Mathematical Forecasting Methods

Лекция 3

МФТИ

Осень, 2024

Краткое повторение

- **Временной ряд** — это совокупность значения параметра $\{x_1, x_2, ..., x_T\} = \{x_t\}_{t=1}^T$, изменяющегося во времени, через равные промежутки времени.
- **Также рассматриваем как:** совокупность случайных величин (*дискретный случайный или стохастический процесс*), где для каждого t значение рассматривается как случайная величина.
- **Задача прогнозирования**: найти функции $f_{T,d}$:

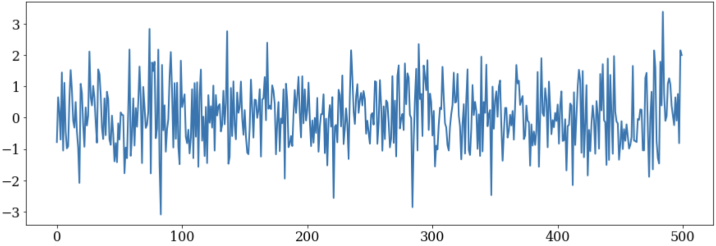
$$x_{T+d} \approx f_{T,d}(x_1, ...x_T; w) =: \hat{x}_{T+d},$$

где $f_{T,d}$ — модель временного ряда, $d = 1, ..., D$ — горизонт прогнозирования.

Краткое повторение

Определение. Временной ряд $\{x_i\}_{i=1}^{T}$ называется слабо стационарным (или стационарным в широком смысле), если

- $\mathrm{E}[x_t] = \mathrm{const}$ (т.е. временной ряд не имеет *тренда*),
- $\mathrm{Cov}(x_t, x_{t+k}) = \mathrm{E}[(x_t - \mathrm{E}x_t)(x_{t+k} - \mathrm{E}x_{t+k})] = \gamma(k)$ (ковариация зависит только от разницы во времени).



Белый шум

Краткое повторение

Модель SARIMAX($p, q, d; P, Q, D$) приводит временной ряд к стационарному виду и прогнозирует с наперед заданной точностью (по Теореме Вольда).

$$x_t = \mu + \underbrace{\phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots}_{p \text{ слагаемых AR}} + \underbrace{u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots}_{q \text{ слагаемых MA}}$$
$$+ \underbrace{\alpha_1 x_{t-S} + \alpha_2 x_{t-2S} + \ldots}_{P \text{ сезонных слагаемых AR}} + \underbrace{\eta_1 u_{t-S} + \eta_2 u_{t-2S} + \ldots}_{Q \text{ сезонных слагаемых MA}}$$
$$+ \underbrace{\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \ldots}_{\text{внешние переменные, не зависящие от } x_t},$$

где $\phi, \theta, \alpha, \eta, \beta$ — настраиваемые параметры модели,
x_t — значения временного ряда (продифференцированные d и D сезонных раз),
u_t — значения шумов,
z_i — внешние (экзогенные) переменные.

Подбор параметров модели

- Если все гиперпараметры модели $p, q, d; P, Q, D$ подобраны, то коэффициенты авторегрессии подбираются методом МНК;
- Для выбора θ шумовая компонента предварительно оценивается на остатках «простой» авторегрессии;
- Если шум гауссовский, то МНК дает оценки максимального правдоподобия.

Вопрос: как подбирать приближение для $p, q, d; P, Q, D$?

