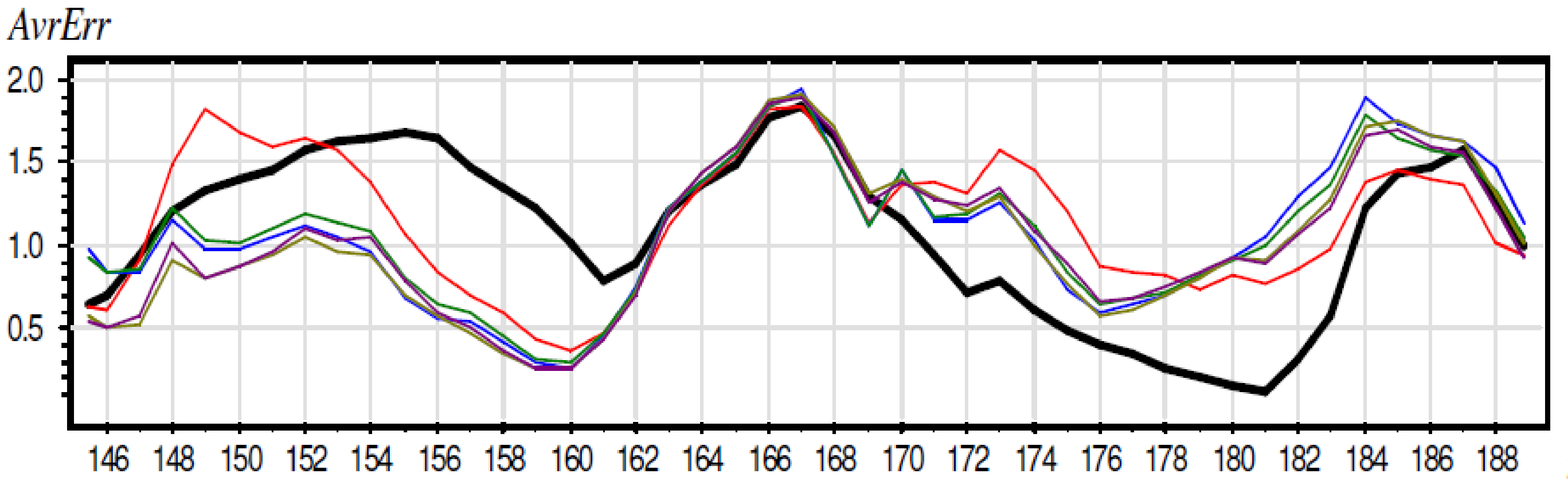
принимается на уровне значимости α, если

$$|K_t| \leq 1.2\Phi_{1-\frac{\alpha}{2}}\sqrt{\gamma/(2-\gamma)},$$

где $\Phi_{1-\alpha/2}$ - квантиль нормального распределения.

АДАПТИВНАЯ СЕЛЕКЦИЯ МОДЕЛЕЙ

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



Адаптивная селекция моделей

Дано: k моделей прогнозирования

- $\hat{y}_{j,t+d}$ - прогноз j-й модели в момент $t+d$
- $\varepsilon_{j,t} = y_t - \hat{y}_{j,t}$ - ошибка прогноза j-й модели в момент времени t
- $\varepsilon^*_{j,t} = \gamma\left|\varepsilon_{j,t}\right| + (1-\gamma)\varepsilon^*_{j,t-1}$ - ЭСС модуля ошибки

Лучшая модель в момент времени t:

$$j^*_t = \operatorname*{argmin}_{j=1,\dots,k} \varepsilon^*_{j,t}$$

Адаптивная селективная модель – прогноз по лучшей модели:

$$\hat{y}_{j,t+d} := \hat{y}_{j^*_t,t+d}$$

Как правило, $\gamma \in (0.01, 0.1)$

АДАПТИВНАЯ КОМПОЗИЦИЯ МОДЕЛЕЙ

Дано: k моделей прогнозирования

- $\hat{y}_{j,t+d}$ - прогноз j-й модели в момент $t+d$
- $\varepsilon_{j,t} = y_t - \hat{y}_{j,t}$ - ошибка прогноза j-й модели в момент времени t
- $\varepsilon^*_{j,t} = \gamma\left|\varepsilon_{j,t}\right| + (1-\gamma)\varepsilon^*_{j,t-1}$ - ЭСС модуля ошибки

Линейная комбинация моделей:

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

Адаптивный подбор весов (идея - обратно пропорционально ошибке модели): $w_{jt} = \left(\varepsilon^*_{j,t}\right)^{-1} / \sum_{s=1}^{k}\left(\varepsilon^*_{s,t}\right)^{-1}$

