Математические основы анализа временных рядов

Русский

Вольфганг Бёме

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

25 октября 2026 г.

Документация на этом сайте разработана при поддержке ИИ Claude (Anthropic).

Содержание

1 Введение **2**
1.1 Обозначения . . . 2

2 Описательная статистика **3**
2.1 Меры положения . . . 3
2.1.1 Среднее значение (арифметическое среднее) . . . 3
2.1.2 Медиана . . . 3
2.1.3 Мода . . . 3
2.2 Меры разброса . . . 3
2.2.1 Дисперсия . . . 3
2.2.2 Стандартное отклонение . . . 3
2.3 Квартили . . . 3
2.3.1 Межквартильный размах (IQR) . . . 3
2.4 Меры формы распределения . . . 4
2.4.1 Асимметрия (Skewness) . . . 4
2.4.2 Эксцесс (Kurtosis) . . . 4

3 Трендовые функции **5**
3.1 Линейная регрессия . . . 5
3.2 Полиномиальная регрессия . . . 5
3.3 Экспоненциальная регрессия . . . 5
3.4 Логарифмическая регрессия . . . 6
3.5 Показатели качества . . . 6
3.5.1 Коэффициент детерминации (R^2) . . . 6
3.5.2 Скорректированный R^2 . . . 6
3.5.3 Среднеквадратичная ошибка (RMSE) . . . 6
3.5.4 Средняя абсолютная ошибка (MAE) . . . 6

4 Анализ автокорреляции **7**
4.1 Функция автокорреляции (ACF) . . . 7
4.2 Частичная функция автокорреляции (PACF) . . . 7
4.3 Тест Льюнга–Бокса . . . 7

5 Методы сглаживания **8**
5.1 Простое скользящее среднее (SMA) . . . 8
5.2 Экспоненциальное сглаживание (EMA) . . . 8
5.3 Взвешенное скользящее среднее (WMA) . . . 8
5.4 Метод Холта–Уинтерса . . . 8
5.5 LOESS (локальная полиномиальная регрессия) . . . 8

6 Декомпозиция временных рядов **9**
6.1 Аддитивная модель . . . 9
6.2 Мультипликативная модель . . . 9
6.3 Алгоритм декомпозиции . . . 9

7 Методы фильтрации **10**
7.1 Фильтр Ходрика–Прескотта 10
7.2 Фильтр Баттерворта 10
7.3 Фильтр Савицкого–Голея 10

8 Сезонная корректировка **11**
8.1 STL-декомпозиция 11
8.2 X-13 ARIMA-SEATS (упрощённый) 11
8.3 Метод Census I 12
8.4 Метод Census II 12

9 Полный пример **13**
9.1 Описательная статистика 13
9.2 Линейная регрессия 13
9.3 Сезонная декомпозиция (период = 12) 13
9.4 HP-фильтр при $\lambda = 14400$ 14

10 Резюме **15**
10.1 Обзор методов 15
10.2 Рекомендации 15
10.3 Рекомендуемая литература 15

1 Введение

Данный документ описывает математические основы всех методов, реализованных в Анализаторе временных рядов. Для каждого метода приводятся лежащие в его основе формулы, алгоритмы и — там, где это целесообразно — примеры.

1.1 Обозначения

Используются следующие обозначения:

- y_t: наблюдение в момент времени t
- n: количество наблюдений
- μ: среднее значение (математическое ожидание)
- σ^2: дисперсия
- σ: стандартное отклонение

2 Описательная статистика

2.1 Меры положения

2.1.1 Среднее значение (арифметическое среднее)

$$\mu = \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t \tag{1}$$

Пример: Для временного ряда $y = [100, 102, 101, 105, 103]$:

$$\mu = \frac{100 + 102 + 101 + 105 + 103}{5} = \frac{511}{5} = 102{,}2$$

2.1.2 Медиана

Медиана — это среднее значение упорядоченного ряда:

$$\text{Медиана} = \begin{cases} y_{\frac{n+1}{2}} & \text{если } n \text{ нечётное} \\ \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2} & \text{если } n \text{ чётное} \end{cases} \tag{2}$$

2.1.3 Мода

Мода — наиболее часто встречающееся значение в ряду.

