Временные ряды [2]

На прошлой неделе

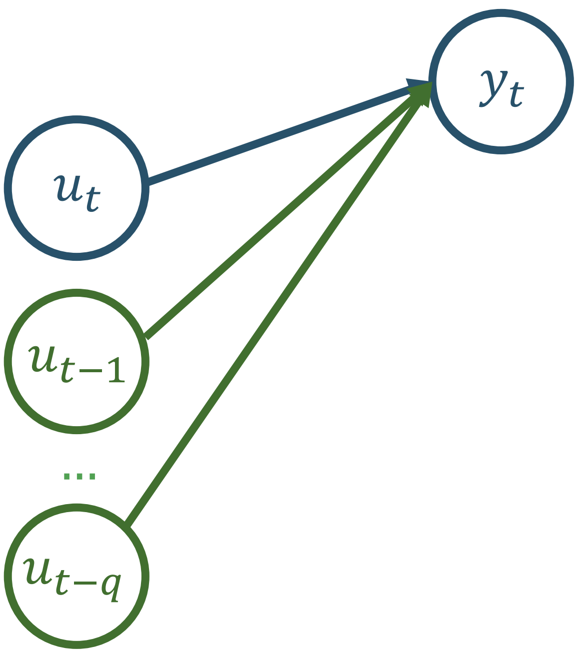
- Особенности временных рядов: стационарность, кросс-валидация
- Обсудили ETS-модель
- Поговорили про декомпозицию ряда
- Обсудили AR и MA модели, начали говорить про SARIMA

План

- Достроим SARIMA модель
- Поговорим о том, как можно усложнять модели: TBATS, векторные модели
- Немного поговорим про проверку гипотез и кластеризацию временных рядов
- Обсудим как использовать временные ряды в качестве признаков для других моделей

SARMA

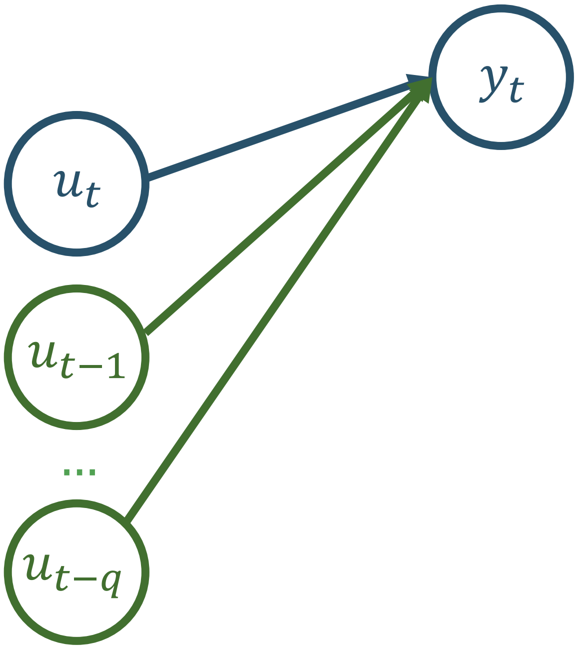
MA(q) (скользящее среднее)



MA(q) (скользящее среднее)

$$u_t \sim iid\ N(0, \sigma^2)$$

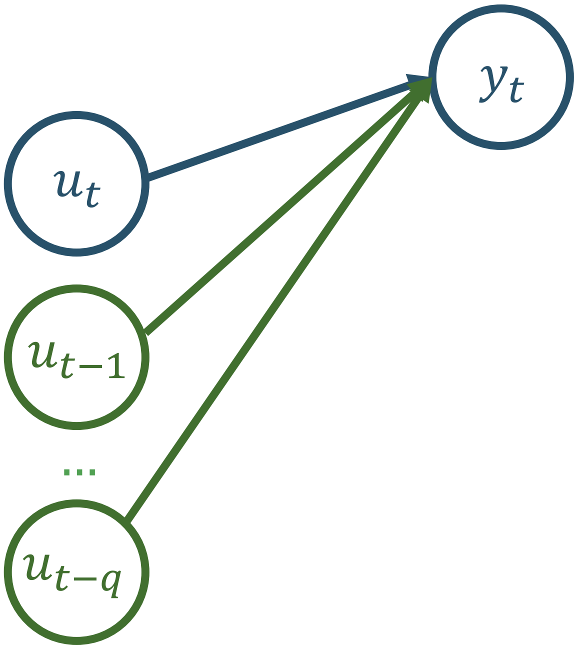
$$y_t = \mu + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \cdots + \alpha_q u_{t-q}$$



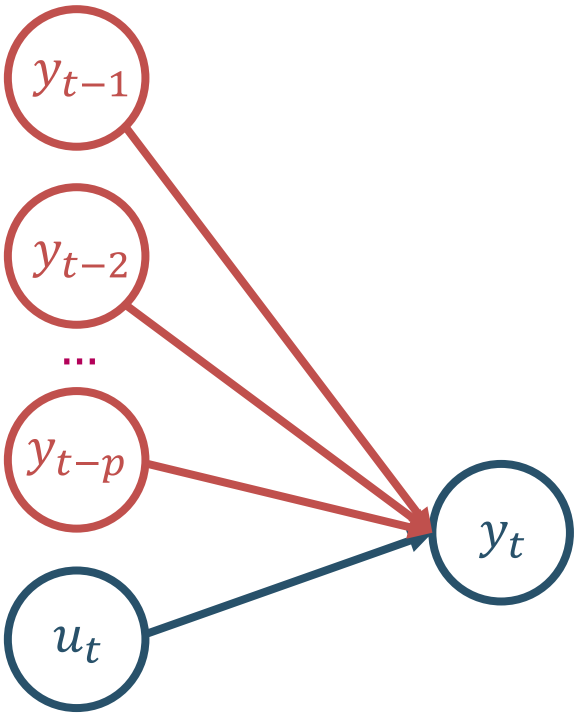
MA(q) (скользящее среднее)

$$u_t \sim iid\, N(0, \sigma^2)$$

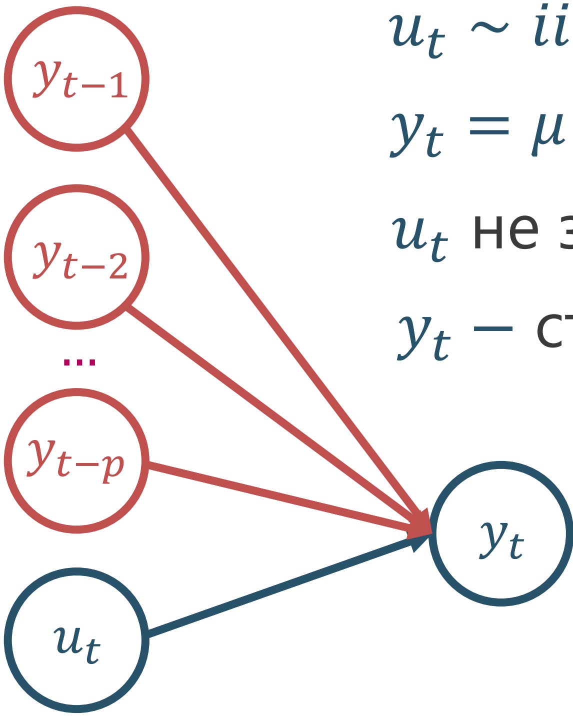
$$y_t - \mu = A(L) \cdot u_t$$



AR(p) (авторегрессия)



AR(p) (авторегрессия)



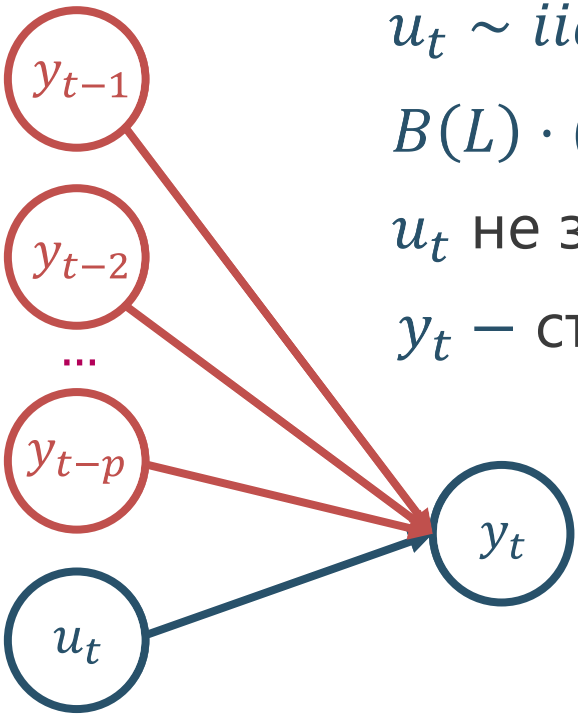
$u_t \sim iid\, N(0, \sigma^2)$

$y_t = \mu + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + u_t$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – стационарный процесс

AR(p) (авторегрессия)



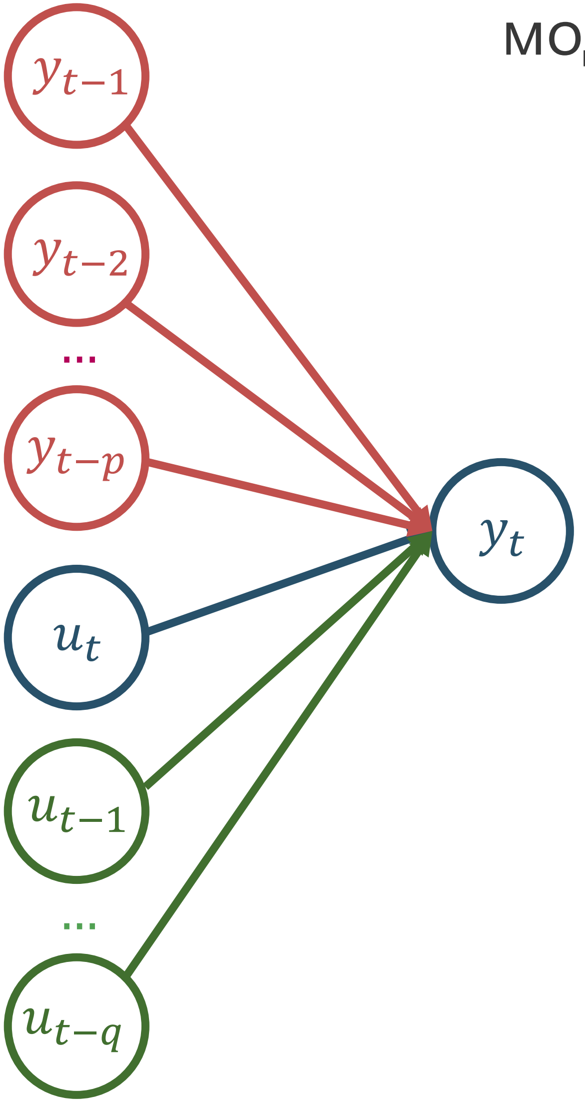
$u_t \sim iid\ N(0, \sigma^2)$

$B(L) \cdot (y_t - \mu^*) = u_t$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

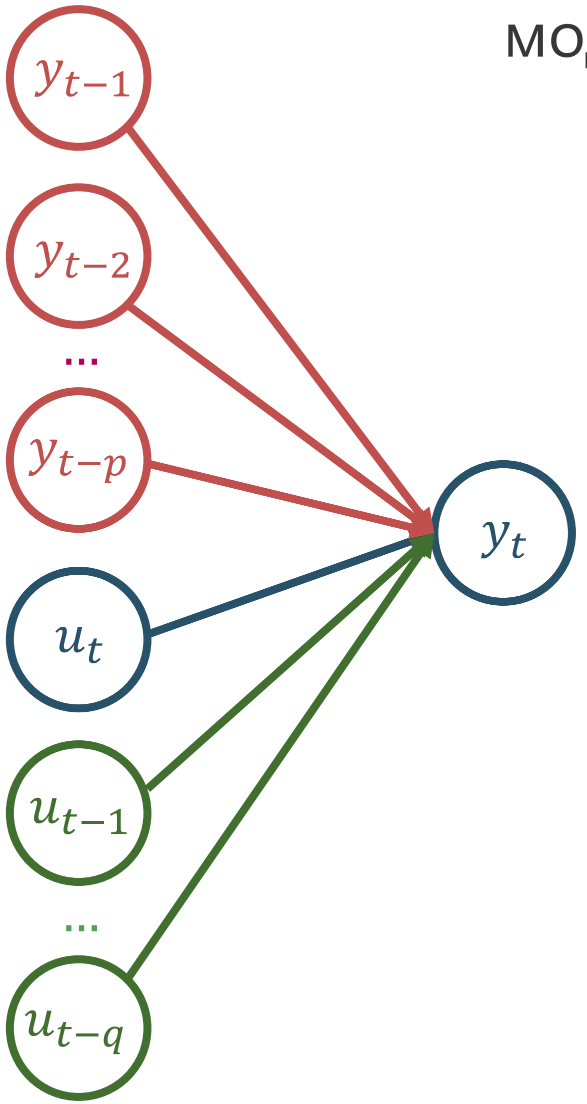
y_t – стационарный процесс

ARMA(p,q)



Соединим $AR(p)$ и $MA(q)$ в одну модель:

ARMA(p,q)



Соединим $AR(p)$ и $MA(q)$ в одну модель:

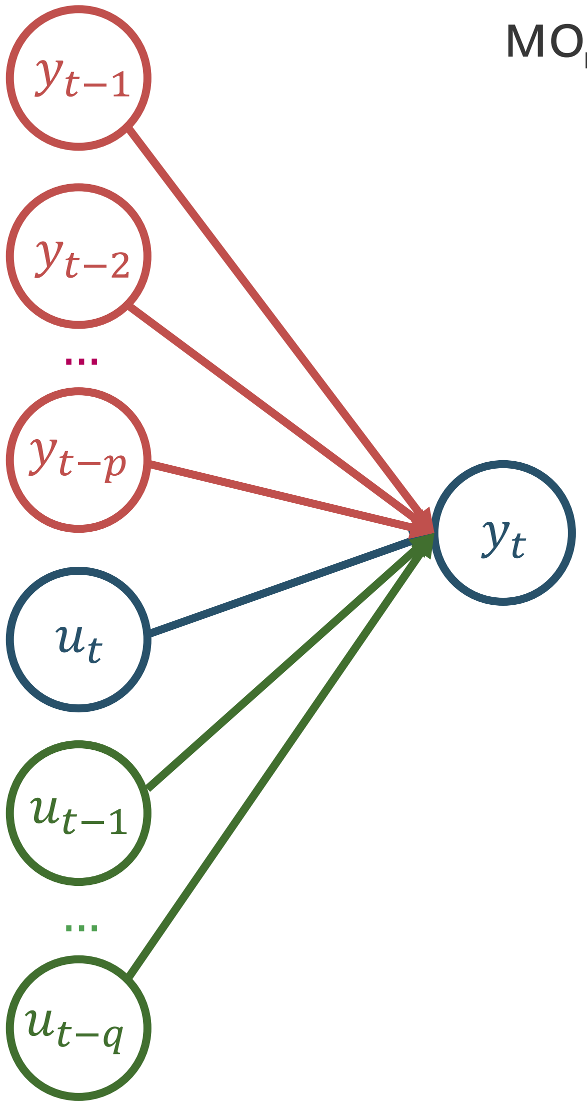
$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

ARMA(p,q)



Соединим $AR(p)$ и $MA(q)$ в одну модель:

$$u_t \sim iid\ N(0, \sigma^2)$$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

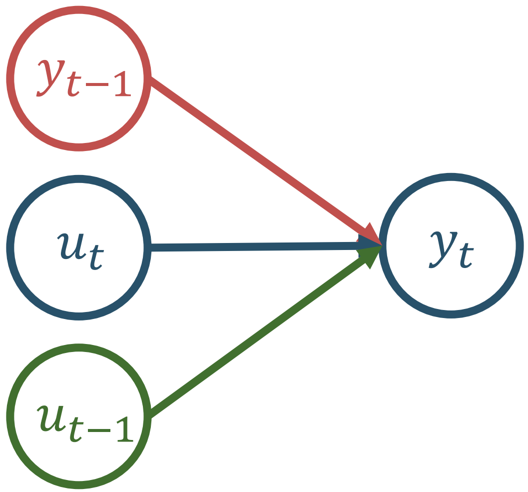
y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

У многочленов $A(L)$ и $B(L)$ нет общих корней

ARMA(1,1)

Соединим $AR(p)$ и $MA(q)$ в одну модель:



$$u_t \sim iid\ N(0, \sigma^2)$$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