Информационные критерии

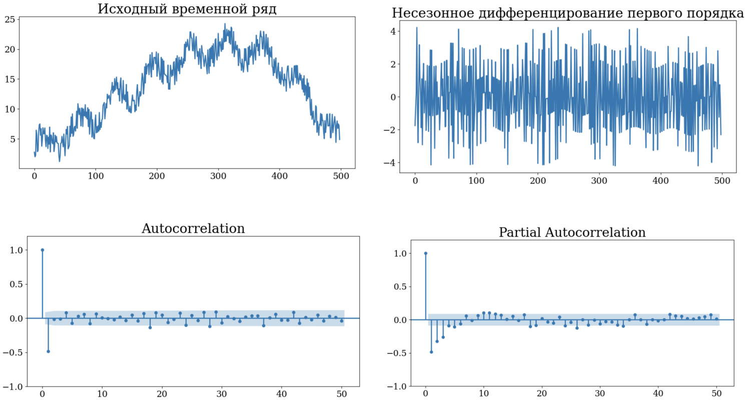
Критерий Акаике (Akaike's information criterion, AIC) — содержит функцию штрафа, линейно зависящую от числа параметров:

$$AIC = \left(\sum_{t=1}^{T} (x_t - \hat{x}_t)^2 \right) + 2(p + q + P + Q + 1)$$

Байесовский информационный критерий (Bayesian information criterion, BIC):

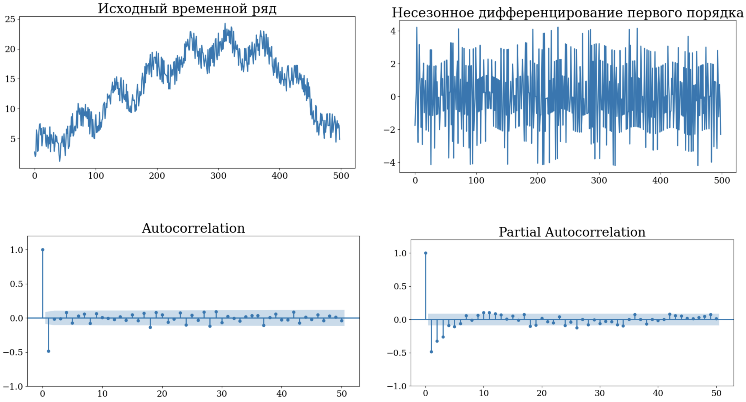
$$BIC = \left(\sum_{t=1}^{T} (x_t - \hat{x}_t)^2 \right) + (p + q + P + Q + 1)(\ln T - 2)$$

Начальное приближение гиперпараметров



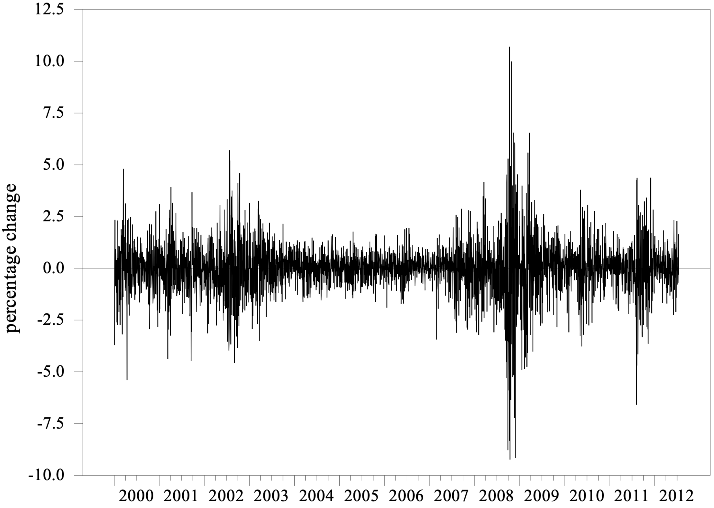
- $Q * S$ — номер последнего **сезонного** лага, при котором автокорреляция статистически значима;
- q — номер последнего **несезонного** лага,при котором автокорреляция статистически значима, меньше сезонного значения;

Начальное приближение гиперпараметров



- $P * S$ — номер последнего **сезонного** лага, при котором **частичная** автокорреляция статистически значима;
- p — номер последнего **несезонного** лага,при котором **частичная** автокорреляция статистически значима, меньше сезонного значения;

Проблема непостоянной дисперсии



Дневные изменения фондового индекса биржи

ARCH и GARCH модели

В традиционных эконометрических моделях дисперсия шума считается постоянной (гомоскедастичность).
Если временные ряды демонстрируют периоды необычно большой волатильности* , за которыми следуют периоды относительного спокойствия, предположение о постоянной дисперсии не выполняется. Для учета такого эффекта используются модели:

- ARCH (autoregressive conditional heteroskedastic model),
- GARCH (generalized autoregressive conditional heteroskedastic model).