2.2 Меры разброса

2.2.1 Дисперсия

$$\sigma^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \mu)^2 \tag{3}$$

2.2.2 Стандартное отклонение

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \mu)^2} \tag{4}$$

Пример: Для $y = [100, 102, 101, 105, 103]$ при $\mu = 102{,}2$:

$$\begin{aligned} \sigma^2 &= \frac{(100 - 102{,}2)^2 + (102 - 102{,}2)^2 + \cdots + (103 - 102{,}2)^2}{5} \\ &= \frac{4{,}84 + 0{,}04 + 1{,}44 + 7{,}84 + 0{,}64}{5} = \frac{14{,}8}{5} = 2{,}96 \\ \sigma &= \sqrt{2{,}96} \approx 1{,}72 \end{aligned}$$

2.3 Квартили

2.3.1 Межквартильный размах (IQR)

$$Q_1 = 25\%\text{-квантиль} \tag{5}$$

$$Q_3 = 75\%\text{-квантиль} \tag{6}$$

$$\text{IQR} = Q_3 - Q_1 \tag{7}$$

2.4 Меры формы распределения

2.4.1 Асимметрия (Skewness)

Измеряет асимметричность распределения:

$$\text{Асимметрия} = \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \mu)^3}{\sigma^3} \tag{8}$$

Интерпретация:

- Асимметрия $= 0$: симметричное распределение
- Асимметрия > 0: правосторонняя (длинный хвост справа)
- Асимметрия < 0: левосторонняя (длинный хвост слева)

2.4.2 Эксцесс (Kurtosis)

Измеряет «остроконечность» распределения:

$$\text{Эксцесс} = \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \mu)^4}{\sigma^4} - 3 \tag{9}$$

Интерпретация (избыточный эксцесс):

- Эксцесс $= 0$: нормальное распределение
- Эксцесс > 0: лептокуртическое (островершинное, тяжёлые хвосты)
- Эксцесс < 0: платикуртическое (плосковершинное, лёгкие хвосты)

3 Трендовые функции

3.1 Линейная регрессия

Линейная трендовая функция имеет вид:

$$\hat{y}_t = a + b \cdot t \tag{10}$$

Параметры оцениваются методом наименьших квадратов:

$$b = \frac{n\sum_{t=1}^{n} t \cdot y_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2} \tag{11}$$

$$a = \frac{\sum_{t=1}^{n} y_t - b\sum_{t=1}^{n} t}{n} \tag{12}$$

Пример: Для $y = [100, 102, 105, 108, 110]$ при $t = [0, 1, 2, 3, 4]$:

$$b = \frac{5 \cdot 1130 - 10 \cdot 525}{5 \cdot 30 - 100} = \frac{5650 - 5250}{150 - 100} = \frac{400}{50} = 2{,}5$$

$$a = \frac{525 - 2{,}5 \cdot 10}{5} = \frac{500}{5} = 100$$

Трендовая функция: $\hat{y}_t = 100 + 2{,}5t$

3.2 Полиномиальная регрессия

Общая форма полинома степени p:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_p t^p = \sum_{k=0}^{p} a_k t^k \tag{13}$$

Коэффициенты определяются из системы нормальных уравнений:

$$\mathbf{X}^T\mathbf{X}\mathbf{a} = \mathbf{X}^T\mathbf{y} \tag{14}$$

где $\mathbf{X}$ — матрица Вандермонда:

$$\mathbf{X} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^p \\ 1 & t_2 & t_2^2 & \cdots & t_2^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^p \end{pmatrix} \tag{15}$$

3.3 Экспоненциальная регрессия

Форма:

$$\hat{y}_t = a \cdot e^{bt} \tag{16}$$

Логарифмирование сводит задачу к линейной:

$$\ln(\hat{y}_t) = \ln(a) + bt \tag{17}$$

3.4 Логарифмическая регрессия

Форма:

$$\hat{y}_t = a + b \cdot \ln(t+1) \tag{18}$$

3.5 Показатели качества

3.5.1 Коэффициент детерминации (R^2)

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{SS_{res}}{SS_{tot}} \tag{19}$$

$R^2 \in [0, 1]$; более высокие значения означают лучшее соответствие.

3.5.2 Скорректированный R^2

Учитывает количество параметров:

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n-1)}{n - k - 1} \tag{20}$$

где k — число независимых переменных.