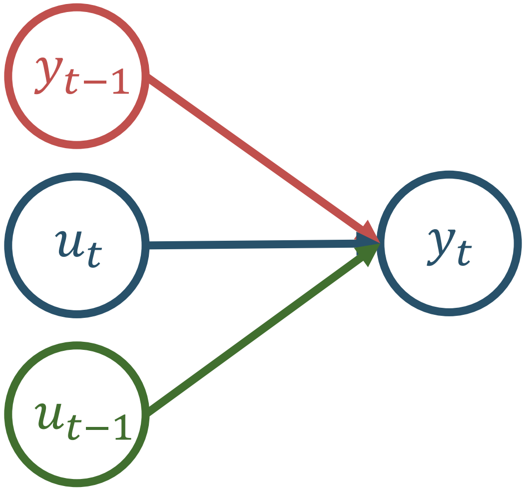
У многочленов $A(L)$ и $B(L)$ нет общих корней

$ARMA(1,1)$:

$$(1 - \beta L) \cdot (y_t - \mu^*) = (1 + \alpha L) u_t$$

ARMA(1,1)

Соединим $AR(p)$ и $MA(q)$ в одну модель:



$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \dots$

y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

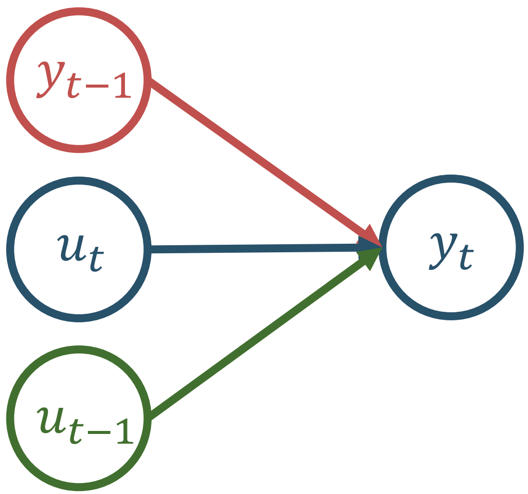
У многочленов $A(L)$ и $B(L)$ нет общих корней

$ARMA(1,1)$:

$$(1 - \beta L) \cdot (y_t - \mu^*) = (1 + \alpha L) u_t$$

$$y_t = \mu + \beta y_{t-1} + u_t + \alpha u_{t-1}$$

ARMA(1,1)



Соединим $AR(p)$ и $MA(q)$ в одну модель:

$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

У многочленов $A(L)$ и $B(L)$ нет общих корней

$ARMA(1,1)$:

$$(1 - \beta L) \cdot (y_t - \mu^*) = (1 + \alpha L) u_t$$

$$y_t = \mu + \beta y_{t-1} + u_t + \alpha u_{t-1}$$

$$\mu^* = \mathbb{E}(y_t) = \frac{\mu}{1 - \beta}$$

Стационарность

Теорема:

Чтобы у ARMA процесса

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

существовало единственное стационарное решение, которое “не заглядывает в будущее”

Стационарность

Теорема:

Чтобы у ARMA процесса

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

существовало единственное стационарное решение, которое “не заглядывает в будущее”, необходимо и достаточно, чтобы все корни характеристического уравнения

$$B(\lambda) = 0$$

$$1 - \beta_1 \lambda - \beta_2 \lambda^2 - \cdots - \beta_p \lambda^p = 0$$

были по модулю **больше единицы**

Сезонная ARMA

$SARMA(p,q)-(P,Q)[12]$

- p – порядок AR-части, q – порядок MA-части
- P – порядок SAR-части, Q – порядок SMA-части
- $[12]$ – период сезонности (месячные данные)

Сезонная ARMA

$SARMA(p, q) - (P, Q)[12]$

- p – порядок AR-части, q – порядок MA-части
- P – порядок SAR-части, Q – порядок SMA-части
- $[12]$ – период сезонности (месячные данные)

$SARMA(1,1) - (1,1)[12]$:

$$(1 - \beta L)(1 - \gamma L^{12})(y_t - \mu^*) = (1 + \alpha L)(1 - \delta L^{12})u_t$$

Сезонная ARMA

$SARMA(p, q) - (P, Q)[12]$

- p – порядок AR-части, q – порядок MA-части
- P – порядок SAR-части, Q – порядок SMA-части
- $[12]$ – период сезонности (месячные данные)

$SARMA(1,1) - (1,1)[12]$:

$$(1 - \beta L)(1 - \gamma L^{12})(y_t - \mu^*) = (1 + \alpha L)(1 - \delta L^{12})u_t$$

$$(1 - \beta L - \gamma L^{12} + \beta\gamma L^{12}) \qquad (1 + \alpha L - \delta L^{12} - \alpha\delta L^{13})$$

Сезонная ARMA

$SARMA(p, q) - (P, Q)[12]$

- p – порядок AR-части, q – порядок MA-части
- P – порядок SAR-части, Q – порядок SMA-части
- $[12]$ – период сезонности (месячные данные)

$SARMA(1,1) - (1,1)[12]$:

$$(1 - \beta L)(1 - \gamma L^{12})(y_t - \mu^*) = (1 + \alpha L)(1 - \delta L^{12})u_t$$

$$(1 - \beta L - \gamma L^{12} + \beta\gamma L^{12}) \qquad (1 + \alpha L - \delta L^{12} - \alpha\delta L^{13})$$

$$y_t = \mu + \beta y_{t-1} + \gamma y_{t-12} - \beta\gamma y_{t-13} +$$

$$+ u_t + \alpha u_{t-1} - \delta u_{t-12} - \alpha\delta u_{t-13}$$

Сезонная ARMA

$SARMA(p,q)-(P,Q)[12]$

- p – порядок AR-части, q – порядок MA-части
- P – порядок SAR-части, Q – порядок SMA-части
- $[12]$ – период сезонности (месячные данные)

$SARMA(p,q)-(P,Q)[12]$:

$$B(L)\cdot B_s(L^{12})\,(y_t-\mu^*)=A(L)\cdot A_s(L^{12})u_t$$

SARMA(p,q)-(P,Q)[12]

Параметры модели: μ, σ^2, коэффициенты в полиномах

SARMA(p,q)-(P,Q)[12]

Параметры модели: μ, σ^2, коэффициенты в полиномах

- Можно аккуратно понять, как выглядит ковариационная матрица, и оценить модель методом максимального правдоподобия

SARMA(p,q)-(P,Q)[12]

Параметры модели: μ, σ^2, коэффициенты в полиномах

❗ Можно аккуратно понять, как выглядит ковариационная матрица, и оценить модель методом максимального правдоподобия

- Точечные и интервальные прогнозы строятся по аналогии с тем, что мы делали с $AR(p)$ и $MA(q)$

SARMA(p,q)-(P,Q)[12]

Параметры модели: μ, σ^2, коэффициенты в полиномах

❗ Можно аккуратно понять, как выглядит ковариационная матрица, и оценить модель методом максимального правдоподобия

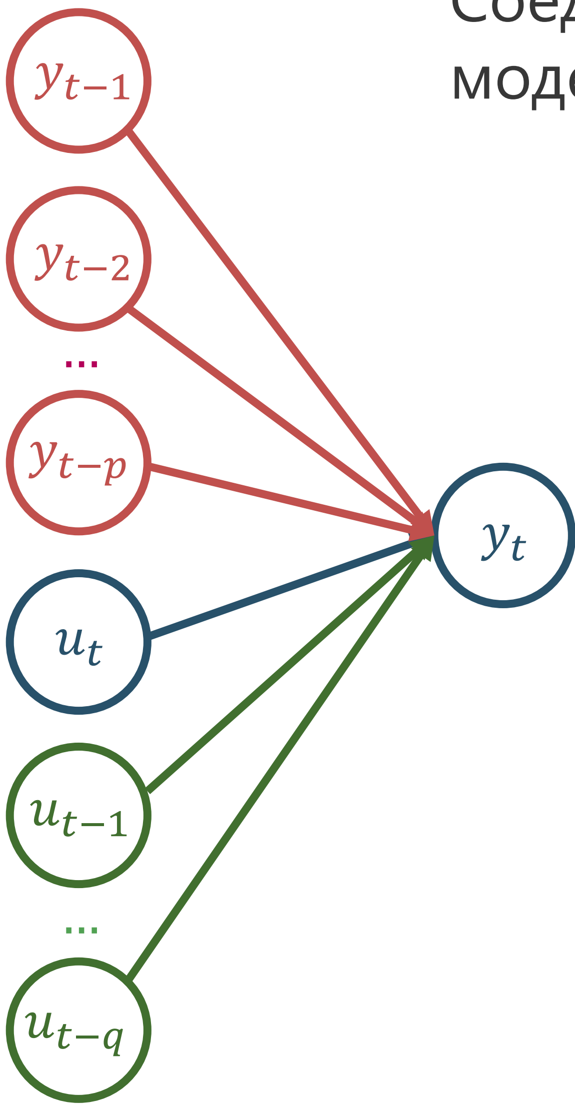
- Точечные и интервальные прогнозы строятся по аналогии с тем, что мы делали с $AR(p)$ и $MA(q)$
- Модель хороша для краткосрочных прогнозов, так как оптимальный прогноз на h шагов вперёд, при $h \to \infty$ довольно быстро сходится к математическому ожиданию процесса

SARMA(p,q)-(P,Q)[12]

Из-за того, что корреляции довольно сложно выражаются через коэффициенты модели, она неинтерпретируема

SARIMA

ARMA(p,q)



Соединим $AR(p)$ и $MA(q)$ в одну модель:

$u_t \sim iid\ N(0, \sigma^2)$

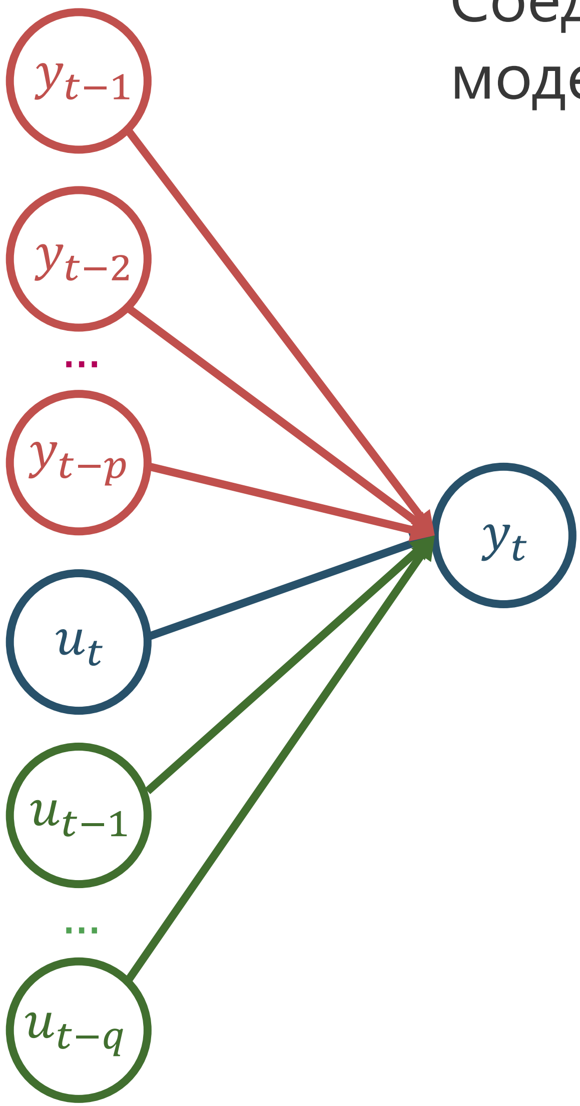
u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

У многочленов $A(L)$ и $B(L)$ нет общих корней

ARMA(p,q)



Соединим $AR(p)$ и $MA(q)$ в одну модель:

$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – стационарный процесс

$$B(L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

У многочленов $A(L)$ и $B(L)$ нет общих корней

А что делать, если нет?

Стационарность и разности

- Ряд y_t может быть нестационарным
- Если взять первую разность, Δy_t, он может стать стационарным

Пример:

$$y_t = y_{t-1} + u_t;\ \ u_t \sim iid\ N(0, \sigma^2);\ \ y_0 = 0$$

Мы с вами до этого выяснили, что такая модель нестационарна, так как $Var(y_t)$ зависит от времени

$$z_t = \Delta y_t = y_t - y_{t-1} = u_t$$

Процесс z_t будет стационарен

Стационарность и разности

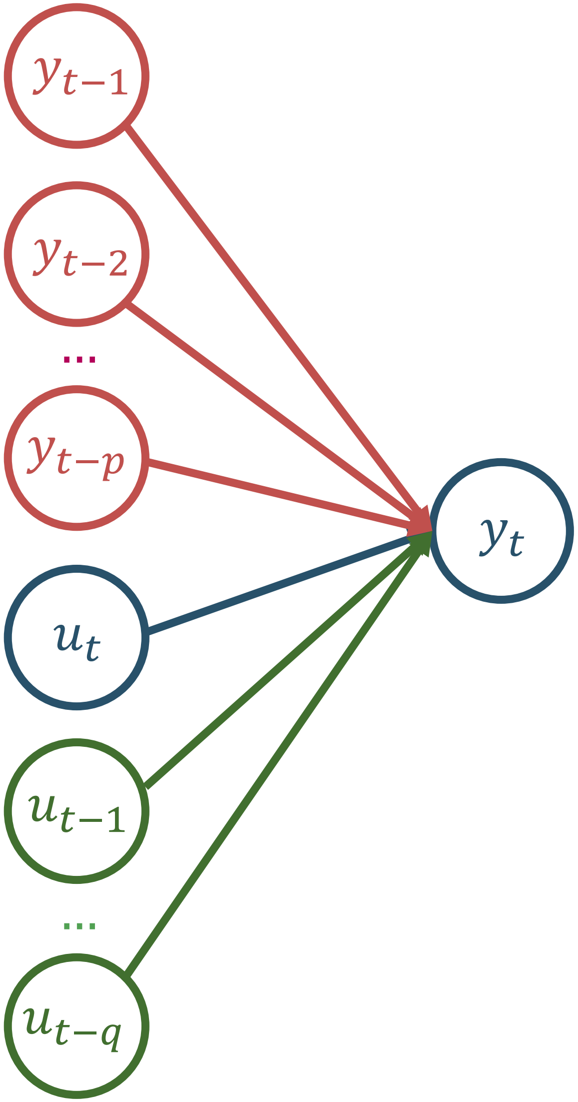
- Ряд y_t может быть нестационарным
- Если взять первую разность, Δy_t, он может стать стационарным
- Такой приём позволяет дёшево добавить нестационарные ряды в мир ARMA-моделей
- Конечно же, взятие разности помогает не всегда

Стационарность и разности

y_t

Δy_t

ARIMA(p,1,q)

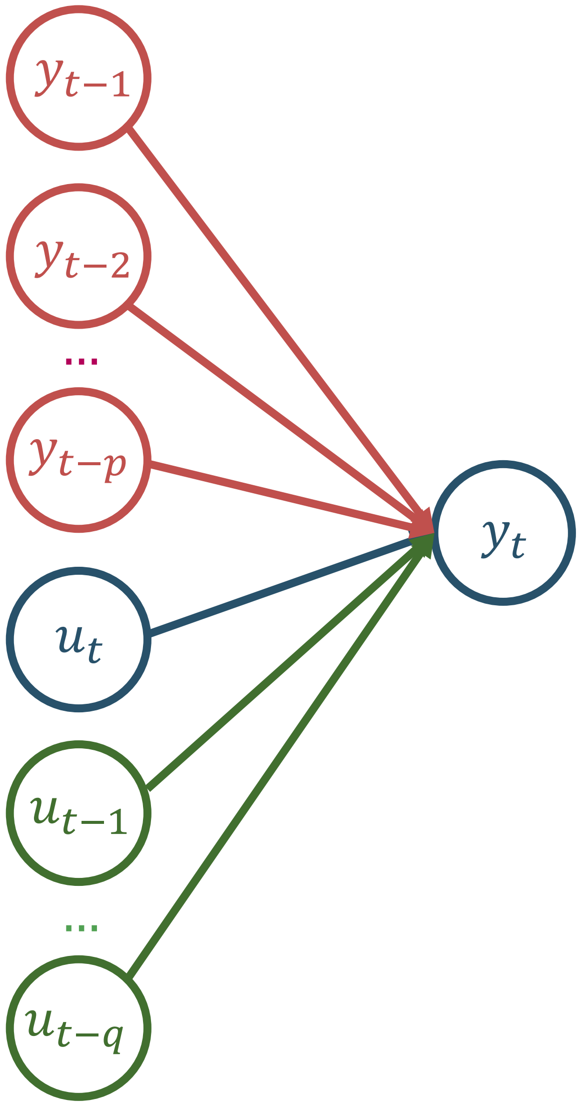


$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – **не**стационарный процесс

ARIMA(p,1,q)



