21. Временные ряды. Виды временных рядов. Компоненты временного ряда. Алгоритм скользящей средней.

1. Понятие и модели временного ряда

Временной ряд представляет собой последовательность значений показателя, зафиксированных в определенные моменты времени. Значения уровней временного ряда могут быть представлены в виде различных математических моделей в зависимости от взаимодействия его компонентов:

- **Аддитивная модель:** уровни ряда представляются как сумма компонентов:

  $$y_t = T_t + S_t + C_t + E_t$$

  .
- **Мультипликативная модель:** уровни ряда представляются как произведение компонентов:

  $$y_t = T_t \cdot S_t \cdot C_t \cdot E_t$$

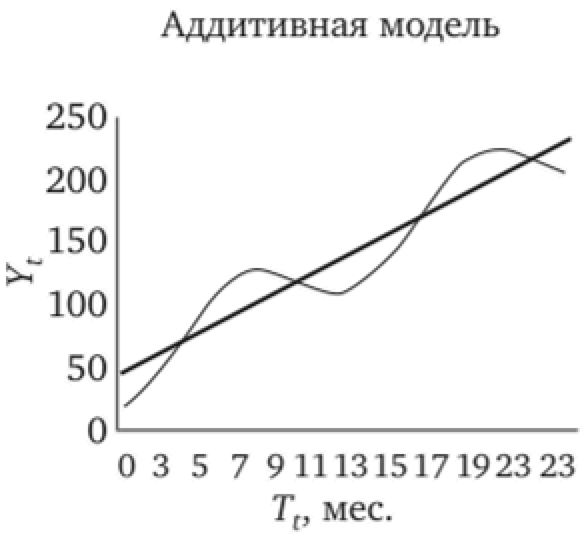
  .
- **Смешанная модель:**

  $$y_t = T_t \cdot S_t \cdot C_t + E_t$$

  .

Где y_t — фактический уровень ряда, T_t — тренд, S_t — сезонная компонента, C_t — циклическая компонента, E_t — случайная составляющая.

Ключевое свойство: Характер сезонности (аддитивный или мультипликативный) можно определить **на стадии графического анализа**.
В **аддитивной модели** амплитуда сезонных колебаний остается примерно постоянной во времени. В **мультипликативной модели** амплитуда меняется пропорционально уровню тренда.



$$Y_t = T_t + SD_t + ED_t \qquad (5.2)$$

$$Y_t = Tt \cdot SI_t + EI_t \qquad (5.3)$$

2. Компоненты временного ряда

В структуре временного ряда принято выделять четыре основные составляющие:

1. **Тренд (T):** Систематическая составляющая долговременного действия, определяющая общую тенденцию и направление развития ряда.
2. **Сезонная компонента (S):** Регулярные колебания с периодом, не превышающим одного года (например, связанные с климатом или социальными факторами). В аддитивной модели амплитуда сезонности постоянна, в мультипликативной — меняется со временем.
3. **Циклическая компонента (C):** Колебания с периодом более одного года (например, экономические циклы или 11-летние циклы солнечной активности).
4. **Случайная (нерегулярная) компонента (E_t)**: Остаток после удаления тренда и периодических составляющих. Формируется под действием факторов:
   - Резкого действия: катастрофы, войны, кризисы («выбросы»).
   - Текущих факторов: сумма множества мелких побочных причин.

3. Алгоритм простой скользящей средней

Сглаживание — это замена фактических уровней расчетными, которые меньше подвержены колебаниям, для четкого проявления тенденции.

Последовательность шагов:

1. **Определение интервала сглаживания L:** Содержит последовательные уровни ряда ($L < n$). Чем сильнее колебания, тем шире должен быть интервал.
2. **Разбиение на участки:** Исходный ряд разбивается на «активные участки» длиной L, при этом интервал «скользит» по ряду с шагом 1.
3. **Замена центрального значения:** Фактическое значение в центре участка заменяется на среднее арифметическое уровней этого участка.

Формулы:

- При нечетном интервале $L = 2p + 1$:

$$\bar{y}_t = \frac{1}{2p+1} \sum_{i=-p}^{p} y_{t+i}$$

.

- При четном интервале (например, $L = 12$ для месяцев), крайние точки берутся с весом 0.5:

$$\bar{y}_t = \frac{0.5y_{t-p} + y_{t-p+1} + \cdots + y_{t+p-1} + 0.5y_{t+p}}{2p}$$

.

Использование интервалов разной длины в методе скользящей средней обусловлено необходимостью центрирования среднего значения относительно конкретного момента времени.

1. Нечетный интервал ($L = 2p + 1$)

Смысл: Позволяет получить симметричное окно с четко выраженным центральным элементом.
Когда лучше использовать:

- Когда в данных нет выраженной сезонности или когда период сезонности (например, недельный цикл в 7 дней) является нечетным числом.
- Преимущество: Формула проста, а результат $\bar{y}_t$ точно соответствует моменту времени t, так как слева и справа от него находится одинаковое количество точек (p).

