

# **Временные ряды: компоненты и наивные модели**

# Данные и задачи

# Данные и задачи: план

- Временные ряды — тип данных.

# Данные и задачи: план

- Временные ряды — тип данных.
- Задачи для одного ряда.

# Данные и задачи: план

- Временные ряды — тип данных.
- Задачи для одного ряда.
- Задачи для множества рядов.

# Заговор рептилоидов

Математический анализ:

**Последовательность**

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$$

# Заговор рептилоидов

Математический анализ:

**Последовательность**

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$$

**Ряд**

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

# Заговор рептилоидов

Математический анализ:

**Последовательность**

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$$

**Ряд**

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

Временные ряды — не ряды!



# Что такое временной ряд?

## Временной ряд

Последовательность наблюдений, упорядоченных во времени.

0, 0, 5, 7, 102, 53, 23.

# Что такое временной ряд?

## Временной ряд

Последовательность наблюдений, упорядоченных во времени.

$0, 0, 5, 7, 102, 53, 23.$

## Временной ряд

Последовательность случайных величин, упорядоченных во времени.

$y_1, y_2, y_3, y_4, \dots, y_T.$

# Задачи для одного ряда

- Спрогнозировать следующие значения.

# Задачи для одного ряда

- Спрогнозировать следующие значения.
- Восстановить пропущенные значения в середине ряда.

# Задачи для одного ряда

- Спрогнозировать следующие значения.
- Восстановить пропущенные значения в середине ряда.
- Восстановить отдельные наблюдения по агрегированным.

# Задачи для одного ряда

- Спрогнозировать следующие значения.
- Восстановить пропущенные значения в середине ряда.
- Восстановить отдельные наблюдения по агрегированным.
- Обнаружить момент разладки.

# Задачи для одного ряда

- Спрогнозировать следующие значения.
- Восстановить пропущенные значения в середине ряда.
- Восстановить отдельные наблюдения по агрегированным.
- Обнаружить момент разладки.
- Выделить составляющие ряда.