$u_t \sim iid\ N(0, \sigma^2)$

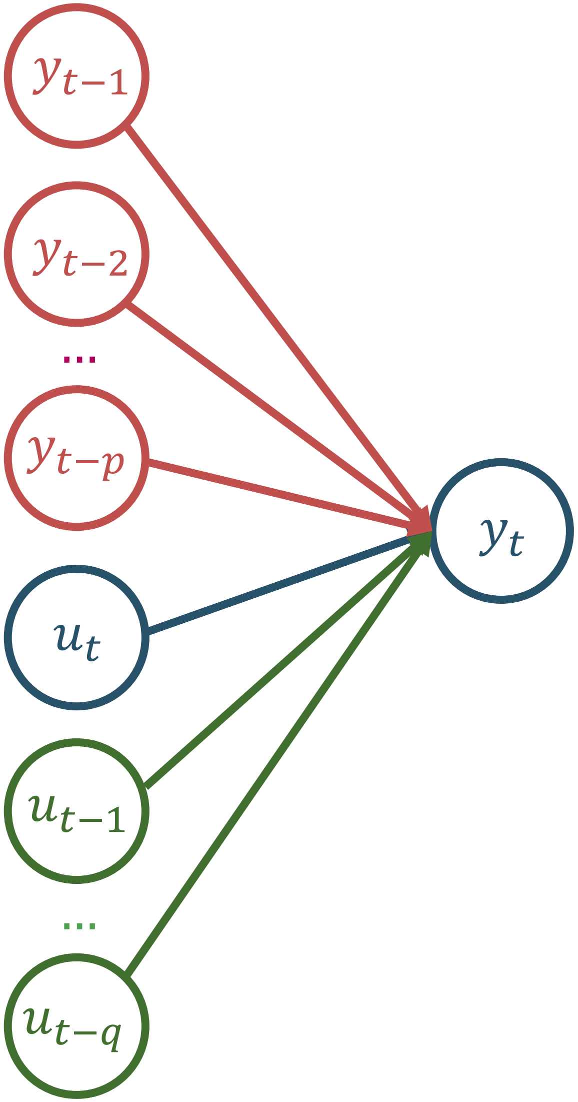
u_t не зависит от $y_{t-1}, y_{t-2}, \dots$

y_t – **не**стационарный процесс

Δy_t – стационарный процесс

$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$

ARIMA(p,1,q)



$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

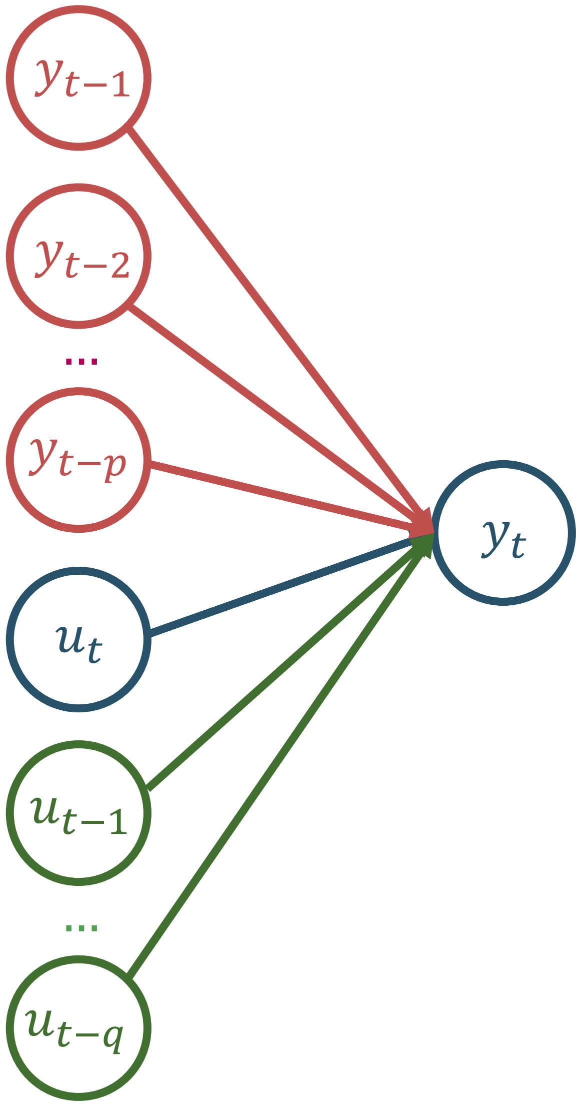
y_t – **не**стационарный процесс

Δy_t – стационарный процесс

$\Delta y_t = y_t - y_{t-1} = (1 - L) y_t$

$$B(L) \cdot (1 - L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

ARIMA(p,1,q)



$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – **не**стационарный процесс

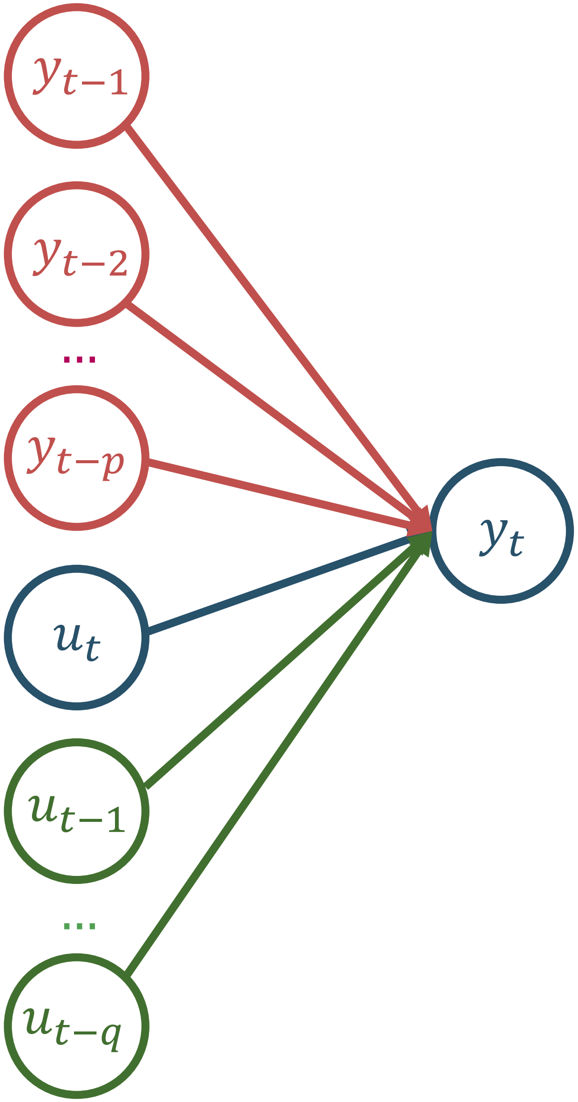
Δy_t – стационарный процесс

$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$

$$B(L) \cdot (1 - L) \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

$$B(L)\,\Delta y_t = A(L) \cdot u_t$$

ARIMA(p,d,q)



$u_t \sim iid\ N(0, \sigma^2)$

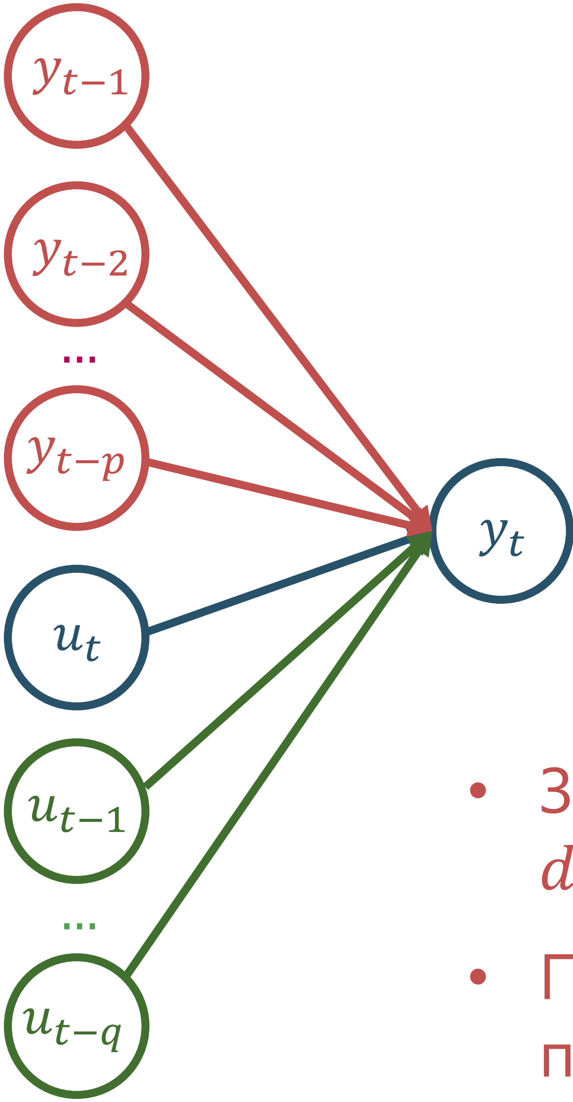
u_t не зависит от $y_{t-1}, y_{t-2}, \dots$

y_t – **не**стационарный процесс

$\Delta^d y_t$ – стационарный процесс

$$B(L) \cdot (1 - L)^d \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

ARIMA(p,d,q)



$u_t \sim iid\ N(0, \sigma^2)$

u_t не зависит от $y_{t-1}, y_{t-2}, \ldots$

y_t – **не**стационарный процесс

$\Delta^d y_t$ – стационарный процесс

$$B(L) \cdot (1 - L)^d \cdot (y_t - \mu^*) = A(L) \cdot u_t$$

- Здравый смысл подсказывает, что d не может превышать 2
- Первые разности – переход к приростам, вторые – к ускорениям

SARIMA(p, d, q) – (P,D,Q)[12]

$$B(L)\cdot B_s(L^{12})(1-L)^d(1-L^{12})^D(y_t-\mu^*)=A(L)\cdot A_s(L^{12})u_t$$

Можно добавить в модель
ещё и сезонные разности

SARIMA(p, d, q) – (P,D,Q)[12]

$$B(L) \cdot B_s(L^{12})(1-L)^d(1-L^{12})^D(y_t-\mu^*) = A(L) \cdot A_s(L^{12})u_t$$

Можно добавить в модель
ещё и сезонные разности

$$(1-L^{12})y_t = y_t - L^{12}y_t = y_t - y_{t-12}$$

Отличие от аналогичного
периода прошлого года

SARIMA(p, d, q) – (P,D,Q)[12]

$$B(L) \cdot B_s(L^{12})(1 - L)^d(1 - L^{12})^D(y_t - \mu^*) = A(L) \cdot A_s(L^{12})u_t$$

- p – порядок AR-части, q – порядок MA-части
- P – порядок SAR-части, Q – порядок SMA-части
- $[12]$ – период сезонности (месячные данные)
- d – то, сколько раз нам пришлось брать разность, чтобы ряд стал стационарным **(порядок интегрированности ряда)**
- D – то, сколько раз нам пришлось брать сезонную разность, чтобы ряд стал стационарным **(порядок сезонной интегрированности ряда)**

Как SARIMA работает на практике

1. Проверяем гипотезы о стационарности ряда:
   - Если ряд нестационарен, берём первую разность и проверяем на стационарность её
   - Делаем так, пока не получим стационарный ряд
   - Для подбора d обычно используют KPSS-тест
   - Для подбора D обычно используют OCSB-тест

Как SARIMA работает на практике

2* Как только мы получили стационарный ряд, перебираем гиперпараметры:

- Можно попробовать построить выборочные оценки автокорреляций и частных автокорреляций
- По их графикам принять решения о том, какими должны быть гиперпараметры p, q, P, Q
- Мы не будем говорить про этот способ

Как SARIMA работает на практике

2. Как только мы получили стационарный ряд, перебираем гиперпараметры:
   - Выбираем p, q, P, Q либо минимизируя информационный критерий, либо с помощью кросс-валидации
   - По аналогии решаем надо ли включать в модель константу

Как SARIMA работает на практике

3. Проверяем модель на адекватность: стабильная дисперсия, нормальность остатков

Как SARIMA работает на практике

3. Проверяем модель на адекватность: стабильная дисперсия, нормальность остатков

4. Если модель неадекватна, преобразовываем исходный ряд (например, преобразование Бокса-Кокса)

Гипотезы о стационарности

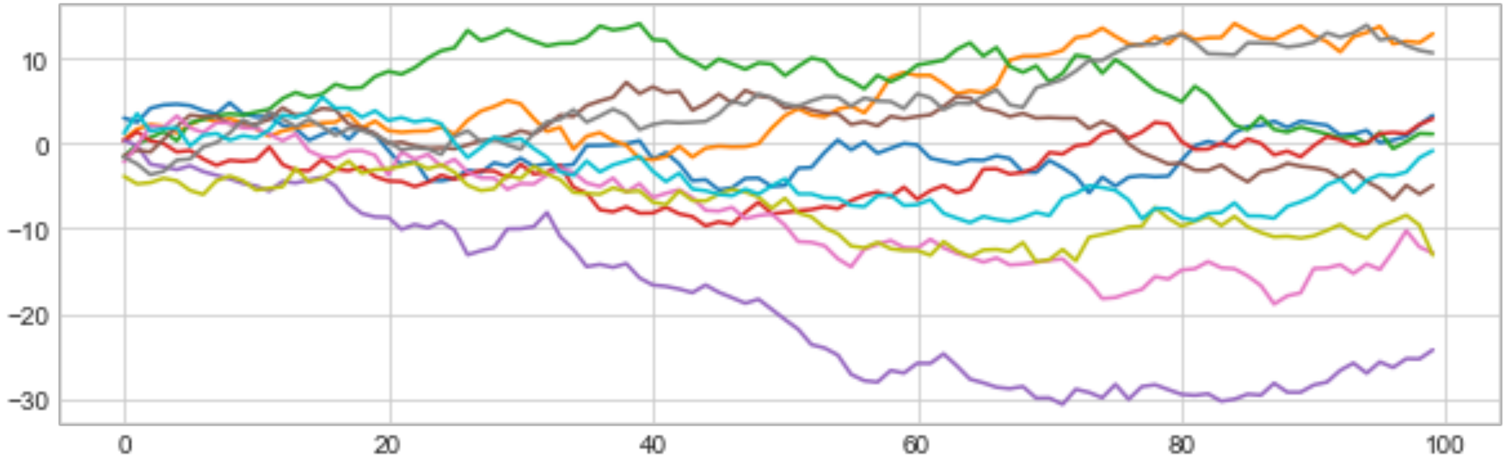
Зачем их проверять

- Мы хотим дёшево и сердито добавить в мир SARIMA-моделей нестационарные процессы
- Для этого нам нужен инструмент, который позволит искать нестационарные ряды

Спецификация теста

- Одна из самых распространенных нестационарных моделей – случайное блуждание:

$$y_t = y_{t-1} + u_t$$



- Дисперсия случайного блуждания растёт со временем
- Тест на стационарность, основанный на случайном блуждании называется **тестом на единичный корень**

Спецификация теста

- Тест на стационарность, основанный на случайном блуждании называется **тестом на единичный корень**

$$y_t = 1 \cdot y_{t-1} + u_t$$

Спецификация теста

- Тест на стационарность, основанный на случайном блуждании называется **тестом на единичный корень**

$$y_t = 1 \cdot y_{t-1} + u_t$$

В случайном блуждании коэффициент перед запаздыванием равен 1

Спецификация теста

- Тест на стационарность, основанный на случайном блуждании называется **тестом на единичный корень**

$$y_t = 1 \cdot y_{t-1} + u_t$$

В случайном блуждании коэффициент перед запаздыванием равен 1

$$(1 - L) \cdot y_t = u_t$$

$$A(\lambda) = 1 - \lambda = 0$$

$$\lambda = 1$$

- У характеристического многочлена AR(1)-процесса есть единичный корень, значит процесс нестационарен

Спецификация теста

Нестационарная модель может выглядеть сложнее:

- У неё может быть ненулевое среднее

$$y_t = \mu + y_{t-1} + u_t$$

- В модели может быть тренд

$$y_t = k \cdot t + y_{t-1} + u_t$$

Тест Дики-Фуллера

- Тест Дики-Фуллера проверяет гипотезу о том, что в AR-части процесса есть единичный корень:

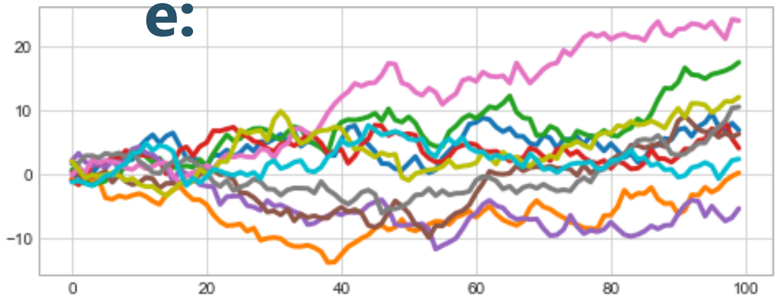
H_0: процесс нестационарен (есть единичный корень)