3.5.3 Среднеквадратичная ошибка (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{21}$$

3.5.4 Средняя абсолютная ошибка (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{22}$$

4 Анализ автокорреляции

4.1 Функция автокорреляции (ACF)

ACF измеряет линейную зависимость между наблюдениями с лагом k:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(y_t - \mu)(y_{t+k} - \mu)}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \mu)^2} \tag{23}$$

где γ_k — автоковариация для лага k.
Свойства:

- $\rho_0 = 1$ (полная корреляция с самим собой)
- $-1 \leq \rho_k \leq 1$
- $\rho_k = \rho_{-k}$ (симметрия)

4.2 Частичная функция автокорреляции (PACF)

PACF измеряет корреляцию между y_t и y_{t+k} после исключения влияния всех промежуточных лагов:

$$\phi_{kk} = \text{PACF}(k) \tag{24}$$

Вычисляется через уравнения Юла–Уокера или алгоритм Дарбина–Левинсона.

4.3 Тест Льюнга–Бокса

Проверяет наличие значимой автокорреляции:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\rho_k^2}{n-k} \tag{25}$$

Q асимптотически следует распределению χ^2 с h степенями свободы.
Гипотезы:

- H_0: отсутствие автокорреляции до лага h
- H_1: значимая автокорреляция присутствует

При $Q > \chi^2_{0{,}05,\,h}$ нулевая гипотеза отвергается (уровень значимости 5%).

5 Методы сглаживания

5.1 Простое скользящее среднее (SMA)

Простое скользящее среднее с размером окна m:

$$\hat{y}_t = \frac{1}{m} \sum_{i=0}^{m-1} y_{t-i} \tag{26}$$

Пример: Для $y = [100, 102, 105, 108, 110]$ при $m = 3$:

$$\hat{y}_3 = \frac{100 + 102 + 105}{3} = 102{,}33$$
$$\hat{y}_4 = \frac{102 + 105 + 108}{3} = 105{,}00$$
$$\hat{y}_5 = \frac{105 + 108 + 110}{3} = 107{,}67$$

5.2 Экспоненциальное сглаживание (EMA)

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}, \quad \alpha = \frac{2}{m+1} \tag{27}$$

где $\alpha \in (0, 1)$ — параметр сглаживания: более высокие значения придают больший вес свежим наблюдениям.

5.3 Взвешенное скользящее среднее (WMA)

$$\hat{y}_t = \frac{\sum_{i=0}^{m-1}(m-i) \cdot y_{t-i}}{\sum_{i=0}^{m-1}(m-i)} = \frac{\sum_{i=0}^{m-1}(m-i) \cdot y_{t-i}}{\frac{m(m+1)}{2}} \tag{28}$$

Более свежим наблюдениям присваиваются линейно возрастающие веса.

5.4 Метод Холта–Уинтерса

Учитывает уровень и тренд:

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \tag{29}$$
$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \tag{30}$$
$$\hat{y}_t = L_t + T_t \tag{31}$$

где α — параметр уровня, β — параметр тренда.

5.5 LOESS (локальная полиномиальная регрессия)

Для каждой точки t используются $h = \lfloor \text{span} \cdot n \rfloor$ ближайших соседей с трикубическими весами:

$$w_j = \left(1 - \left(\frac{|t - j|}{d_{\max}}\right)^3\right)^3 \tag{32}$$

где $d_{\max}$ — расстояние до самого дальнего соседа.

6 Декомпозиция временных рядов

Временной ряд разлагается на компоненты.

6.1 Аддитивная модель

$$y_t = T_t + S_t + R_t \tag{33}$$

Используется, когда амплитуда сезонных колебаний постоянна.

6.2 Мультипликативная модель

$$y_t = T_t \cdot S_t \cdot R_t \tag{34}$$

Используется, когда сезонные колебания пропорциональны уровню ряда.

6.3 Алгоритм декомпозиции

Шаг 1. Оценка тренда с помощью центрированного скользящего среднего:

$$T_t = \frac{1}{p} \sum_{i=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor} y_{t+i} \tag{35}$$

Шаг 2. Удаление тренда:

$$D_t = \begin{cases} y_t - T_t & \text{(аддитивная)} \\ y_t / T_t & \text{(мультипликативная)} \end{cases} \tag{36}$$

Шаг 3. Оценка сезонной компоненты усреднением по периодам.
Шаг 4. Вычисление остатка:

$$R_t = \begin{cases} y_t - T_t - S_t & \text{(аддитивная)} \\ y_t / (T_t \cdot S_t) & \text{(мультипликативная)} \end{cases} \tag{37}$$

7 Методы фильтрации

7.1 Фильтр Ходрика–Прескотта

HP-фильтр разделяет временной ряд на тренд (τ_t) и цикл (c_t):

$$y_t = \tau_t + c_t \tag{38}$$

Тренд определяется минимизацией целевой функции:

$$\min_{\{\tau_t\}} \left\{ \sum_{t=1}^{n} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{n-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right\} \tag{39}$$