2. Четный интервал (например, $L = 12$ или $L = 4$)

Смысл: Используется для устранения сезонных колебаний, когда период сезонности четный.

- Проблема: Если взять простое среднее за 12 месяцев, «центр» попадет в промежуток между 6-м и 7-м месяцами. Это создаст временной сдвиг (фазовое смещение).
- Решение (центрирование): Чтобы вернуть среднее точно на момент t, используется двухэтапное сглаживание или формула с весами 0.5 для крайних точек.
    - Крайние точки (y_{t-p} и y_{t+p}) берутся с весом 0.5, чтобы суммарно они давали «целую» точку, а окно оставалось симметричным относительно центрального момента.
- Когда лучше использовать: При наличии сезонности с четным периодом (квартальные данные — $L = 4$, месячные — $L = 12$).

Длина интервала	Тип данных	Цель
Нечетная ($3, 5, 7, \ldots$)	Недельные циклы (7), произвольный шум	Простота, отсутствие временного сдвига.
Четная ($4, 12, \ldots$)	Квартальные, месячные данные	Точное центрирование тренда при исключении четной сезонности.

L — Длина интервала сглаживания

Это количество последовательных уровней ряда, которые вы «берете в охапку» для расчета одного среднего значения.

- **Смысл:** Она определяет степень «сглаживания». Чем больше L, тем более плавной и ровной будет линия тренда, но тем больше информации о мелких колебаниях вы потеряете.
- **Пример:** Если вы анализируете данные по месяцам и хотите убрать сезонность, вы выбираете $L = 12$.

p — Параметр «радиуса» или полуширина интервала

Этот параметр используется в формулах, чтобы показать, на сколько шагов влево и вправо от центральной точки мы отступаем.

- **Пример:** Если вы выбрали интервал из 5 точек ($L = 5$), то центральная точка — это вы сами, а вокруг вас по 2 соседа с каждой стороны. Значит, $p = 2$.
- **Значение для расчетов:** Именно определяет, сколько значений будет потеряно в начале и в конце временного ряда после сглаживания.

t — Момент времени (индекс)

Это порядковый номер наблюдения в вашем ряду данных.

- **Смысл:** Он указывает на «адрес» конкретного значения. y_t — это значение показателя в момент времени t.
- **В формулах сглаживания:** Когда мы пишем $\bar{y}_t$, это означает сглаженное значение, которое мы вычислили для конкретного момента t, используя его «соседей» (от $t - p$ до $t + p$).

Наглядная таблица связей

Символ	Название	Пример ($L = 5$)
L	Длина окна (сколько точек берем)	**5**
p	Сколько точек берем слева и справа от центра	**2** (так как $2 \cdot 2 + 1 = 5$)
t	Текущий момент времени, который мы сглаживаем	Если считаем для среды, то $t = 3$ (в ряду Пн-Пт)

1. Пример для нечетного интервала ($L = 3$)

Предположим, у нас есть ряд значений курса акций. Для расчета трехчленной скользящей средней ($L = 3$, следовательно, $p = 1$) в момент времени $t = 2$, мы берем значения в точках $t = 1, 2, 3$.
$y_1 = 510$, $y_2 = 497$, $y_3 = 504$
Расчет:

$$\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3} = \frac{510 + 497 + 504}{3} = 503.7$$

Это значение записывается напротив периода $t = 2$. При этом значение для $t = 1$ теряется, так как у него нет «соседа» слева для усреднения.

2. Пример для четного интервала ($L = 4$)

Такой интервал часто используется для квартальных данных, чтобы устранить сезонность.

$$\bar{y}_t = \frac{0.5y_{t-2} + y_{t-1} + y_t + y_{t+1} + 0.5y_{t+2}}{4}$$

Почему так? Если просто взять среднее четырех кварталов, результат окажется «между» 2-м и 3-м кварталом. Чтобы «центрировать» значение точно на конкретный момент времени, мы берем 5 уровней, но крайние учитываем лишь наполовину.
Пример расчета: Допустим, значения за 5 кварталов: $10, 20, 30, 40, 50$. Расчет для центрального (3-го) квартала будет таким:

$$\bar{y}_3 = \frac{(0.5 \cdot 10) + 20 + 30 + 40 + (0.5 \cdot 50)}{4} = \frac{5 + 20 + 30 + 40 + 25}{4} = \frac{120}{4} = 30$$

3. Пример взвешенной скользящей средней ($L = 5$)

Веса для пяти точек будут: $-3, 12, 17, 12, -3$, а сумма весов (знаменатель) равна 35.
Расчет для $t = 3$:

$$\bar{y}_3 = \frac{-3y_1 + 12y_2 + 17y_3 + 12y_4 - 3y_5}{35}$$

$$\bar{y}_3 = \frac{-3(510) + 12(497) + 17(504) + 12(510) - 3(509)}{35} = 502.7$$

Главный вывод: Простая средняя «сглаживает» всё одинаково, а взвешенная средняя отдает приоритет центральному значению (вес 17), что позволяет точнее отслеживать изгибы графика.