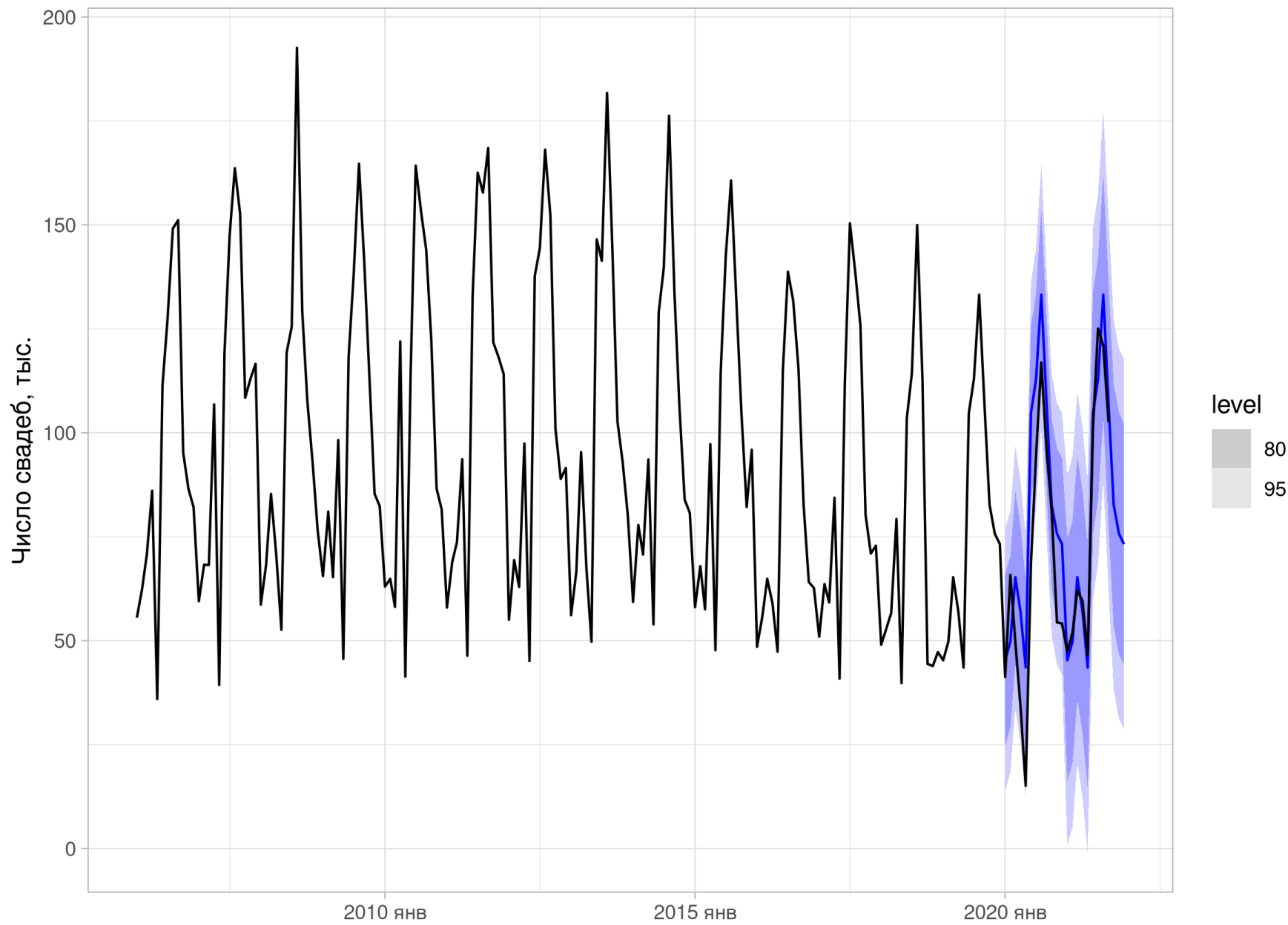
# Задачи для одного ряда

- Спрогнозировать следующие значения.
- Восстановить пропущенные значения в середине ряда.
- Восстановить отдельные наблюдения по агрегированным.
- Обнаружить момент разладки.
- Выделить составляющие ряда.
- ...



# Прогнозируем

Сезонный наивный прогноз числа свадеб на 2 года



# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.

# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.

# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.

# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.

# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.

# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

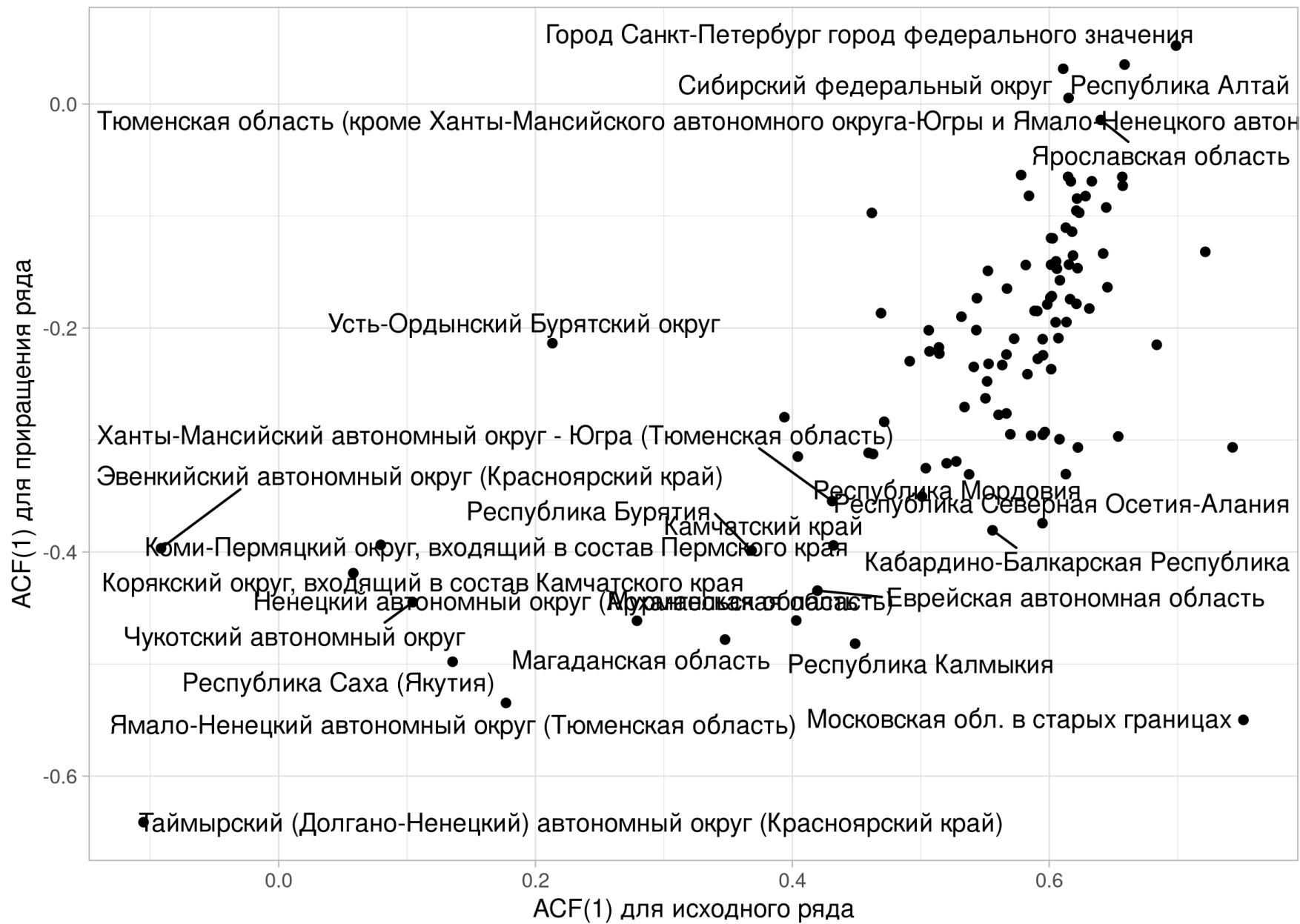
# Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.
- ...



# Измеряем близость рядов

Близость регионов России по динамике числа свадеб



# Модели и алгоритмы

## Модели

- Явные предположения про величины  $y_1, y_2, \dots, y_T$ .

ETS, ARIMA, ORBIT, PROPHET, ...

# Модели и алгоритмы

## Модели

- Явные предположения про величины  $y_1, y_2, \dots, y_T$ .
- Метод оценивания: максимальное правдоподобие, байесовский подход.

ETS, ARIMA, ORBIT, PROPHET, ...

# Модели и алгоритмы

## Модели

- Явные предположения про величины  $y_1, y_2, \dots, y_T$ .
- Метод оценивания: максимальное правдоподобие, байесовский подход.
- Точечные и интервальные прогнозы, проверка гипотез.

ETS, ARIMA, ORBIT, PROPHET, ...

# Модели и алгоритмы

## Алгоритмы

- Размытые предположения про величины  $y_1, y_2, \dots, y_T$ .

STL, градиентный бустинг, случайный лес, ...

# Модели и алгоритмы

## Алгоритмы

- Размытые предположения про величины  $y_1, y_2, \dots, y_T$ .
- Особая инструкция.

STL, градиентный бустинг, случайный лес, ...

# Модели и алгоритмы

## Алгоритмы

- Размытые предположения про величины  $y_1, y_2, \dots, y_T$ .
- Особая инструкция.
- Точечные результаты без доверительных интервалов.

STL, градиентный бустинг, случайный лес, ...

# Фокус курса

Прогнозирование одномерных рядов с помощью моделей.



# Компоненты ряда

# Компоненты ряда: план

- Тренд, цикличность и сезонность.

# Компоненты ряда: план

- Тренд, цикличность и сезонность.
- Аддитивное и мультипликативное разложение.

# Компоненты ряда: план

- Тренд, цикличность и сезонность.
- Аддитивное и мультипликативное разложение.
- Откуда взять формальное определение?

# Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

# Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Тренд — плавно изменяющаяся составляющая ряда.

# Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

**Тренд** — плавно изменяющаяся составляющая ряда.

**Сезонная составляющая** — составляющая с чёткой периодичностью и стабильной интенсивностью.

# Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

**Тренд** — плавно изменяющаяся составляющая ряда.

**Сезонная составляющая** — составляющая с чёткой периодичностью и стабильной интенсивностью.

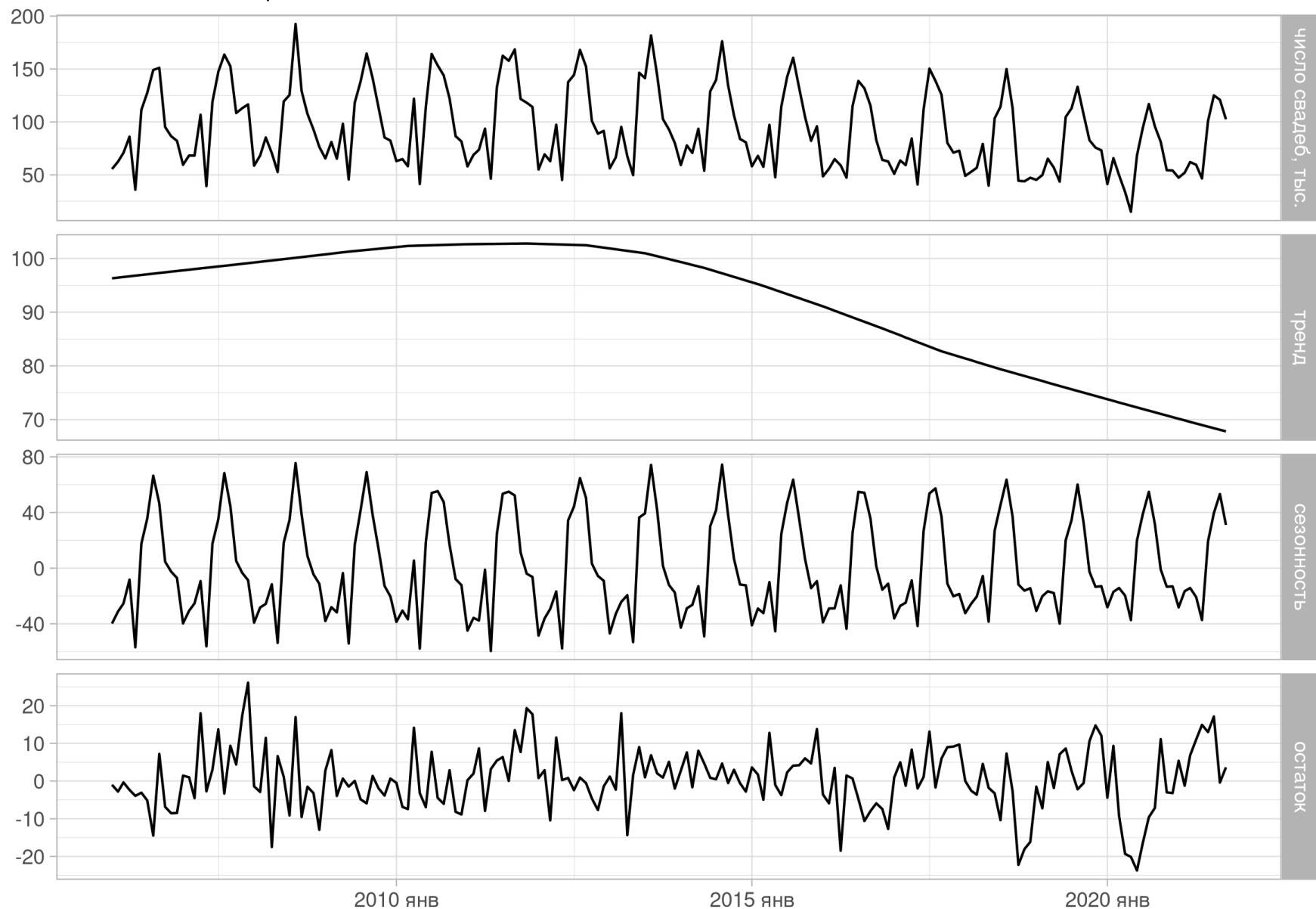
**Случайная компонента** (остаток) — всё остальное.