H_a: процесс стационарен (единичного корня нет)

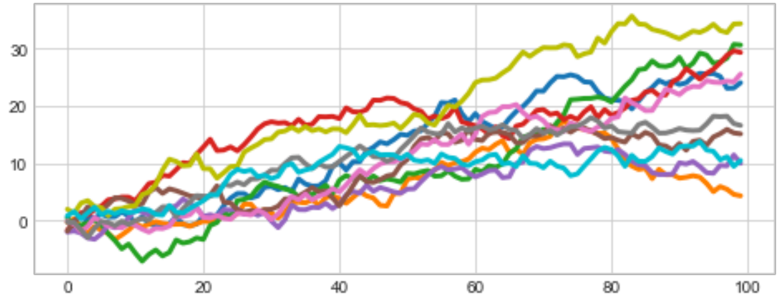
Тест Дики-Фуллера

Для каждой спецификации есть свой тест:

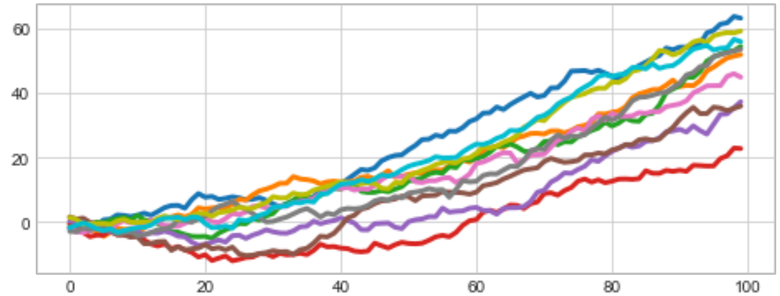
Нестационарные:



$$y_t = y_{t-1} + u_t$$



$$y_t = \mu + y_{t-1} + u_t$$

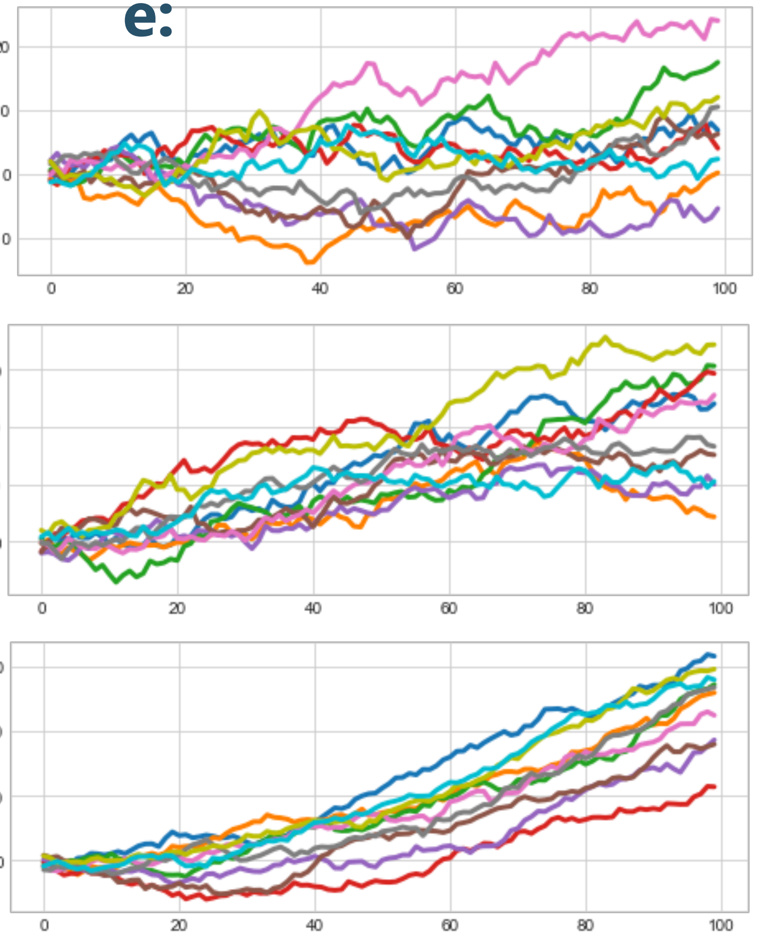


$$y_t = \mu + at + y_{t-1} + u_t$$

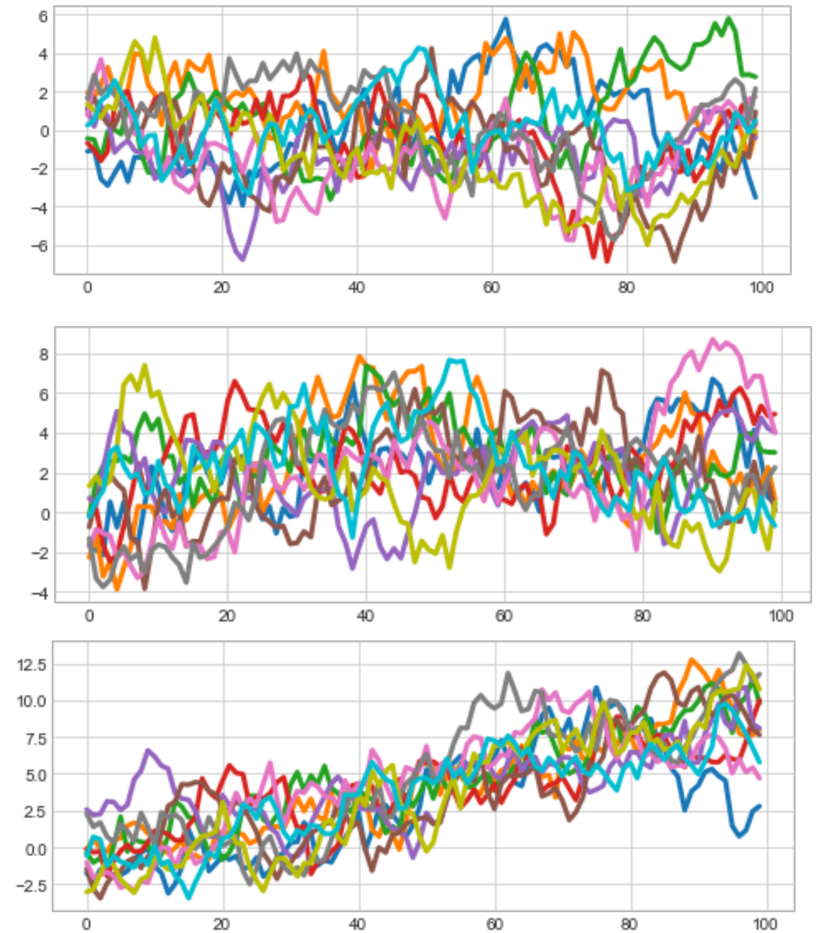
Тест Дики-Фуллера

Для каждой спецификации есть свой тест:

Нестационарные:



Стационарные:

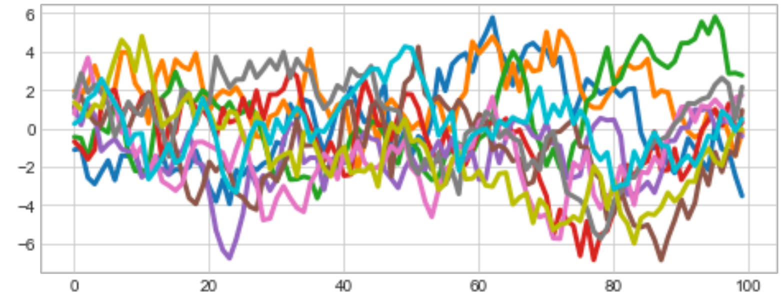


Тест Дики-Фуллера

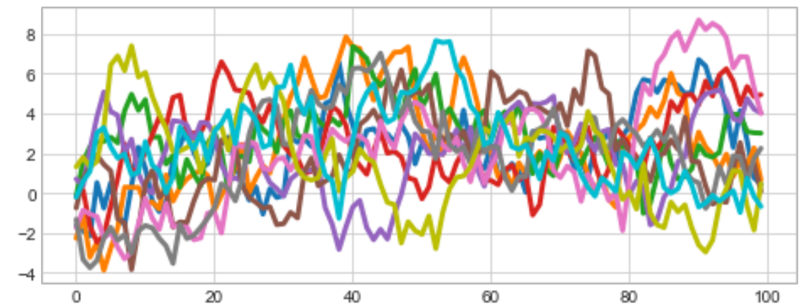
Для каждой спецификации есть свой тест:

Стационарные:

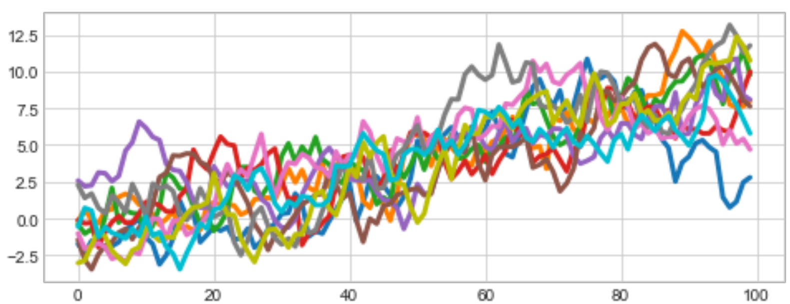
$$y_t = 0.9 \cdot y_{t-1} + u_t$$



$$y_t = \mu + 0.9 \cdot y_{t-1} + u_t$$



$$y_t = \mu + at + 0.9 \cdot y_{t-1} + u_t$$



Тест Дики-Фуллера

- Спецификацию теста можно выбрать с помощью визуального анализа
- Также есть более формальные процедуры, например многовариантная процедура проверки гипотезы единичного корня Доладо

Расширенный тест Дики-Фуллера (ADF)

- Если процесс специфицирован как AR(p), работают все те же принципы:

$$B(L) \cdot y_t = u_t,$$

- Если у характеристического многочлена есть единичные корни, процесс нестационарный

$$B(\lambda) = 0$$

$$1 - \beta_1\lambda - \beta_2\lambda^2 - \cdots - \beta_p\lambda^p = 0$$

- Тест, специфицированный для такой модели называется расширенным критерием Дики-Фуллера (Adjusted Dickey-Fuller, ADF)

KPSS-тест

- У ADF-теста низкая мощность, из-за этого придумано много других тестов на стационарность
- Один из самых популярных, KPSS-тест, его придумали Квятовский, Филлипс, Шмидт и Шин (Kwiatkowski, Phillips, Schmidt, Shin)
- В этом тесте нулевая и альтернативная гипотеза меняются местами
- В нулевой гипотезе предполагается, что процесс стационарен, в альтернативной, что у него есть единичный корень

OCSB-тест

- В сезонной составляющей тоже могут быть единичные корни
- Один из самых популярных тестов для гипотезы о сезонной стационарности – OCSB-тест
- Его придумали Осборн, Чуи, Смит, Бирченхол (Osborn, Chui, Smith, Birchenhall)
- В нулевой гипотезе предполагается, что процесс нестационарен

Как сделать ряд стационарным

Есть несколько разных приёмов, которыми можно пользоваться, перечислим самые простые:

- Взять ряд в разностях
- Очистить ряд от тренда, либо в явном виде включить его в модель как отдельную переменную
- Если есть проблемы с дисперсией остатков, её можно стабилизировать с помощью преобразования Бокса-Кокса
- Также преобразование Бокса-Кокса часто используют, чтобы добиться нормальности остатков и корректных интервальных прогнозов

Резюме

- Для проверки гипотезы о стационарности временного ряда разработано довольно большое число разных тестов
- Каждый тест это теорема, которая описывает как себя ведёт распределение тестовой статистики
- Концептуально мы пытаемся понять, какая из двух моделей (стационарная или нестационарная) более правдоподобна для данных
- Мы будем использовать тест Дики-Фуллера и KPSS-тест. Для распределения их статистик составлены таблицы критических значений.

TBATS

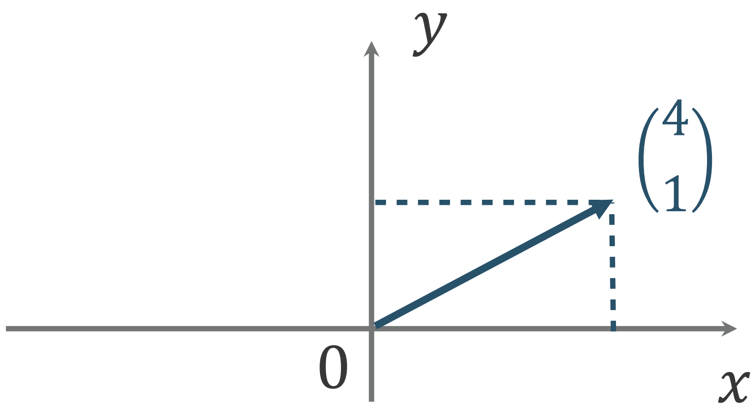
Тригонометрическое моделирование сезонности

- При моделировании сезонности возникают проблемы, связанные с длительностью периодов
- В году в среднем 52.18 недели
- В году в среднем 365.25 дней
- Даже с месячной сезонностью возникают проблемы, так как месяца обычно разной длины

Тригонометрическое моделирование сезонности позволяет решить эту проблему

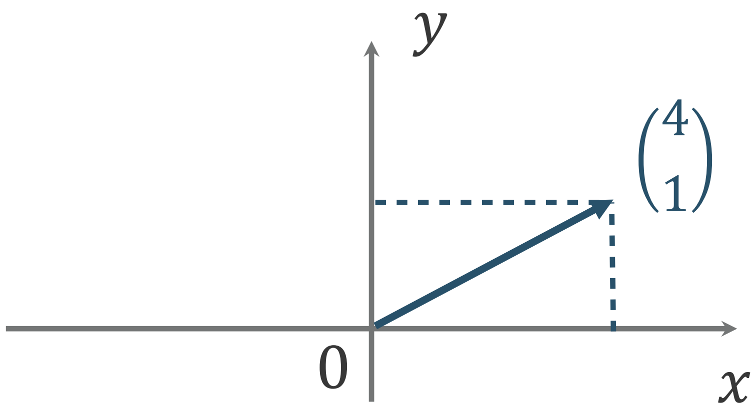
Матрица поворота

- Любая матрица задаёт какое-то преобразование линейного пространства



Матрица поворота

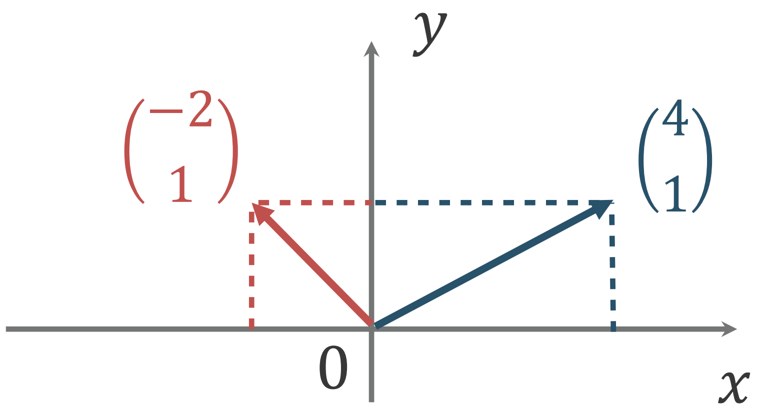
- Любая матрица задаёт какое-то преобразование линейного пространства



$$\begin{bmatrix} -0.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

Матрица поворота

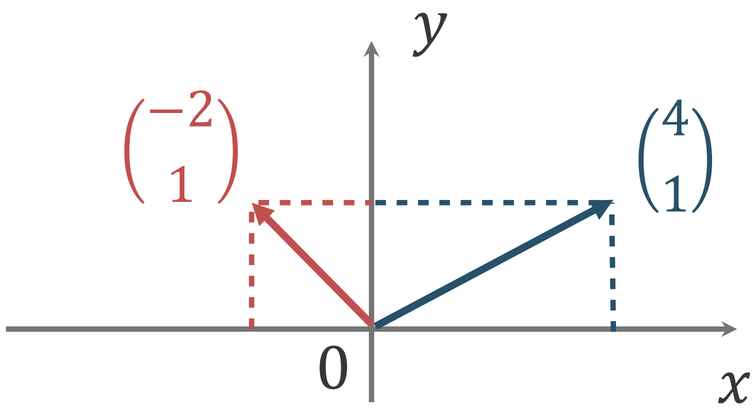
- Любая матрица задаёт какое-то преобразование линейного пространства



$$\begin{bmatrix}-0.5 & 0\\0 & 1\end{bmatrix}\begin{pmatrix}4\\1\end{pmatrix}=\begin{pmatrix}-2\\1\end{pmatrix}$$

Матрица поворота

- Любая матрица задаёт какое-то преобразование линейного пространства



$$\begin{bmatrix}-0.5 & 0\\ 0 & 1\end{bmatrix}\begin{pmatrix}4\\1\end{pmatrix}=\begin{pmatrix}-2\\1\end{pmatrix}$$