4. Теоретическое обоснование (Доказательство)

Процедуры скользящих средних опираются на **теорему Вейерштрасса**: любая гладкая функция на ограниченном интервале может быть представлена полиномом.

При использовании простой скользящей средней осуществляется аппроксимация неслучайной составляющей линейной функцией $y = a_0 + a_1 t$.
Для нахождения коэффициентов используется **метод наименьших квадратов (МНК)**. Если перенести начало координат в середину интервала ($t = -p, \dots, 0, \dots, p$), то сумма моментов времени $\sum t = 0$.

Минимизация функционала:

$$F = \sum_{t=-p}^{p} (a_0 + a_1 t - y_t)^2 \to \min$$

.

Функционал в данном случае — это математическое выражение, которое описывает общую ошибку (сумму квадратов отклонений) между нашими реальными данными (y_t) и линией, которую мы строим ($a_0 + a_1 t$).

Система нормальных уравнений: (ищем минимум функции)

1. $\frac{\partial F}{\partial a_0} = 2 \sum_{t=-p}^{p} (a_0 + a_1 t - y_t) = 0 \implies (2p+1)a_0 + a_1 \sum_{t=-p}^{p} t = \sum_{t=-p}^{p} y_t$
2. $\frac{\partial F}{\partial a_1} = 2 \sum_{t=-p}^{p} (a_0 + a_1 t - y_t)t = 0 \implies a_0 \sum_{t=-p}^{p} t + a_1 \sum_{t=-p}^{p} t^2 = \sum_{t=-p}^{p} y_t t.$

Так как $\sum t = 0$, из первого уравнения получаем:

$$a_0 = \frac{\sum_{t=-p}^{p} y_t}{2p+1}$$

.

Поскольку при $t = 0$ расчетное значение $\hat{y}_0 = a_0$, это доказывает, что сглаженное значение в центре участка — это **среднее арифметическое** его уровней.

1. Связь с методом наименьших квадратов (МНК)

Вывод доказывает, что когда вы считаете обычное среднее в «окошке» скользящей средней, вы на самом деле **строите наилучшую прямую линию** (регрессию) для этого участка.

- Математика подтверждает: значение средней арифметической — это не просто удобное число, а наиболее вероятное значение «очищенного» показателя в центре этого интервала с точки зрения минимизации ошибок.

2. Смысл коэффициента a_0

В уравнении прямой $y = a_0 + a_1 t$, коэффициент a_0 — это точка пересечения прямой с осью y (значение в момент времени $t = 0$).

- Поскольку мы хитро назначили центр нашего интервала нулевой точкой ($t = 0$), то найденное значение a_0 и есть наше «улучшенное», сглаженное значение для центральной точки.

3. Обоснование «сглаживания»

Вывод говорит нам: «Хотите узнать реальную тенденцию в этой точке, отбросив случайный шум? Просто возьмите среднее вокруг неё».

- Формула $a_0 = \frac{\sum y_t}{2p+1}$ буквально означает: **Сглаженное значение = Сумма всех уровней в окне / Количество этих уровней.**

4. Почему это работает только для прямой?

Этот конкретный вывод доказывает справедливость **простой** скользящей средней. Он подразумевает, что на маленьком участке (внутри окна L) тренд ведет себя как **прямая линия**.

- Если тренд на этом участке сильно изгибается (как парабола), простая средняя начнет «врать» (давать систематическую ошибку). Именно поэтому для сложных кривых в документе далее предлагается использовать **взвешенную** скользящую среднюю (где веса подбираются уже под полиномы 2-й и 3-й степени).

5. Взвешенная скользящая средняя

Если тренд нелинеен, простая средняя дает искажения. Тогда используют сглаживание полиномами 2-го или 3-го порядка. Веса w_i зависят от удаленности от центра:

$$\bar{y}_t = \frac{\sum w_i y_i}{\sum w_i}$$

Сумма весовых коэффициентов всегда равна 1. Это позволяет лучше сохранять изгибы тренда.

6. Свойства и недостатки метода

1. **Потеря краевых значений:** При интервале $L = 2p + 1$ теряются по p уровней в начале и в конце ряда.
2. **Эффект Слуцкого-Юла:** Усреднение может искусственно создавать видимость циклов там, где были только случайные колебания.
3. **Фильтрация:** Метод идеально подавляет периодику (сезонность), если длина L равна или кратна периоду колебаний.

Аналогия:

Сглаживание скользящей средней похоже на взгляд на шумную толпу с высоты птичьего полета. Вы перестаете видеть резкие движения отдельных людей (случайные колебания), но отчетливо видите, в какую сторону медленно движется вся масса людей (тренд).