# Тренд, сезонность и остаток

STL разложение числа свадеб в России

Число свадеб = тренд + сезонность + остаток



**Строгое определение?**

# Строгое определение?

Единого строгого определения **не** будет!

# Строгое определение?

Единого строгого определения **не** будет!

Некоторые модели и алгоритмы формально **определяют** данные составляющие.

# Циклическая составляющая

Иногда ряд раскладывают дальше

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

# Циклическая составляющая

Иногда ряд раскладывают дальше

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

**Циклическая составляющая** — составляющая с плавающей периодичностью и нестабильной интенсивностью.

# Циклическая составляющая

Иногда ряд раскладывают дальше

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

**Циклическая составляющая** — составляющая с плавающей периодичностью и нестабильной интенсивностью.

**Тренд** (в узком смысле) — плавно изменяющаяся монотонная составляющая ряда.

# Аддитивное и мультипликативное разложение

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$



# Аддитивное и мультипликативное разложение

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Мультипликативное разложение ряда:

$$y_t = trend_t \cdot seas_t \cdot remainder_t.$$

# Аддитивное и мультипликативное разложение

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Мультипликативное разложение ряда:

$$y_t = trend_t \cdot seas_t \cdot remainder_t.$$

Превращаем одно в другое:

$$\ln y_t = \ln trend_t + \ln seas_t + \ln remainder_t.$$

# Какие единороги лучше?

Формальное определение составляющих **зависит от модели.**

# Какие единороги лучше?

Формальное определение составляющих **зависит от модели**.

**Алгоритм STL**: одно разложение

$$y_t = trend_t + seas_t + remainder_t.$$

# Какие единороги лучше?

Формальное определение составляющих **зависит от модели**.

**Алгоритм STL**: одно разложение

$$y_t = trend_t + seas_t + remainder_t.$$

**Модель ETS(AAA)**: другое разложение

$$y_t = trend_t + seas_t + remainder_t.$$

# Какие единороги лучше?

Формальное определение составляющих **зависит от модели**.

**Алгоритм STL**: одно разложение

$$y_t = trend_t + seas_t + remainder_t.$$

**Модель ETS(AAA)**: другое разложение

$$y_t = trend_t + seas_t + remainder_t.$$

Важно понимать **цель построения** разложения.

# А зачем разложение?

- Интересно **само по себе**.

# А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.



# А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

# А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

# А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

А характеристики зачем?

- Чтобы классифицировать новый ряд в один из заданных классов.

# А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

А характеристики зачем?

- Чтобы классифицировать новый ряд в один из заданных классов.
- Чтобы выявить в рядах неизвестные кластеры.

# Компоненты ряда: итоги

- Тренд **плавно меняется** и включает циклическую составляющую.

# Компоненты ряда: итоги

- Тренд **плавно меняется** и включает циклическую составляющую.
- Сезонная составляющая имеет **чёткую периодичность** и **стабильную амплитуду**.

# Компоненты ряда: итоги

- Тренд **плавно меняется** и включает циклическую составляющую.
- Сезонная составляющая имеет **чёткую периодичность** и **стабильную амплитуду**.
- Точная формализация компонент **зависит от модели**.

# Алгоритм STL



# Алгоритм STL: план

- Локальная регрессия.

# Алгоритм STL: план

- Локальная регрессия.
- Внешний цикл STL.

# Алгоритм STL: план

- Локальная регрессия.
- Внешний цикл STL.
- Внутренний цикл STL.

# Алгоритм STL: план

- Локальная регрессия.
- Внешний цикл STL.
- Внутренний цикл STL.
- Параметры STL.

# STL

STL — Seasonal Trend decomposition with Loess.

STL — разложение на сезонность и тренд с использованием LOESS.