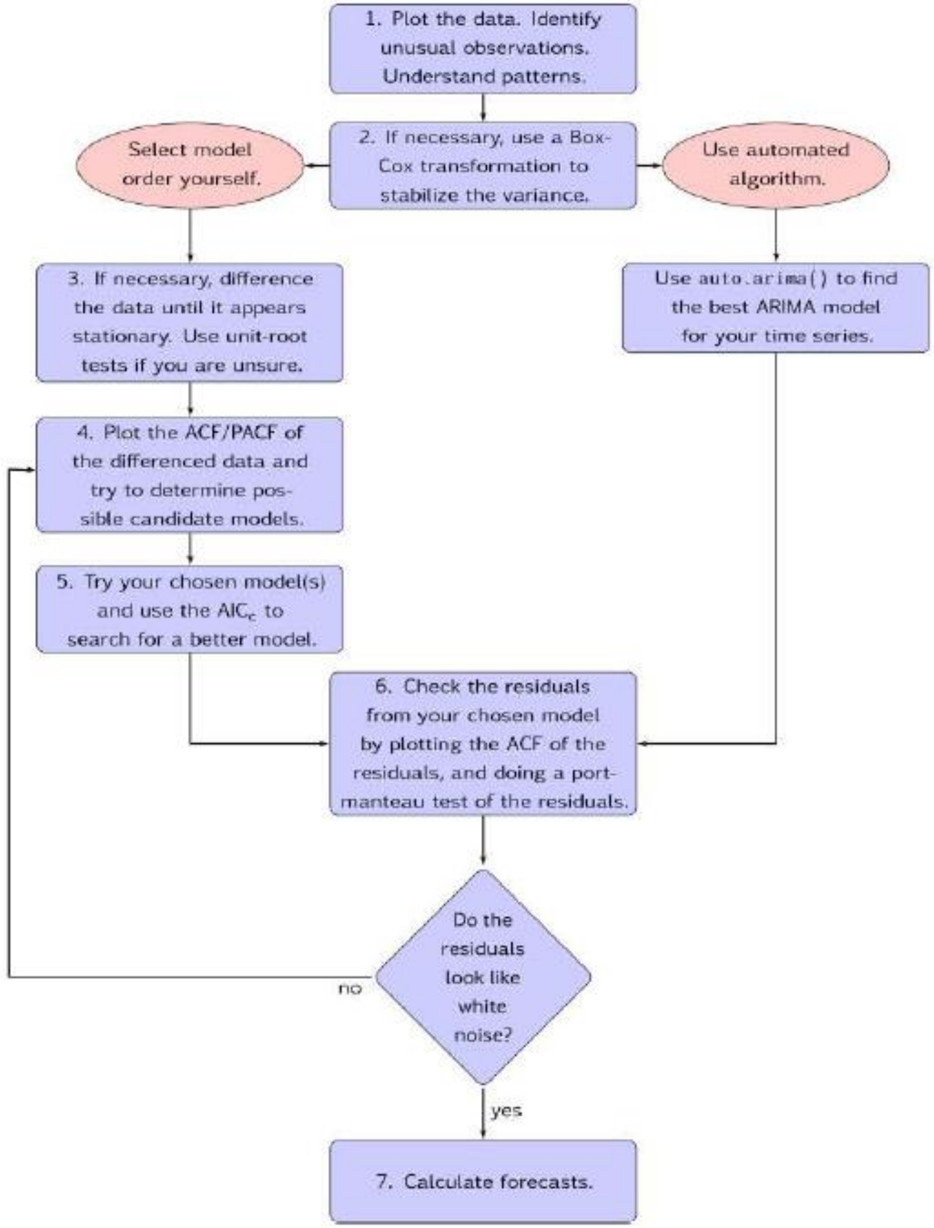
Требования к задаче прогнозирования временных рядов

- 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

МАШИННОЕ ОБУЧЕНИЕ: ПРИЗНАКИ

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

СХЕМА ПОЛУЧЕНИЯ ПРЕДСКАЗАНИЯ С ПОМОЩЬЮ ARIMA/SARIMA



ПОДБОР ПАРАМЕТРОВ ARIMA/SARIMA

Пусть мы получили предсказание с помощью некоторой модели, и остатки этой модели - $\{e_i\}_{i=1}^{T}$.

- Пусть $\sigma^2 = \frac{1}{T}\sum e_i^2$ - оценка дисперсии ошибок.

Пусть мы используем модель ARIMA(p,d,q).