Такая матрица задаёт зеркальное отображение относительно оси y и сжатие по оси x в два раза

Матрица поворота

- По аналогии можно задать матрицу для поворотов:

$$\begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix}$$

- λ – угол поворота
- Если мы хотим сезонность с периодом 12, мы можем взять углы:

$$\frac{2\pi}{12}, \frac{4\pi}{12}, \frac{6\pi}{12}, \dots, \frac{24\pi}{12}$$

- Тогда за 12 шагов мы сделаем полный поворот и вернёмся в исходную точку

Матрица поворота

- По аналогии можно задать матрицу для поворотов:

$$\begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix}$$

- Запишем динамику нашей точки через матрицу поворотов:

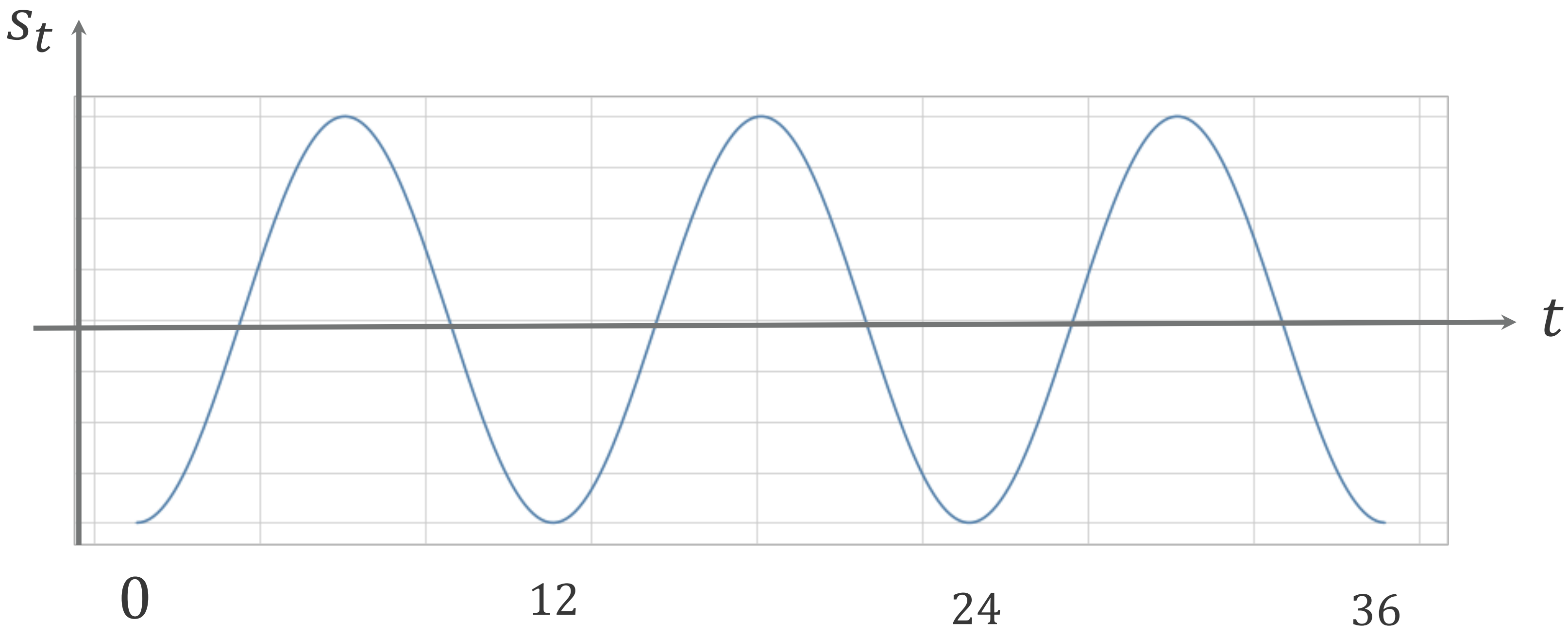
$$\begin{pmatrix} s_t \\ s_t^* \end{pmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \cdot \begin{pmatrix} s_{t-1} \\ s_{t-1}^* \end{pmatrix}$$

- У точки на плоскости две координаты, мы наблюдаем за движением координаты s_t, вторая координата s_t^* для нас важной роли не играет

Матрица поворота

- Тогда динамика s_t будет выглядеть следующим образом:

$$\begin{pmatrix} s_t \\ s_t^* \end{pmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \cdot \begin{pmatrix} s_{t-1} \\ s_{t-1}^* \end{pmatrix}$$



Матрица поворота

- Добавим к уравнению случайную ошибку:

$$\begin{pmatrix} s_t \\ s_t^* \end{pmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \cdot \begin{pmatrix} s_{t-1} \\ s_{t-1}^* \end{pmatrix} + \begin{pmatrix} u_t \\ u_t^* \end{pmatrix}$$

- В реальных данных несколько частот могут накладываться друг на друга:

$$s_t^{total} = s_t(\lambda_1) + s_t(\lambda_2) + s_t(\lambda_3)$$

- В итоговую компоненту войдут полугодовые, месячные, квартальные частоты и так далее

TBATS

Попробуем объединить все те модели и подходы, которые мы знаем, в одного большого монстра

TBATS это

y_t – наблюдаемое значение ряда

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

Трансформация Бокса-Кокса

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\, y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

Темпы роста

Случайная ошибка

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

Долговременный уровень

Разные сезонные составляющие (всего M штук)

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\, y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1 - \phi) b + \phi b_{t-1} + \beta \cdot u_t$$

Долговременный уровень и темпы роста похожи на ETS-модель

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\, y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta \cdot u_t$$

В ETS не было этой части

Она тут для того, чтобы долгосрочный прогноз сходился не к нулю, а к какой-то константе

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\, y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta \cdot u_t$$

Ошибка ведёт себя как ARMA-процесс

$$u_t \sim ARMA(p, q)$$

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\, y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

$$u_t \sim ARMA(p, q)$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1 - \phi) b + \phi b_{t-1} + \beta \cdot u_t$$

$$\begin{pmatrix} s_t \\ s_t^* \end{pmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \cdot \begin{pmatrix} s_{t-1} \\ s_{t-1}^* \end{pmatrix} + \begin{pmatrix} \gamma_1 \cdot u_t \\ \gamma_2 \cdot u_t \end{pmatrix}$$

Тригонометрическое моделирование сезонности

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\ y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases}$$

$$u_t \sim ARMA(p,q)$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta \cdot u_t$$

$$\begin{pmatrix} s_t \\ s_t^* \end{pmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \cdot \begin{pmatrix} s_{t-1} \\ s_{t-1}^* \end{pmatrix} + \begin{pmatrix} \gamma_1 \cdot u_t \\ \gamma_2 \cdot u_t \end{pmatrix}$$

$$s_t = s_{t-1} \cdot \cos\lambda + s_{t-1}^* \cdot \sin\lambda + \gamma_1 u_t$$

$$s_t^* = -s_{t-1} \cdot \sin\lambda + s_{t-1}^* \cdot \cos\lambda + \gamma_2 u_t$$

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\ y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases} \qquad u_t \sim ARMA(p,q)$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta \cdot u_t$$

k_i – число гармоник, необходимое для описания i-ой сезонной компоненты

$$s_t = s_{t-1} \cdot \cos\lambda + s_{t-1}^* \cdot \sin\lambda + \gamma_1 u_t$$

$$s_t^* = -s_{t-1} \cdot \sin\lambda + s_{t-1}^* \cdot \cos\lambda + \gamma_2 u_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_t(\lambda_j^{(i)}) \qquad \lambda_j^{(i)} = \frac{j \cdot 2\pi}{m_i}$$

TBATS это

y_t – наблюдаемое значение ряда

$$y_t^* = \begin{cases} \log\, y_t, & p = 0 \\ \dfrac{y_t^p - 1}{p}, & 0 \le p \le 1 \end{cases} \qquad u_t \sim ARMA(p, q)$$

$$y_t^* = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m_i}^{(i)} + u_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha \cdot u_t$$

$$b_t = (1 - \phi) b + \phi b_{t-1} + \beta \cdot u_t$$

$$s_t = s_{t-1} \cdot \cos \lambda + s_{t-1}^* \cdot \sin \lambda + \gamma_1 u_t$$

$$s_t^* = -s_{t-1} \cdot \sin \lambda + s_{t-1}^* \cdot \cos \lambda + \gamma_2 u_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_t(\lambda_j^{(i)}) \qquad \lambda_j^{(i)} = \frac{j \cdot 2\pi}{m_i}$$

➤ *https://robjhyndman.com/papers/ComplexSeasonality.pdf*

TBATS это

- Trigonometric (тригонометрическая сезонность)
- Box-Cox (преобразование бокса-кокса для стабилизации дисперсии)
- ARMA (описывает краткосрочную динамику остаточной компоненты)
- Trend (в модели есть трендовая компонента)
- Seasonal (сезонная компонента)
- Модель реализована в python в пакете tbats

Усложнение моделей

Модели можно продолжить усложнять:

- Добавлять дополнительные уравнения, которые объясняют динамику дисперсии: ARCH и GARCH – модели (полезны для финансовых рядов)
- Сделать взаимосвязи нелинейными, добавить другие преобразования: prophet от facebook
- Добавлять новые “ненаблюдаемые” переменные вроде долгосрочного уровня и сезонности
- Включать в модели экзогенные переменные
- Те модели, которые мы рассмотрели – частный случай моделей пространства-состояния (state-space models)

Векторные модели

Несколько временных рядов

- Все модели, о которых мы говорили до этого, работают с одним рядом
- В ARIMA в качестве экзогенной переменной можно добавлять значения другого ряда
- Такая модель будет строить прогнозы только для основного ряда
- Чтобы получать прогнозы сразу для обоих рядов, нам нужно второе уравнение $\Rightarrow$ векторные модели

Векторная авторегрессия (VAR)

- Векторная авторегрессия – обобщение авторегрессии на несколько рядов

Пример: VAR(1)

$$\begin{cases} y_{1,t} = \mu_1 + \gamma_{11} \cdot y_{1,t-1} + \gamma_{12} \cdot y_{2,t-1} + u_{1,t} \\ y_{2,t} = \mu_2 + \gamma_{21} \cdot y_{1,t-1} + \gamma_{22} \cdot y_{2,t-1} + u_{2,t} \end{cases}$$

Векторная авторегрессия (VAR)

- Векторная авторегрессия – обобщение авторегрессии на несколько рядов

Пример: VAR(1)

$$\begin{cases} y_{1,t} = \mu_1 + \gamma_{11} \cdot y_{1,t-1} + \gamma_{12} \cdot y_{2,t-1} + u_{1,t} \\ y_{2,t} = \mu_2 + \gamma_{21} \cdot y_{1,t-1} + \gamma_{22} \cdot y_{2,t-1} + u_{2,t} \end{cases}$$

В каждом уравнении есть запаздывания обоих рядов

К каждому уравнению добавляются свои случайные ошибки

Векторная авторегрессия (VAR)

- Векторная авторегрессия – обобщение авторегрессии на несколько рядов

Пример: VAR(1)

$$\begin{cases} y_{1,t} = \mu_1 + \gamma_{11} \cdot y_{1,t-1} + \gamma_{12} \cdot y_{2,t-1} + u_{1,t} \\ y_{2,t} = \mu_2 + \gamma_{21} \cdot y_{1,t-1} + \gamma_{22} \cdot y_{2,t-1} + u_{2,t} \end{cases}$$

$$y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} \quad \Gamma_1 = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \quad u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

$$y_t = \mu + \Gamma_1 \cdot y_{t-1} + u_t$$

Можно переписать систему в более удобном матричном виде

Векторная авторегрессия (VAR)

- Векторная авторегрессия – обобщение авторегрессии на несколько рядов

Пример: VAR(1)

$$y_t = \mu + \Gamma_1 \cdot y_{t-1} + u_t$$

$$u_t \sim iid\ N(0, \Sigma)$$

y_t – стационарный процесс

u_t не зависит от $y_{t-1}, y_{t-2}, \dots$

Векторная модели

- Векторные модели по аналогии с одномерными тоже можно усложнять

Как устроен мир и что будет завтра

- Мы с вами сконцентрировались на прогнозировании
- Однако для временных рядов также можно пытаться искать ответ на вопрос “Как устроен мир?”

Причинность по Грейнджеру

$$\begin{cases} y_{1,t} = \mu_1 + \gamma_{11} \cdot y_{1,t-1} + \gamma_{12} \cdot y_{2,t-1} + u_{1,t} \\ y_{2,t} = \mu_2 + \gamma_{21} \cdot y_{1,t-1} + \gamma_{22} \cdot y_{2,t-1} + u_{2,t} \end{cases}$$

Что, если этот коэффициент нулевой?

Причинность по Грейнджеру

$$\begin{cases} y_{1,t} = \mu_1 + \gamma_{11} \cdot y_{1,t-1} + \gamma_{12} \cdot y_{2,t-1} + u_{1,t} \\ y_{2,t} = \mu_2 + \gamma_{21} \cdot y_{1,t-1} + \gamma_{22} \cdot y_{2,t-1} + u_{2,t} \end{cases}$$

Предыдущее значение $y_{1,t}$ влияет на текущее значение $y_{2,t}$, но не наоборот

Причинность по Грейнджеру

$$\begin{cases} y_{1,t} = \mu_1 + \gamma_{11} \cdot y_{1,t-1} + \gamma_{12} \cdot y_{2,t-1} + u_{1,t} \\ y_{2,t} = \mu_2 + \gamma_{21} \cdot y_{1,t-1} + \gamma_{22} \cdot y_{2,t-1} + u_{2,t} \end{cases}$$

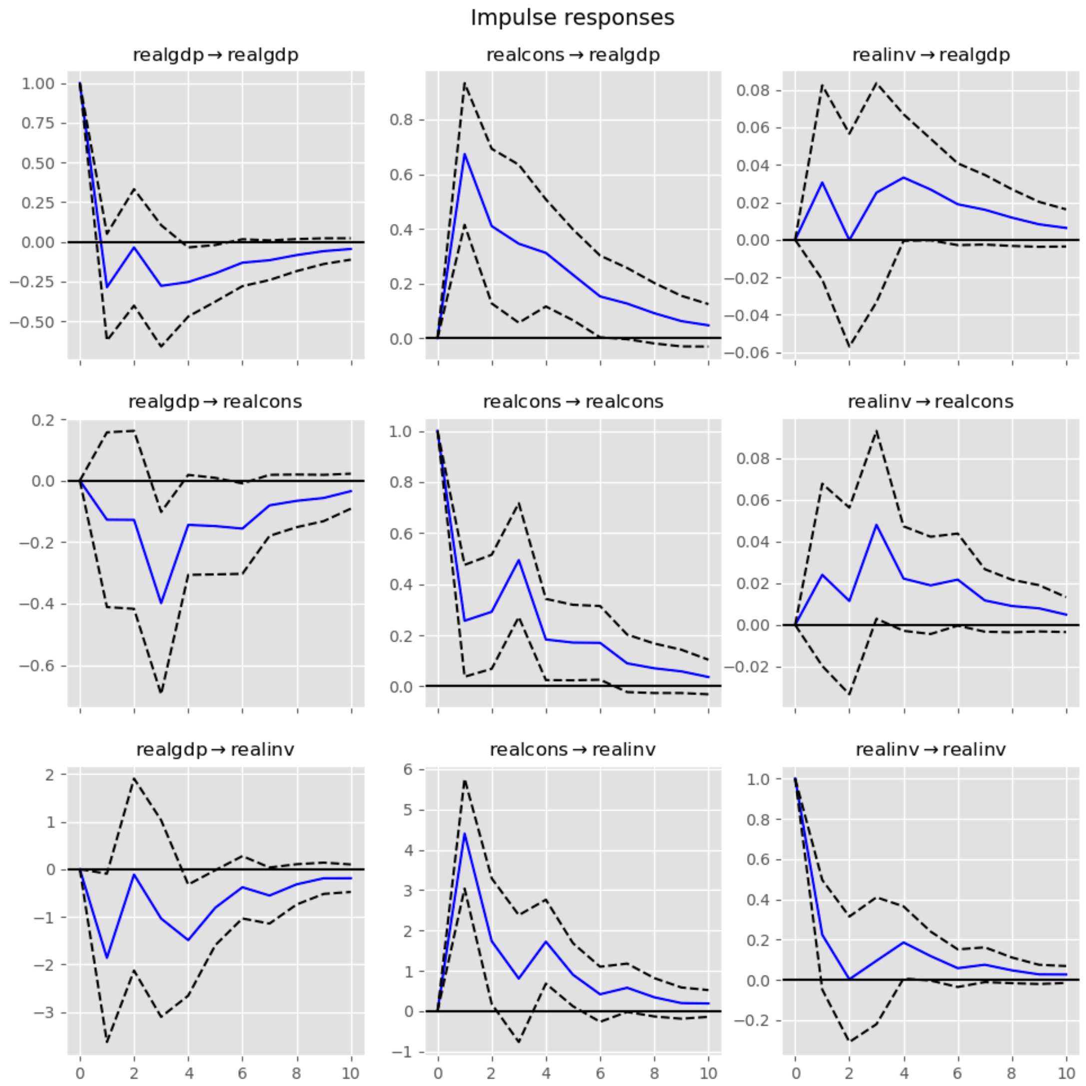
Предыдущее значение $y_{1,t}$ влияет на текущее значение $y_{2,t}$, но не наоборот