# STL

**STL** — Seasonal Trend decomposition with Loess.

STL — разложение на сезонность и тренд с использованием LOESS.

**LOESS** — LOcal regrESSion.

LOESS — локальная линейная регрессия.

# STL как чёрный ящик

На входе:

Ряд  $Y_t$ .

Параметры алгоритма  $n_p, n_i, n_o, n_l, n_s, n_t$ .

# STL как чёрный ящик

На входе:

Ряд  $Y_t$ .

Параметры алгоритма  $n_p, n_i, n_o, n_l, n_s, n_t$ .

На выходе:

Разложение  $Y_t = T_t + S_t + R_t$ .



# STL как чёрный ящик

На входе:

Ряд  $Y_t$ .

Параметры алгоритма  $n_p, n_i, n_o, n_l, n_s, n_t$ .

На выходе:

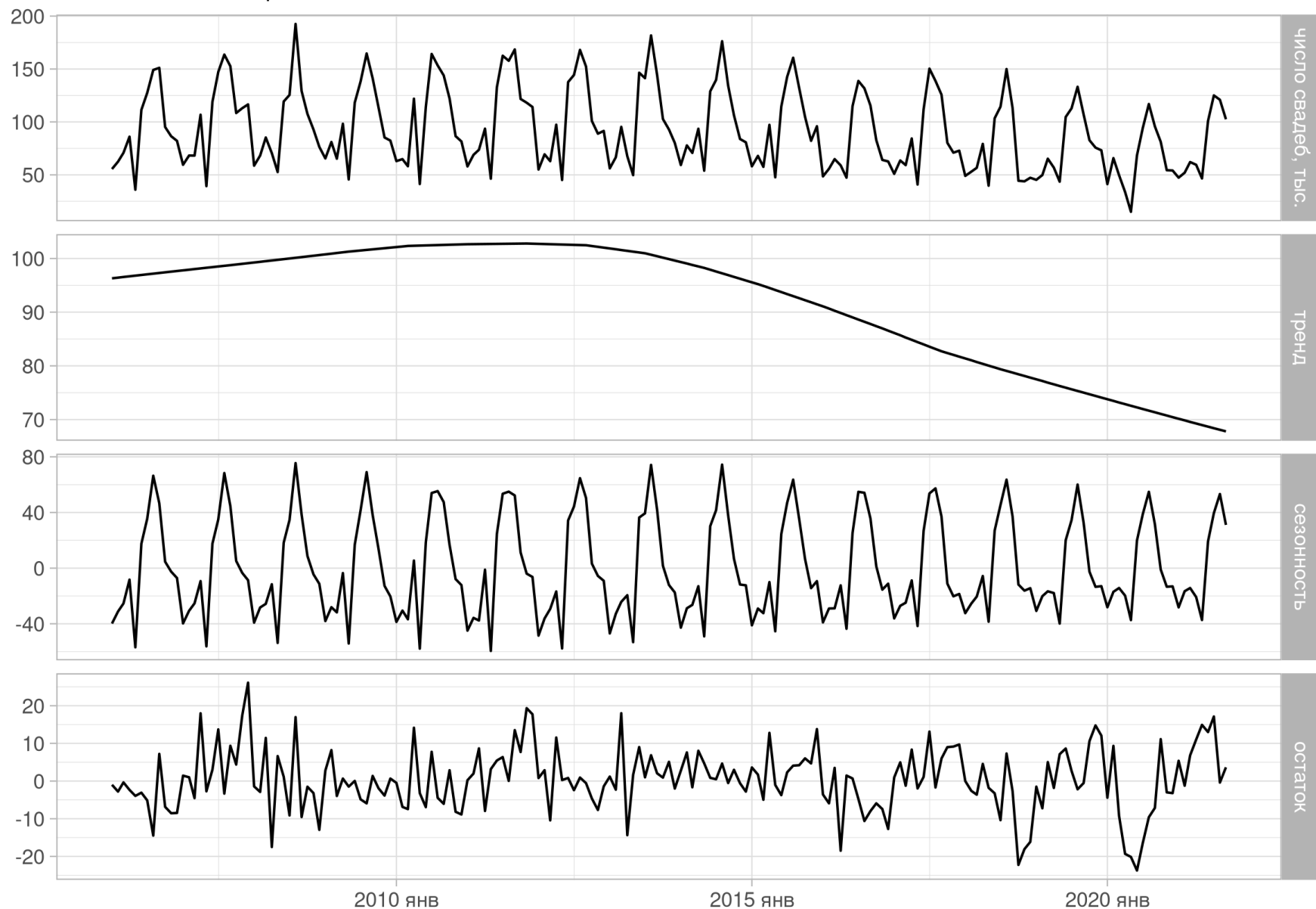
Разложение  $Y_t = T_t + S_t + R_t$ .