* Термин **волатильность** в финансовой аналитике — это отклонение доходности (непредсказуемая часть цен на активы).

Изменяющаяся волатильность

- Модель AR(1): $x_t = \mu + \phi_1 x_{t-1} + u_t = \mu_t + u_t$
- Здесь μ_t и u_t — ожидаемая часть с учетом истории и случайная часть соответственно.
- Если ранее полагалось, что $u_t \sim WN(0,1)$, то теперь возмущения будут моделироваться с меняющейся во времени дисперсией σ_t^2:

  $$u_t = z_t \sqrt{\sigma_t^2},$$

  где $z_t \sim WN(0,1)$
- σ_t^2 является функцией предыдущих значений ошибки $u_{t-1},\ u_{t-2}, \ldots$.

Историческая волатильность

Скользящее среднее.

- $$\sigma_t^2 = \sum_{i=1}^{N} \frac{r_{t-i}^2}{N},$$

  где r_t — ошибка прошлых прогнозов,
- учитывает изменения с задержкой,
- сложности с большим горизонтом прогнозирования.

Экспоненциальное сглаживание

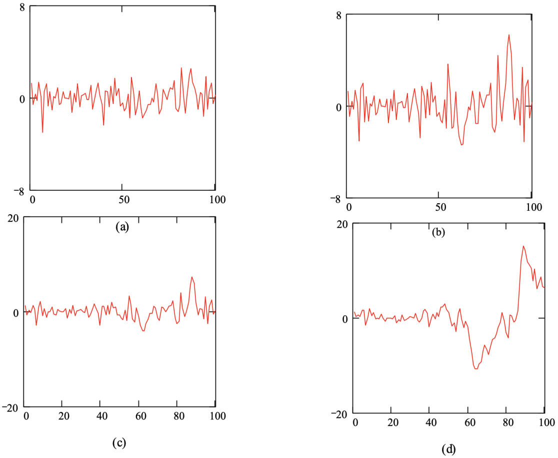
- $$\sigma_t^2 = \lambda r_{t-1}^2 + (1-\lambda) r_{t-2}^2,$$
- новый гиперпараметр λ,
- нестабильный переход от «большой» дисперсии к «малой»,
- сложности с большим горизонтом прогнозирования.

ARCH

- Одна из простых стратегий — смоделировать меняющуюся дисперсию как процесс AR(q), используя квадраты оцененных остатков.
- ARCH(p): $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2$, где $\omega > 0,\ \alpha_i \geq 0\ \forall i = \overline{1, p}$.
- В отличие от скользящего среднего, здесь веса не обязательно равны 1/N.
- Прогнозы:
$\mathrm{E}[u_{t+1}^2] = \sigma_{t+1}^2 = \alpha_0 + \alpha_1 u_t^2 + \alpha_2 u_{t-1}^2 + \cdots + \alpha_p u_{t+1-p}^2$.

Примеры ARCH

$$a) \quad z_t = WN(0,1), \quad b) \quad u_t = z_t\sqrt{1 + 0.8u_{t-1}^2}$$



$$c) \quad x_t = 0.2x_{t-1} + u_t, \qquad d) \quad x_t = 0.9x_{t-1} + u_t$$

GARCH

ARCH-модель предполагает зависимость дисперсии только от квадратов прошлых значений временного ряда.

Обобщить данную модель можно, предположив, что дисперсия зависит также от прошлых значений самой дисперсии (аналог ARMA).

GARCH(p, q):

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Здесь $\omega > 0,\ \alpha_i \geq 0\ \forall i = \overline{1, p},\ \beta_j \geq 0\ \forall j = \overline{1, q}$.