- В таких ситуациях говорят, что ряд $y_{1,t}$ является **причиной по Грейнджеру** для ряда $y_{2,t}$
- В динамике ряда $y_{1,t}$ есть информация, которая помогает спрогнозировать $y_{2,t}$, но не наоборот
- Причинности по Грейнджеру недостаточно для причинно-следственной связи

Импульсные отклики

- Другой подход сделать осмысленные выводы о взаимосвязях между рядами – исследование откликов системы на неожиданные шоки
- Для этого VAR переписывают в более удобной форме, накладывают на неё ограничения и строят **импульсные отклики**
- За этой процедурой скрывается много статистических тонкостей, про которые мы говорить не будем ☹
- Но на примеры откликов посмотрим

Импульсные отклики

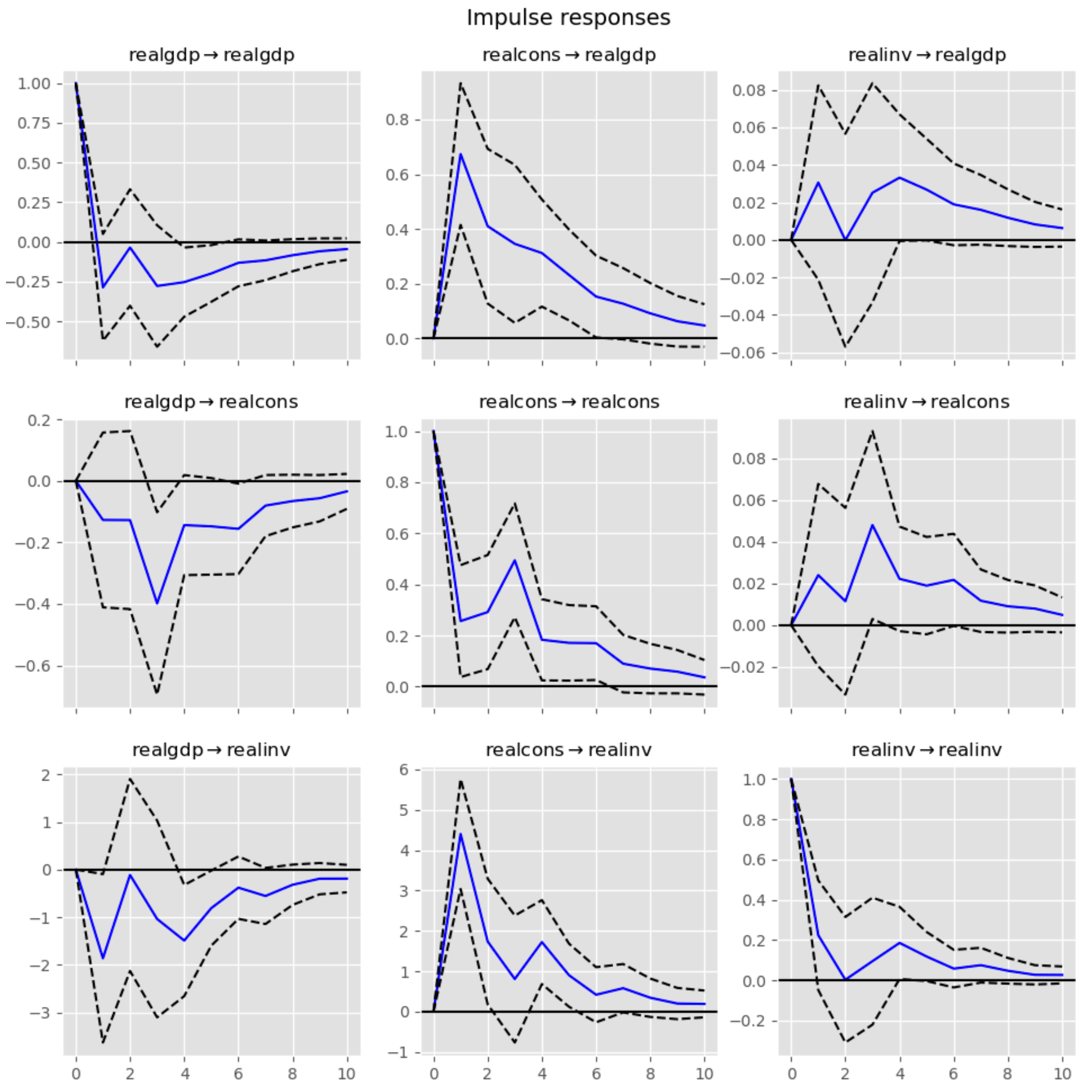


В системе было три уравнения:

- Для реального ВВП (realgdp)
- Для реального потребления (realcons)
- Для реальных инвестиций (realinv)

➤ *https://www.statsmodels.org/dev/vector_ar.html*

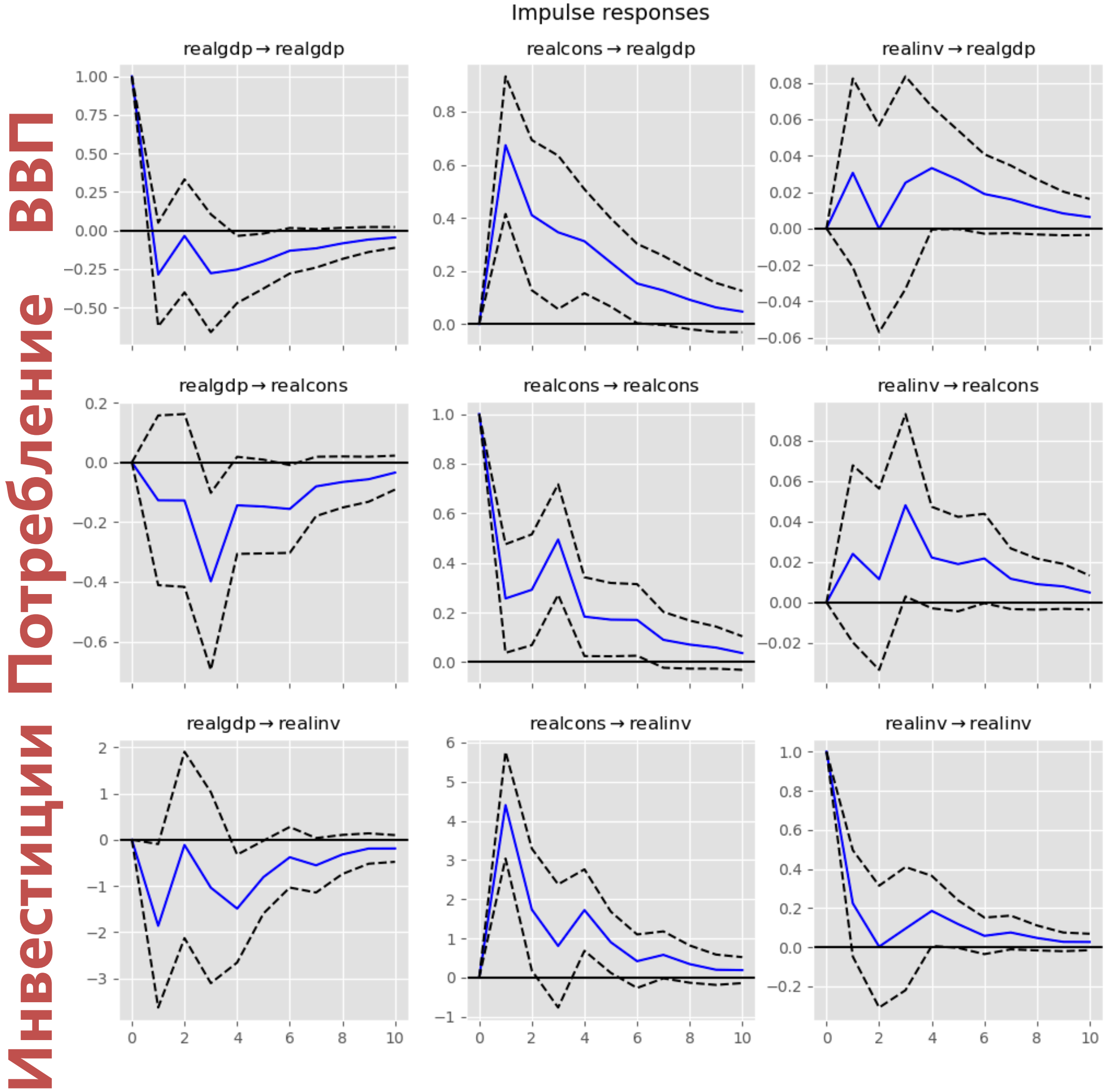
Импульсные отклики



- По оси x отложено время
- По оси y – изменение переменной при случайном шоке

➤ *https://www.statsmodels.org/dev/vector_ar.html*

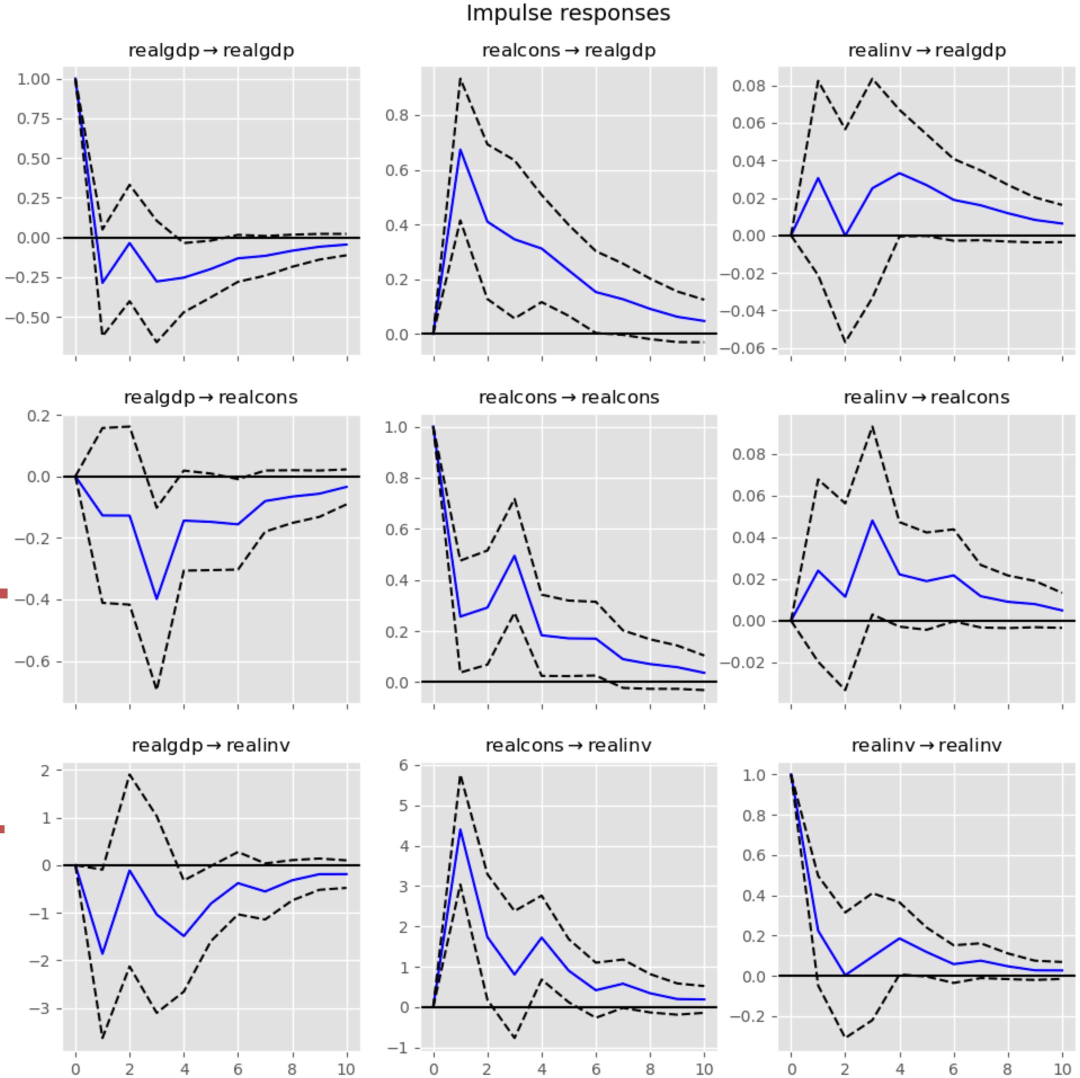
Импульсные отклики



По строкам откладывается переменная, в случайной ошибке которой произошёл шок

➤ *https://www.statsmodels.org/dev/vector_ar.html*

Импульсные отклики

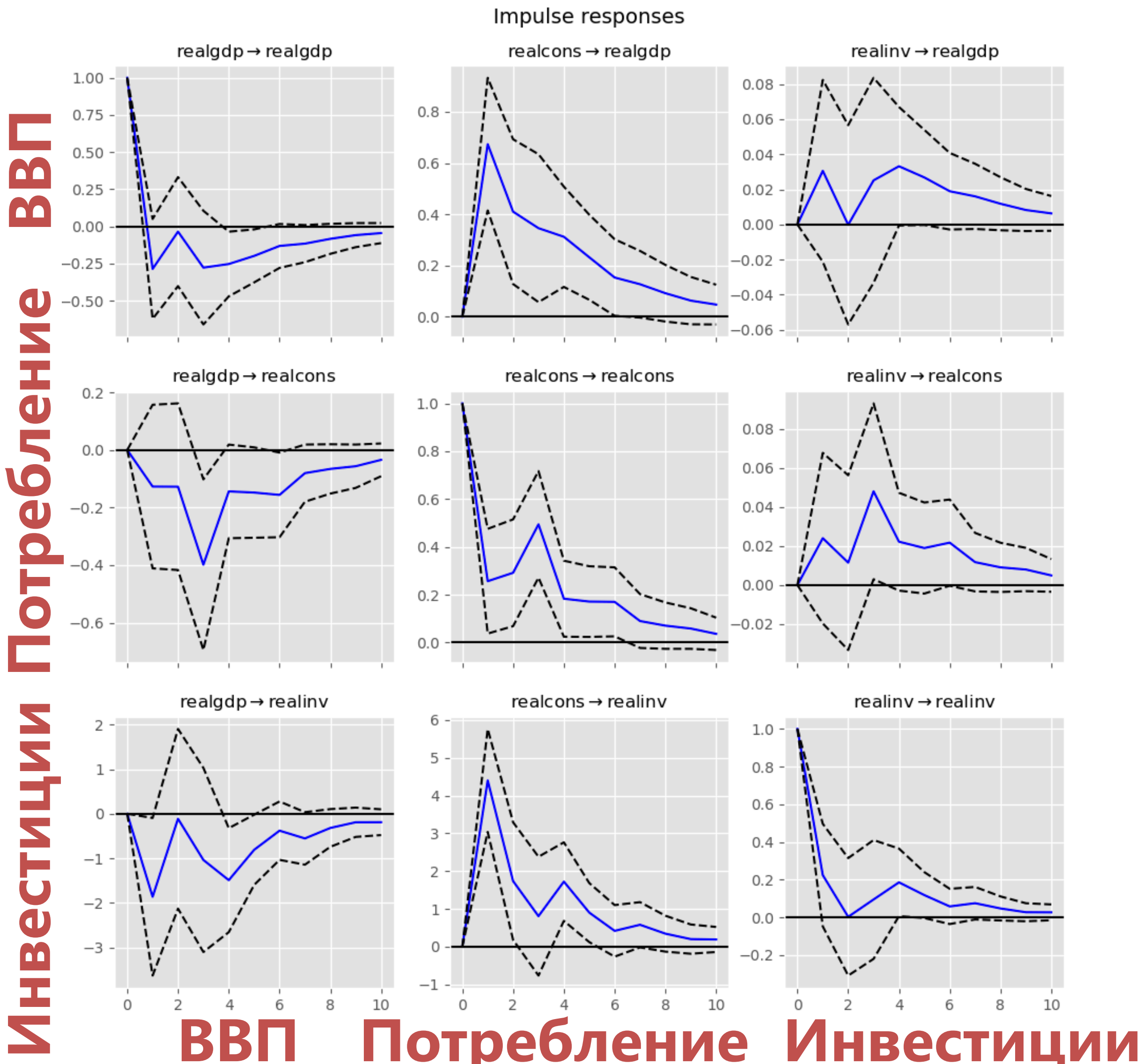


ВВП **Потребление** **Инвестиции**

По столбцам откладываются переменные, которые изменились под воздействием этого шока