Чёрный ящик **долго** настраивали.

# STL: результат

STL разложение числа свадеб в России

Число свадеб = тренд + сезонность + остаток



# LOESS

- Хотим построить прогноз для точки  $x$ .
- Находим **локальные оценки**  $\hat{\beta}_1(x), \hat{\beta}_2(x)$ .

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Прогнозируем:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

# LOESS

- Хотим построить прогноз для точки  $x$ .
- Находим **локальные оценки**  $\hat{\beta}_1(x), \hat{\beta}_2(x)$ .

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Прогнозируем:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

## Ядерная функция

- Функция  $K_h(x_i - x)$  убывает с увеличением расстояния  $|x_i - x|$ ;
- Параметр  $h$  отвечает за ширину окна сглаживания.

# LOESS

- Хотим построить прогноз для точки  $x$ .
- Находим **локальные оценки**  $\hat{\beta}_1(x), \hat{\beta}_2(x)$ .

$$\min \sum_i K_h(x_i - x) (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Прогнозируем:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

## Ядерная функция

- Функция  $K_h(x_i - x)$  убывает с увеличением расстояния  $|x_i - x|$ ;
- Параметр  $h$  отвечает за ширину окна сглаживания.

Например,  $h$  — количество точек  $x_i$  рядом с  $x$ , которые мы учитываем.

# Нюансы локальной регрессии

- Выбор **степени полинома**.

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i - \hat{\beta}_3 x_i^2)^2$$

- Выбор **ядерной функции**.

$$K_h(d) = \frac{1}{\sqrt{2\pi}h} \exp\left(-d^2/2h^2\right)$$

- Выбор **ширины окна**  $h$ .

# STL с высоты птичьего полёта

Цель: разложение  $Y_t = T_t + S_t + R_t$ .

Алгоритм содержит два цикла: **внешний** и **внутренний**.

1. Инициализируем  $T_t = 0, R_t = 0$ .

**Внешний** цикл:

# STL с высоты птичьего полёта

Цель: разложение  $Y_t = T_t + S_t + R_t$ .

Алгоритм содержит два цикла: **внешний** и **внутренний**.

1. Инициализируем  $T_t = 0, R_t = 0$ .

**Внешний** цикл:

2. Посчитаем вес каждого наблюдения,  $\rho_t$ .

На первом проходе  $\rho_t = 1$  у каждого наблюдения.

На последующих проходах  $\rho_t$  отрицательно зависит от свежей величины  $R_t$ .



# STL с высоты птичьего полёта

Цель: разложение  $Y_t = T_t + S_t + R_t$ .

Алгоритм содержит два цикла: **внешний** и **внутренний**.

1. Инициализируем  $T_t = 0, R_t = 0$ .

**Внешний** цикл:

2. Посчитаем вес каждого наблюдения,  $\rho_t$ .

На первом проходе  $\rho_t = 1$  у каждого наблюдения.

На последующих проходах  $\rho_t$  отрицательно зависит от свежей величины  $R_t$ .

3. Обновим текущее разложение  $Y_t = T_t + S_t + R_t$  с учётом весов  $\rho_t$ .

# STL: внутренний цикл

Цель: обновить разложение  $Y_t = T_t + S_t + R_t$ .