<u>Критерий Акаике (AIC):</u>

$$\boldsymbol{AIC_c = ln\sigma^2 + \frac{2(p+q+1)(p+q+2)}{T-q-p-2}}$$

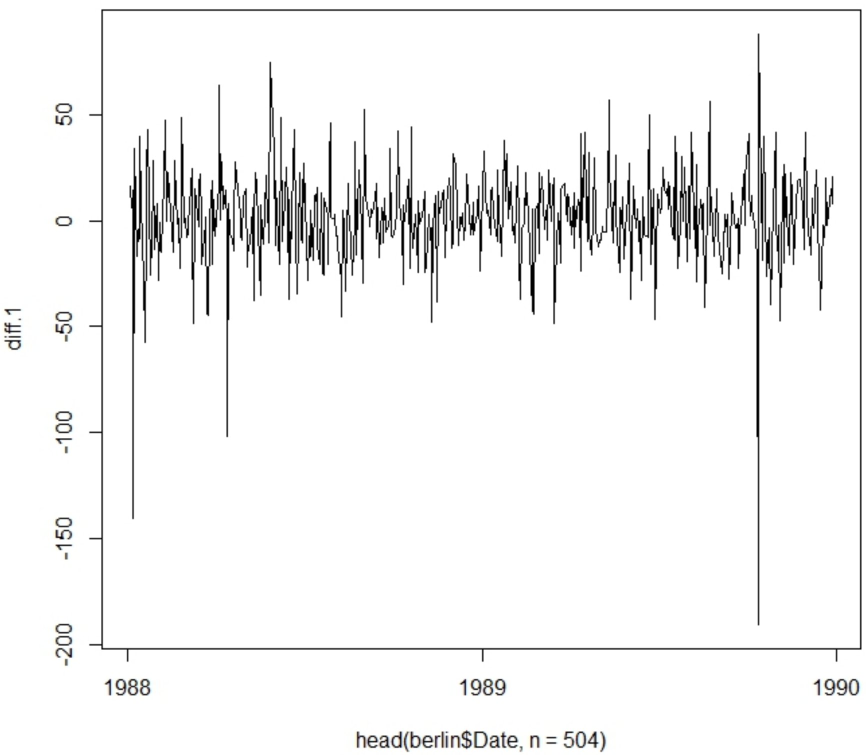
<u>Критерий Шварца (BIC):</u>

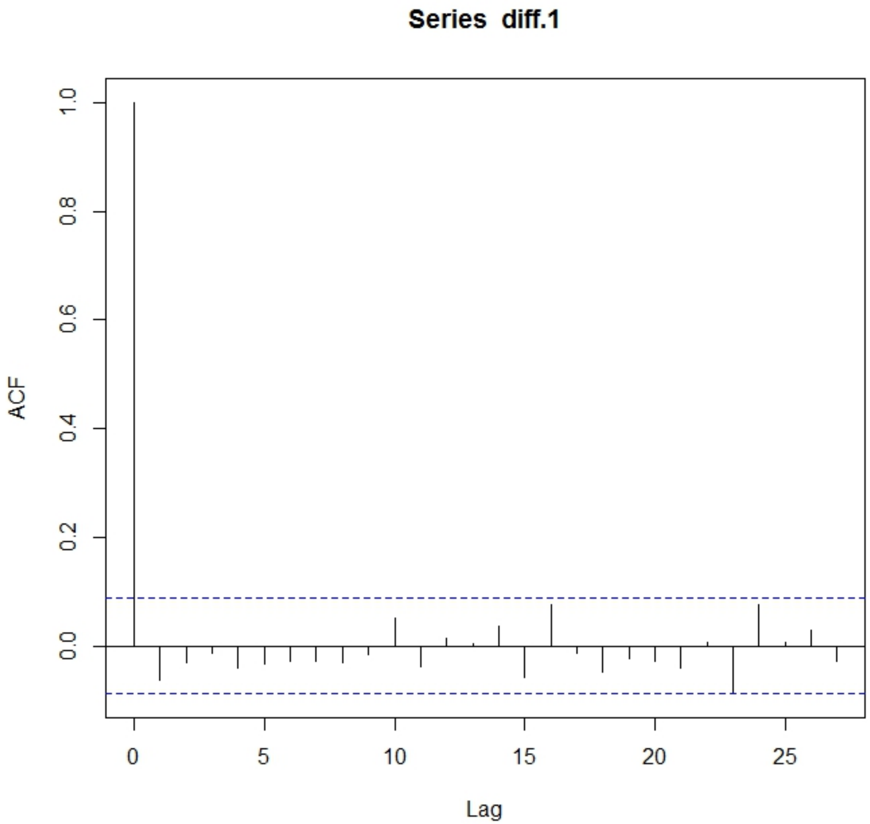
$$\boldsymbol{BIC = ln\sigma^2 + \frac{(lnT-2)(p+q+1)}{T}}$$

Чем меньше AIC (BIC), тем лучше считается модель.

ПРИМЕР ПОДБОРА ПАРАМЕТРОВ ПО AIC

- В моделях ARIMA, SARIMA и др. необходимо сначала перейти к стационарному ряду, и уже для него подбирать значения параметров AR и MA по критерию AIC.





ПРИМЕР ПОДБОРА ПАРАМЕТРОВ ПО AIC

		p in AR(p)					
		0	1	2	3	4	5
q in MA(q)	0	4588.666	4588.472	4589.884	4591.619	4592.181	4593.312
	1	4588.618	4584.675	4586.262	4588.261	4590.172	4592.002
	2	4590.031	4586.263	4588.317	4590.25	4590.726	4594.104
	3	4591.883	4589.089	4583.762	4593.013	4589.644	4590.99
	4	4592.883	4590.161	4592.254	4594.099	4583.88	4586.875
	5	4594.055	4590.793	4594.07	4596.018	4586.779	4587.788

- https://coolstatsblog.com/2013/08/14/using-aic-to-test-arima-models-2/

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.