Первый член измеряет качество подгонки, второй штрафует за кривизну тренда.
Параметр λ:

- $\lambda = 100$: годовые данные
- $\lambda = 1600$: квартальные данные
- $\lambda = 14400$: месячные данные
- $\lambda \to 0$: тренд точно следует данным
- $\lambda \to \infty$: тренд становится линейным

Решение получается методом Гаусса из системы уравнений:

$$(\mathbf{I} + \lambda \mathbf{K}^T \mathbf{K})\boldsymbol{\tau} = \mathbf{y} \tag{40}$$

где $\mathbf{K}$ — матрица вторых разностей.

7.2 Фильтр Баттерворта

Двунаправленное экспоненциальное сглаживание:

$$\text{Прямой проход:} \quad f_t = \alpha y_t + (1-\alpha) f_{t-1} \tag{41}$$

$$\text{Обратный проход:} \quad b_t = \alpha y_t + (1-\alpha) b_{t+1} \tag{42}$$

$$\text{Тренд:} \quad \tau_t = \frac{f_t + b_t}{2} \tag{43}$$

7.3 Фильтр Савицкого–Голея

Локальная полиномиальная аппроксимация с размером окна m и степенью полинома p:

$$\hat{y}_t = \sum_{j=-\lfloor m/2 \rfloor}^{\lfloor m/2 \rfloor} c_j y_{t+j} \tag{44}$$

Коэффициенты c_j определяются МНК-аппроксимацией полинома в локальном окне.

8 Сезонная корректировка

8.1 STL-декомпозиция

STL (Seasonal and Trend decomposition using Loess) — надёжный метод разложения временных рядов.

Алгоритм:

Внутренний цикл:

1. **Удаление тренда:** $D_t = y_t - T_t$ с помощью LOESS
2. **Сглаживание подрядов:** сгладить каждый период-подряд
3. **Фильтр нижних частот:** сгладить сезонную компоненту
4. **Удаление сезонности:** $V_t = y_t - S_t$
5. **Сглаживание тренда:** сгладить V_t методом LOESS

Внешний цикл: робастизация к выбросам

Результат:

$$y_t = T_t + S_t + R_t \tag{45}$$

8.2 X-13 ARIMA-SEATS (упрощённый)

X-13 — стандартный метод официальной статистики.

Упрощённый алгоритм:

Шаг 1. Тренд со скользящим средним периода p:

$$T_t = \frac{1}{p} \sum_{i=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor} y_{t+i} \tag{46}$$

Шаг 2. Сезонные факторы:

$$SF_j = \frac{1}{m} \sum_{i=1}^{m} (y_{j+ip} - T_{j+ip}) \quad \text{для } j = 1, \ldots, p \tag{47}$$

Шаг 3. Нормализация:

$$SF_j^* = SF_j - \frac{1}{p} \sum_{j=1}^{p} SF_j \tag{48}$$

Шаг 4. Скорректированный ряд:

$$SA_t = y_t - SF_{(t \bmod p)} \tag{49}$$

8.3 Метод Census I

Метод отношения к скользящему среднему:

Шаг 1. Центрированное скользящее среднее:

$$MA_t = \frac{1}{p} \sum_{i=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor} y_{t+i} \tag{50}$$

Шаг 2. Отношения к скользящему среднему:

$$r_t = \frac{y_t}{MA_t} \tag{51}$$

Шаг 3. Сезонные индексы усреднением:

$$SI_j = \frac{1}{m} \sum_{i=1}^{m} r_{j+ip} \quad \text{для } j = 1, \ldots, p \tag{52}$$

Шаг 4. Нормализация по периоду p:

$$SI_j^* = \frac{SI_j}{\frac{1}{p}\sum_{j=1}^{p} SI_j} \cdot p \tag{53}$$

Шаг 5. Скорректированный ряд:

$$SA_t = \frac{y_t}{SI^*_{(t \bmod p)}} \tag{54}$$

8.4 Метод Census II

Расширенная версия Census I с многократными итерациями:

Итерация 1: Начальная оценка тренда со скользящим средним

Итерация 2. Вычисление сезонно-нерегулярной компоненты:

$$SI_t = y_t - T_t \tag{55}$$

Итерация 3: Сезонная компонента усреднением

Итерация 4. Сглаживание сезонной компоненты трёхчленным скользящим средним:

$$S_j^* = \frac{S_{j-1} + S_j + S_{j+1}}{3} \tag{56}$$

Итерация 5. Центрирование:

$$S_j^{**} = S_j^* - \frac{1}{p} \sum_{j=1}^{p} S_j^* \tag{57}$$

Итерация 6. Скорректированный ряд:

$$SA_t = y_t - S^{**}_{(t \bmod p)} \tag{58}$$

Итерация 7: Повторная оценка тренда из SA_t

9 Полный пример

Рассмотрим месячный временной ряд за 2 года (24 наблюдения):

$$y = [100, 95, 98, 102, 110, 120, 125, 118, 112, 105, 98, 95,\ 105, 98, 102, 108, 118, 130, 138, 128, 120, 110, 102, 9$$

9.1 Описательная статистика

$$\begin{aligned} n &= 24 \\ \mu &= \frac{1}{24}\sum_{t=1}^{24} y_t = \frac{2653}{24} = 110{,}54 \\ \sigma^2 &= \frac{1}{24}\sum_{t=1}^{24}(y_t - 110{,}54)^2 = 156{,}41 \\ \sigma &= \sqrt{156{,}41} = 12{,}51 \end{aligned}$$

9.2 Линейная регрессия

При $t = 0, 1, 2, \ldots, 23$:

$$b = \frac{24 \cdot 28626 - 276 \cdot 2653}{24 \cdot 4324 - 76176} = \frac{687024 - 732228}{103776 - 76176} = \frac{-45204}{27600} = -1{,}64$$
$$a = \frac{2653 - (-1{,}64) \cdot 276}{24} = \frac{3105{,}64}{24} = 129{,}4$$

Трендовая функция: $\hat{y}_t = 129{,}4 - 1{,}64t$

Отрицательный тренд указывает на лёгкий спад с течением времени (что не соответствует визуальной интерпретации — это демонстрирует ограниченность чисто линейной аппроксимации для сезонных данных).

9.3 Сезонная декомпозиция (период = 12)

Шаг 1. Тренд с 12-месячным скользящим средним (для месяца 7):

$$T_7 = \frac{1}{12}\sum_{i=1}^{12} y_{i+1} = \frac{1}{12}(95 + 98 + \cdots + 105) = 108{,}5$$

Шаг 2. Удаление тренда (аддитивно):

$$D_7 = y_7 - T_7 = 125 - 108{,}5 = 16{,}5$$

Шаг 3. Сезонные компоненты для каждого месяца:

$$S_{\text{Янв}} = \frac{(y_1 - T_1) + (y_{13} - T_{13})}{2} \approx -10{,}5$$

Шаг 4. Скорректированный ряд:

$$SA_7 = y_7 - S_7 = 125 - 16{,}5 = 108{,}5$$

9.4 HP-фильтр при $\lambda = 14400$

Решается система $(\mathbf{I} + \lambda \mathbf{K}^T \mathbf{K})\boldsymbol{\tau} = \mathbf{y}$.

Для $t = 7$, например:

$$\tau_7 \approx 109{,}2$$

Цикл:

$$c_7 = y_7 - \tau_7 = 125 - 109{,}2 = 15{,}8$$

10 Резюме

10.1 Обзор методов

Категория	Методы
Статистика	Среднее, медиана, мода, дисперсия, асимметрия, эксцесс
Тренд	Линейный, полиномиальный (2–6 ст.), экспоненциальный, логарифмический
Автокорреляция	ACF, PACF, тест Льюнга–Бокса
Сглаживание	SMA, EMA, WMA, Холт–Уинтерс, LOESS
Декомпозиция	Аддитивная/Мультипликативная (тренд, сезонность, остаток)
Фильтры	Ходрика–Прескотта, Баттерворта, Савицкого–Голея
Сезонная корр.	STL, X-13, Census I, Census II

10.2 Рекомендации

- **Для определения тренда:** начните с линейной регрессии; увеличивайте степень полинома только при необходимости
- **Для сглаживания:** LOESS и Холт–Уинтерс — для сложных паттернов; SMA/EMA — для простых случаев
- **Для сезонной корректировки:** STL — надёжный и гибкий метод; X-13 — для официальной статистики
- **Для разделения тренда и цикла:** HP-фильтр с подходящим λ

10.3 Рекомендуемая литература

- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting.* Springer.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice.* OTexts.
- Cleveland, R. B., et al. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1), 3–73.
- Hodrick, R. J., & Prescott, E. C. (1997). Postwar US Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.