➤ *https://www.statsmodels.org/dev/vector_ar.html*

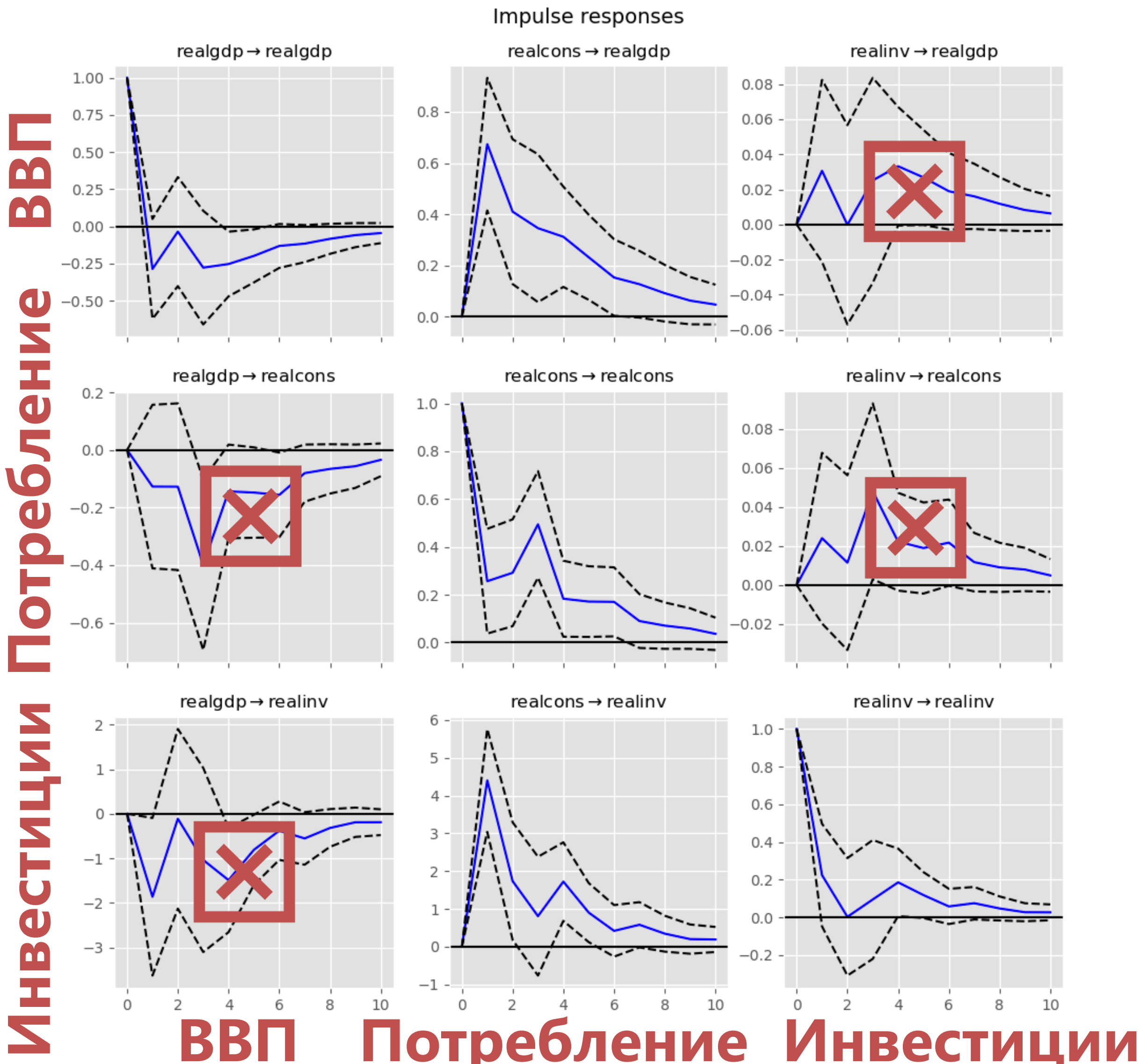
Импульсные отклики



- Голубая линия – импульсный отклик
- Пунктирная линия – доверительный интервал для него
- Если ноль внутри интервала, отклик незначим

➤ *https://www.statsmodels.org/dev/vector_ar.html*

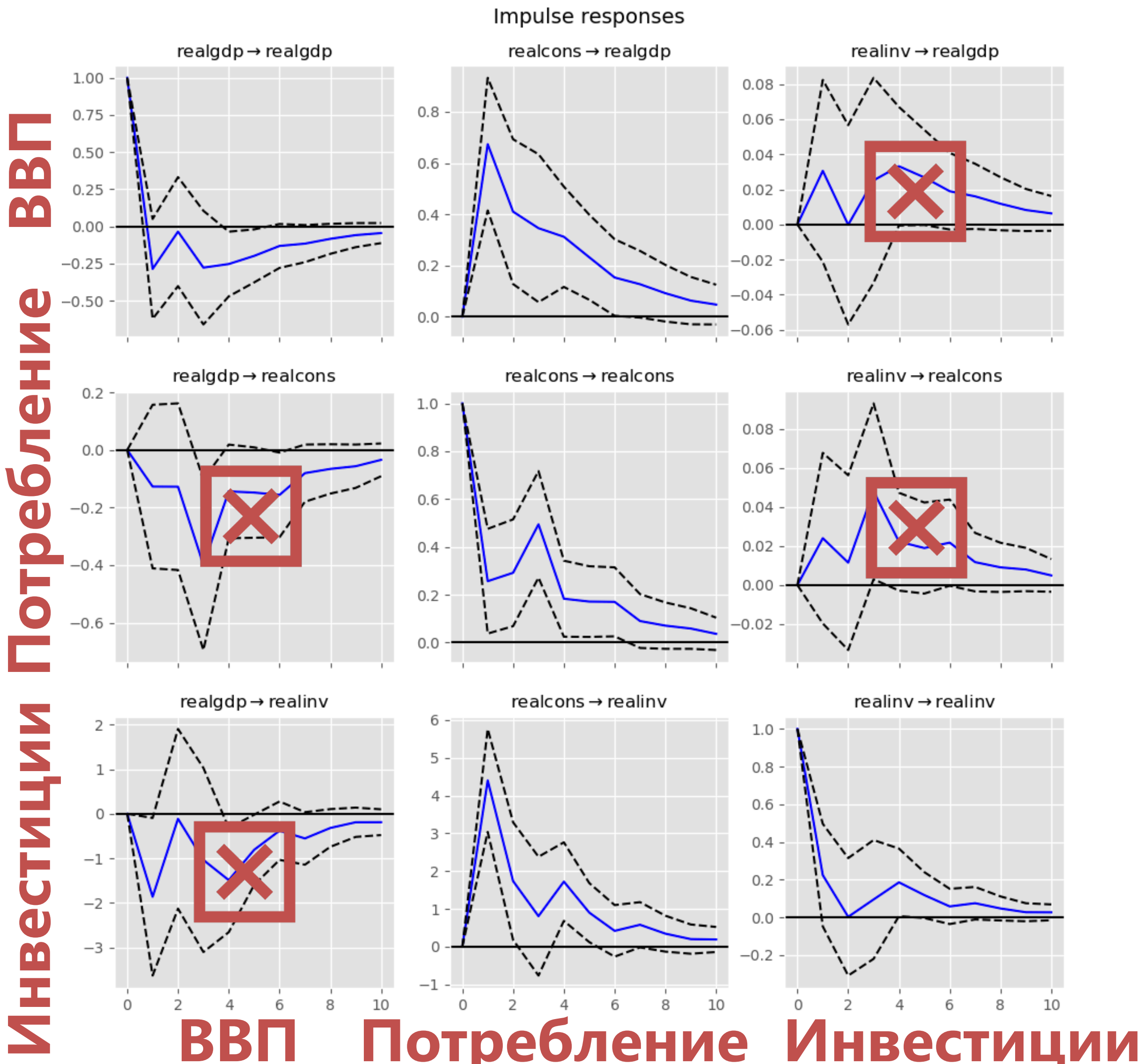
Импульсные отклики



➤ *https://www.statsmodels.org/dev/vector_ar.html*

- Голубая линия – импульсный отклик
- Пунктирная линия – доверительный интервал для него
- Если ноль внутри интервала, отклик незначим

Импульсные отклики

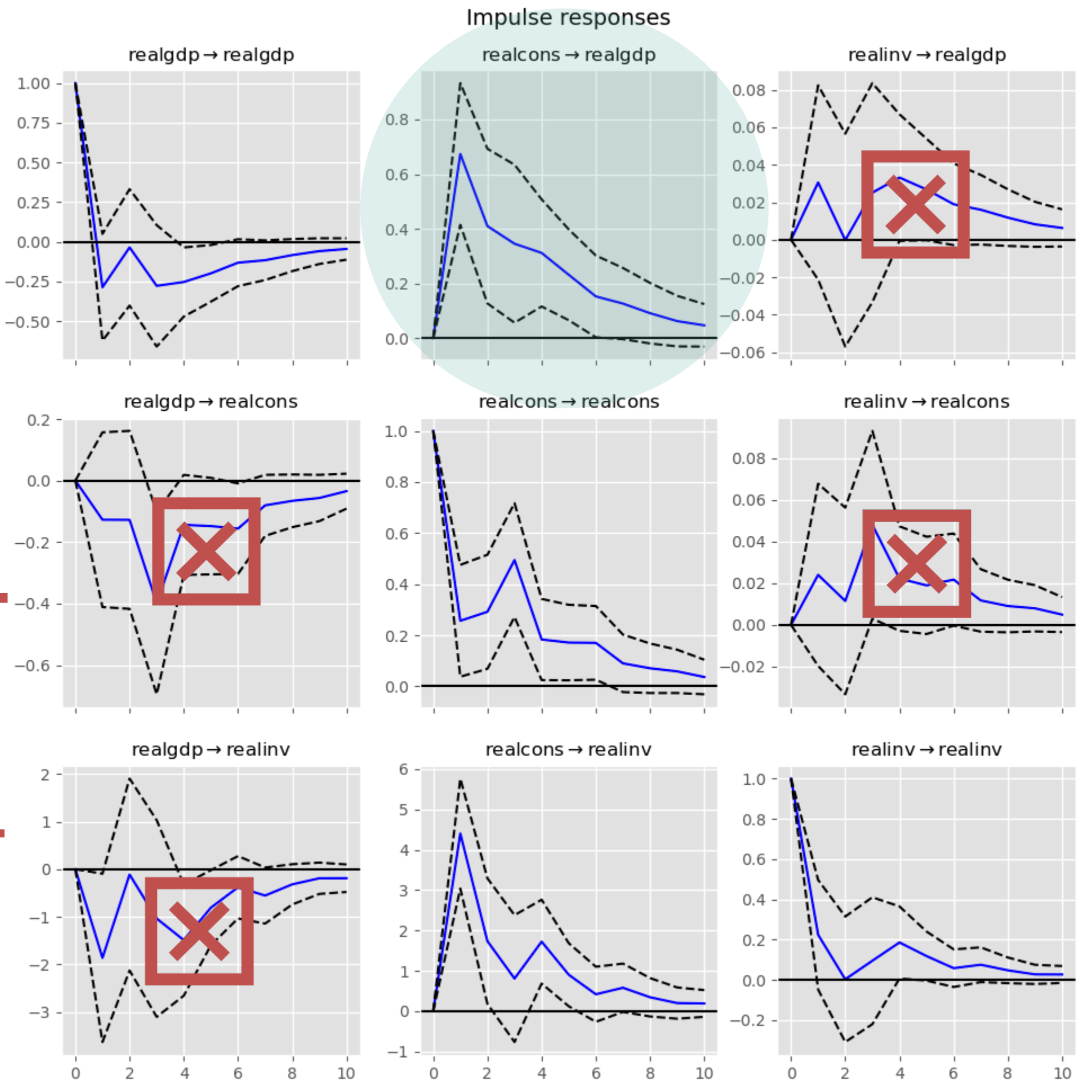


В случайной ошибке, которая соответствует столбцу, происходит шок

Отклик показывает как меняется переменная в строчке

➤ *https://www.statsmodels.org/dev/vector_ar.html*

Импульсные отклики



ВВП Потребление Инвестиции

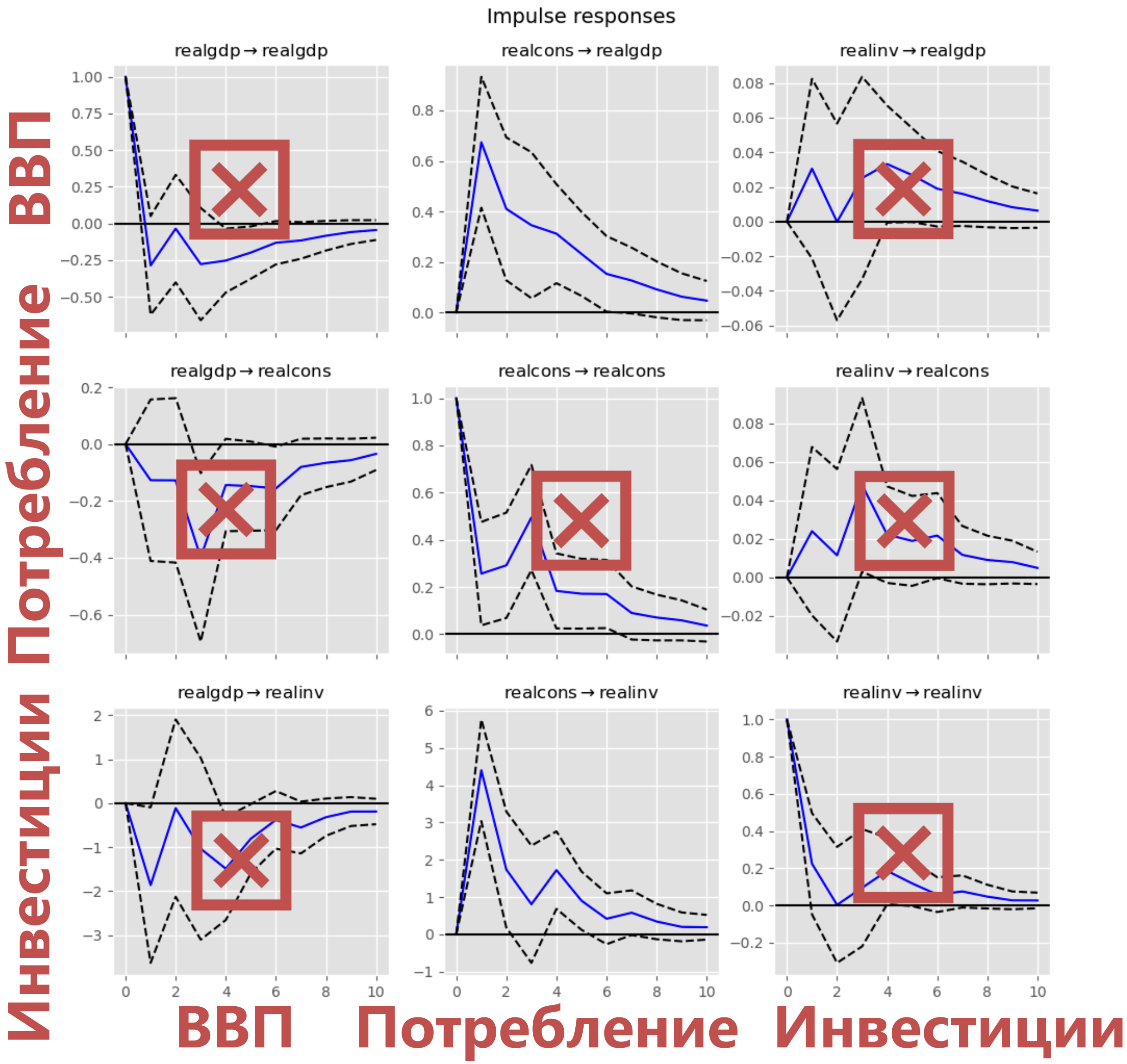
ВВП Потребление Инвестиции

ВВП изменился при шоке в потреблении

В первые несколько месяцев ВВП увеличится, а потом постепенно вернётся к прежнему уровню

➤ *https://www.statsmodels.org/dev/vector_ar.html*

Импульсные отклики



В шоках
по диагонали
нет смысла

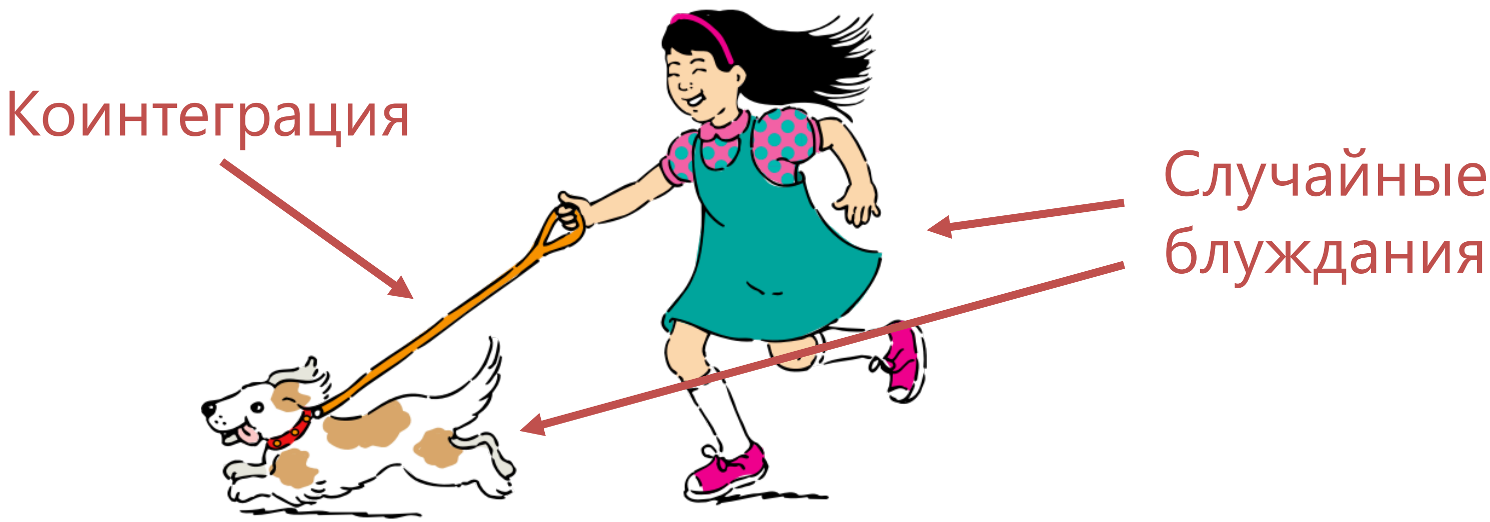
➤ *https://www.statsmodels.org/dev/vector_ar.html*

Что делать, если ряды нестационарны?