GARCH

Основное преимущество:

- Процесс ARCH(p) требует большого количества лагов для определения структуры зависимости, обычно встречающейся в «сложных» временных рядах.
- Для этого необходимо оценивать множество параметров.
- Модель GARCH допускает более гибкую и экономную спецификацию.

GARCH

Некоторые важные недостатки:

- В моделях GARCH положительные и отрицательные возмущения одинаково влияют на дисперсии. Однако на практике такое изменение влияет на динамику временного ряда по-разному.
- В предыдущих моделях автокорреляционная функция экспоненциально убывает, но прикладные задачи показывают, что квадратичное затухание встречается чаще: такое высокое постоянство может быть достигнуто только с помощью сильно параметризованных GARCH моделей.

Краткое повторение. Модель SARIMAX

К модели SARIMA(p,d,q)(P,D,Q) добавляюся *экзогенные* переменные, значение которых формируется вне модели. Экзогенные переменные являются в модели независимыми величинами.

$$x_t = \mu + \sum_{j=1}^{p} \phi_j x_{t-j} + u_t + \sum_{s=1}^{q} \theta_s u_{t-s} + \ldots + \sum_{i=1}^{r} \beta_i x_i^{\text{exog}}$$

Модель векторной авторегрессии

- Динамика сложных явлений описывается несколькими временными рядами.
- Временные ряды изменяются синхронно в определенной взаимозависимости.
- Необходимы методы совместного моделирования двух или более временных рядов.

Vector Autoregression (VAR)

Рассмотрим подходы к методам совместного моделирования двух или более временных рядов.

Модель векторной авторегрессии VAR(p) порядка p:

$$x_t = \mu_1 + \sum_{j=1}^{p} \beta_j x_{t-j} + \sum_{j=1}^{p} \gamma_j y_{t-j} + u_t$$

$$y_t = \mu_2 + \sum_{j=1}^{p} \delta_j x_{t-j} + \sum_{j=1}^{p} \theta_j y_{t-j} + v_t$$

Здесь $\beta, \gamma, \delta, \theta$ и μ - настраиваемые параметры модели, $u_t, v_t \sim WN(0, 1)$.

Важно: в модели VAR все факторы рассматриваются как эндогенные.

Vector Autoregression (VAR)

В терминах матричных обозначений,

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} A_j \mathbf{x}_{t-j} + \mathbf{u}_t,$$

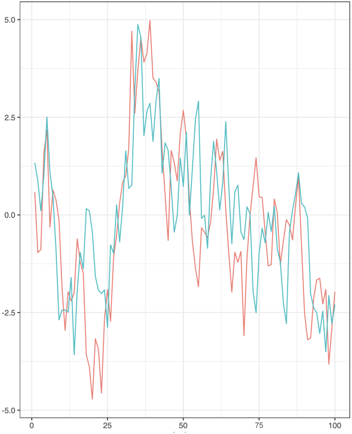
где $\mathbf{x}_t = \begin{pmatrix} x_{1,t} \\ \vdots \\ x_{k,t} \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$, $\mathbf{u}_t = \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{k,t} \end{pmatrix}$,

A_j – матрицы размера $k \times k$.

Vector Autoregression (VAR). Пример.

$$\text{VAR}: x_t = \mu_1 + Ax_{t-j} + u_t$$

$$A = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{pmatrix}$$



Резюме

- При изменяющейся дисперсии ARMA не применима.
- Динамику изменения дисперсии описывает взвешенная сумма предыстории ошибок или оценок дисперсии.
- Модель AR, примененная к истории ошибок, приводит к модели ARCH.
- Модель GARCH — это обобщение модели ARCH, использующее предысторию оценок диспресии σ_t^2.
- Созависимые временные ряды включаются в модель не только как экзогенные переменные.
- Многомерные временные ряды моделируются с помощью векторных моделей.
- Модель VAR — векторная авторегрессия, моделирующая несколько временных рядов.