- Можно попробовать оценивать модели, для которых неважна эта предпосылка
- Можно попробовать взять все ряды в разностях, таким образом сделав их стационарными
- Можно задуматься о коинтеграции

Коинтеграция

- **Коинтеграция** – это свойство нескольких нестационарных рядов, заключающиеся в существовании их стационарной линейной комбинации
- На основе этого свойства можно найти “долгосрочные взаимосвязи” между нестационарными рядами и исследовать краткосрочные отклонения от них ⇒ модели коррекции ошибок



freesvg.org

Резюме

- Есть класс моделей, позволяющих моделировать системы из временных рядов
- В рамках таких моделей можно не только заниматься прогнозированием, но и пытаться проверять гипотезы
  о взаимосвязях между различными переменными

Кластеризация временных рядов

Кластеризация рядов

- Прогнозированием и проверкой гипотез работа с рядами не ограничивается
- Ряды можно кластеризовать и классифицировать

Зачем кластеризовать ряды

- Найти похожие по динамике и паттернам группы временных рядов: поведение пользователей, похожие друг на друга акции
- Поиск объектов с аномальным поведением

Проблемы

- Похожие паттерны могут находиться на разных участках временных рядов
- Временные ряды разной длины
- Сложность вычислений

Методы

Классическая кластеризация:

- На сырых данных, приведённых к одинаковой длине
- На признаках, извлечённых из рядов
- Вычислительно дёшево

Методы

Кластеризация на эмбедингах:

- С помощью нейросетей преобразуем ряды любой длины в сжатые представления фиксированной длины
- Запускаем классические методы кластеризации на получившихся представлениях (эмбедингах)

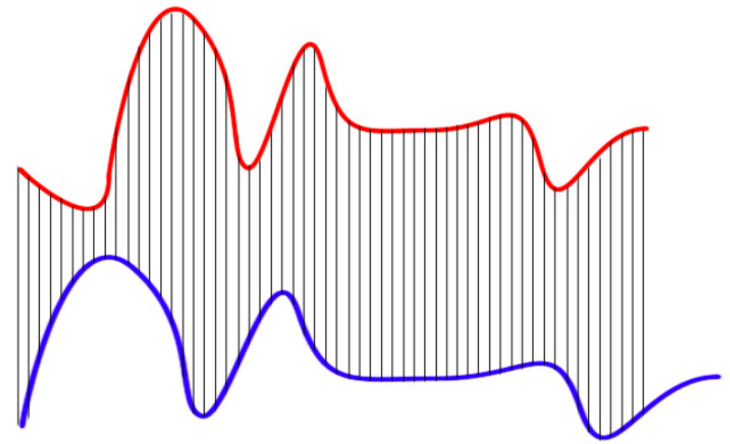
Методы

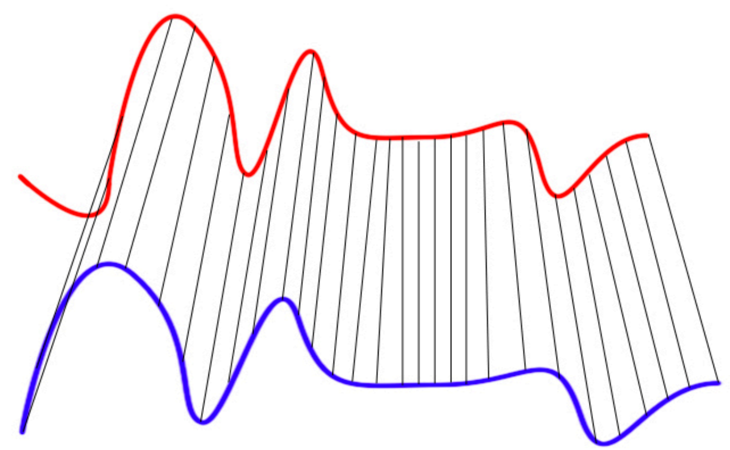
Динамическая трансформация временной шкалы (Dynamic Time Wrapping, DTW):

- Пытается динамически сопоставить ряды между собой
- Позволяет работать с рядами разной длины
- Вычислительно дорогой

DTW

- Деформируем исходные ряды так, чтобы между ними возникло сопоставление
- Сопоставление строится так, чтобы добиться минимизации какой-то метрики расстояния
- Наиболее похожие участки рядов сопоставляются друг с другом независимо от времени их появления

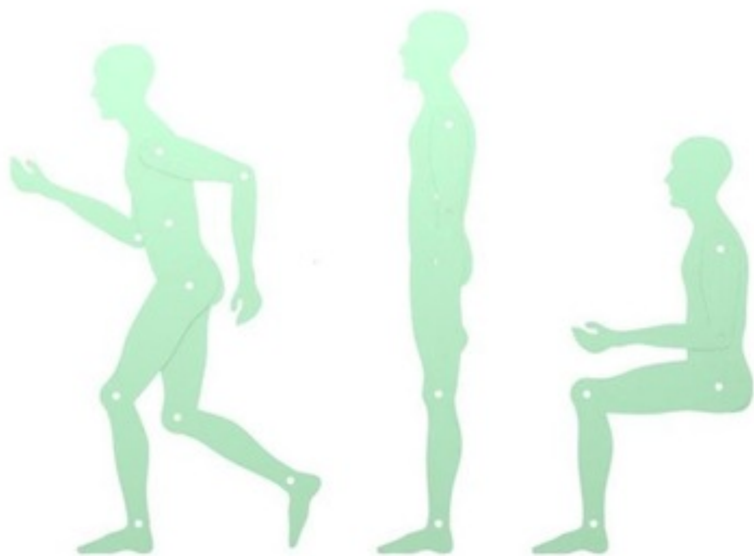

Euclidean Matching


Dynamic Time Warping Matching

➤ *https://tslearn.readthedocs.io/en/stable/user_guide/dtw.html#dtw-softdtw*

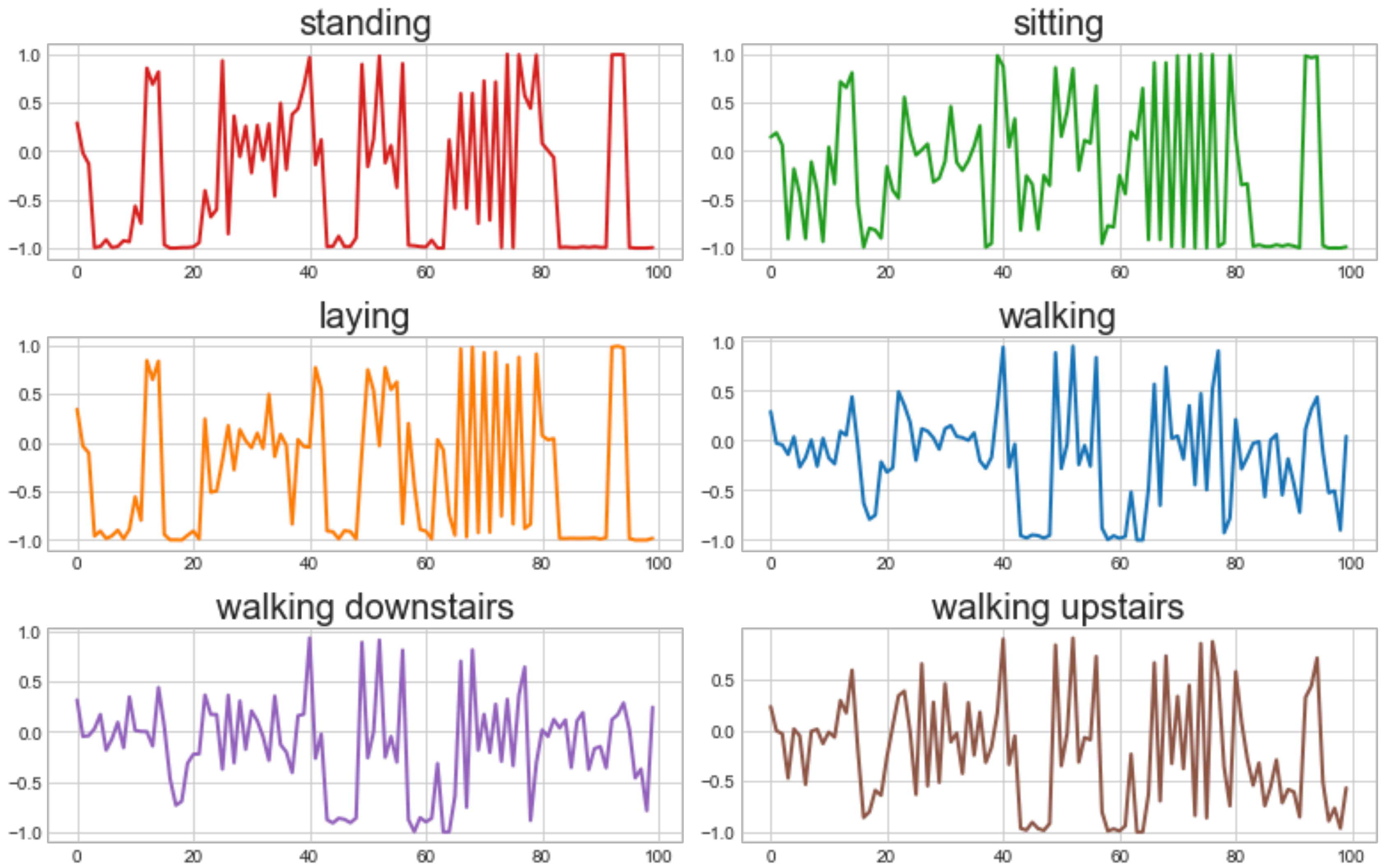
Распознавание физической активности

- Люди что-то делают с телефоном в кармане
- Телефон с помощью встроенных гироскопа и акселерометра записывает информацию об ускорениях людей
- **Задача:** кластеризовать людей по их текущему поведению



alldrawings.ru

Распознавание физической активности



Извлечение признаков из ряда

Извлечение признаков из ряда

Задача: блоггеры пишут статьи в рекомендательную ленту, выводят свой блог на монетизацию, получают деньги за показы рекламы

Проблема: можно попытаться создать много аккаунтов
и выкладывать тексты, сгенерированные автоматически либо автопереводы

❗ Нужен какой-то автоматический алгоритм, который сможет находить таких "блоггеров"

Извлечение признаков из ряда

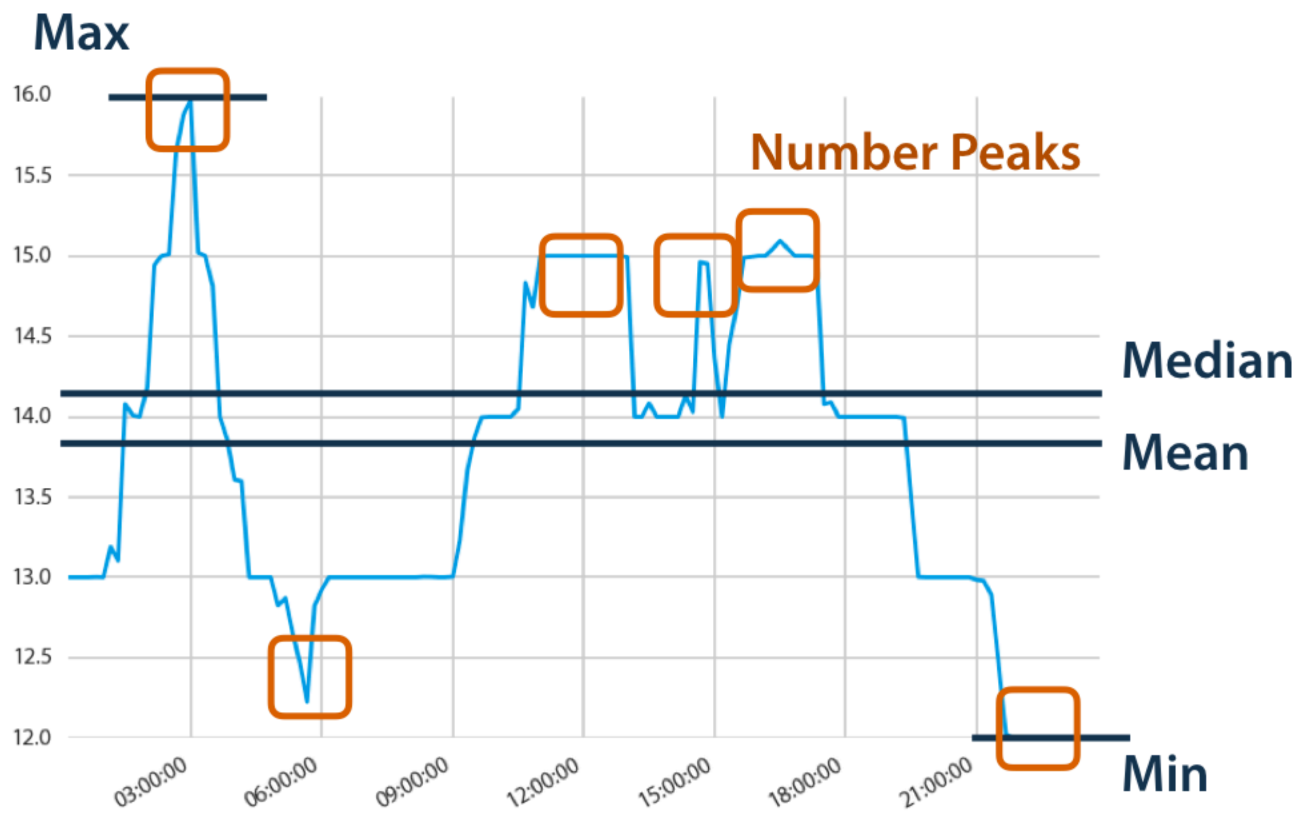
- Поведение каждого пользователя – множество временных рядов:
- Сколько времени тратит на написание статьи, как часто заходит в редактор
- Сколько времени проводит в рекомендательной системе, как часто кликает на статьи, как часто пишет комментарии и т.п.

Извлечение признаков из ряда

- Делаем разметку блоггеров на тех, кто генерирует контент автоматически, и нормальных (создаём обучающую выборку)
- Современные алгоритмы автогенерации контента не так хороши, как о них говорят, и человеку заметна разница ☺
- Обучаем классификатор на основе их поведенческих факторов

Извлечение признаков из ряда

Например, динамика числа кликов за десять минут могла бы выглядеть вот так:



Извлечение признаков из ряда

- Можно выделить такие признаки вручную, а можно какой-то автоматической библиотекой (например tsfresh)
- Главное – делать это аккуратно
- На получившихся признаках можно запустить обучаться свой любимый алгоритм градиентного бустинга или любой другой