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

# STL: внутренний цикл

Цель: обновить разложение  $Y_t = T_t + S_t + R_t$ .

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендированный ряд на 12 подрядов.

# STL: внутренний цикл

Цель: обновить разложение  $Y_t = T_t + S_t + R_t$ .

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендированный ряд на 12 подрядов.
3. Сгладим каждый подряд по отдельности с помощью LOESS:

$$C^{jan} = LOESS_{\rho}(Y_{jan}^{det}), C^{feb} = LOESS_{\rho}(Y_{feb}^{det}), \dots$$

# STL: внутренний цикл

Цель: обновить разложение  $Y_t = T_t + S_t + R_t$ .

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендрованный ряд на 12 подрядов.
3. Сгладим каждый подряд по отдельности с помощью LOESS:

$$C^{jan} = LOESS_{\rho}(Y_{jan}^{det}), C^{feb} = LOESS_{\rho}(Y_{feb}^{det}), \dots$$

4. Выделяем низкочастотную составляющую (дважды скользящее среднее + LOESS):

$$L_t = LOESS(MA(MA(C_t)))$$

# STL: внутренний цикл

Цель: обновить разложение  $Y_t = T_t + S_t + R_t$ .

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендрованный ряд на 12 подрядов.
3. Сгладим каждый подряд по отдельности с помощью LOESS:

$$C^{jan} = LOESS_{\rho}(Y_{jan}^{det}), C^{feb} = LOESS_{\rho}(Y_{feb}^{det}), \dots$$

4. Выделяем низкочастотную составляющую (дважды скользящее среднее + LOESS):

$$L_t = LOESS(MA(MA(C_t)))$$

- 5-6. Получаем новые  $S_t^{new}$  и  $T_t^{new}$ .

# STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.

# STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.
4. Выделяем низкочастотную составляющую.



# STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.
4. Выделяем низкочастотную составляющую.
5. Удаляем низкочастотную составляющую, получаем **новую** сезонную компоненту:

$$S_t^{new} = C_t - L_t.$$

# STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.
4. Выделяем низкочастотную составляющую.
5. Удаляем низкочастотную составляющую, получаем **новую** сезонную компоненту:

$$S_t^{new} = C_t - L_t.$$

6. Удаляем новую сезонность из исходного ряда и сглаживаем с помощью LOESS:

$$T_t^{new} = LOESS_{\rho}(Y_t - S_t^{new}).$$

$y\phi!$

# Параметры STL

- $n_p$  — периодичность сезонности, например,  $n_p = 12$ .

# Параметры STL

- $n_p$  — периодичность сезонности, например,  $n_p = 12$ .
- $n_o$  — число проходов внешнего цикла.

Чем больше число  $n_o$ , тем слабее влияние выбросов.

Значение  $n_o = 1$  часто достаточно.

# Параметры STL

- $n_p$  — периодичность сезонности, например,  $n_p = 12$ .
- $n_o$  — число проходов внешнего цикла.

Чем больше число  $n_o$ , тем слабее влияние выбросов.  
Значение  $n_o = 1$  часто достаточно.

- $n_i$  — число проходов внутреннего цикла.  
Значение  $n_i = 2$  часто достаточно для достижения сходимости.

# Параметры сглаживания STL

- $n_l$  — сила сглаживания низкочастотного фильтра.

Что настроить?

# Параметры сглаживания STL

- $n_l$  — сила сглаживания низкочастотного фильтра.
- $n_s$  — сила сглаживания сезонных подрядов.

Что настроить?



# Параметры сглаживания STL

- $n_l$  — сила сглаживания низкочастотного фильтра.
- $n_s$  — сила сглаживания сезонных подрядов.
- $n_t$  — сила сглаживания при выделении тренда на последнем шаге.

Что настроить?

# Параметры сглаживания STL

- $n_l$  — сила сглаживания низкочастотного фильтра.
- $n_s$  — сила сглаживания сезонных подрядов.
- $n_t$  — сила сглаживания при выделении тренда на последнем шаге.

## Что настроить?

1. Обязательно указать периодичность  $n_p$ .

# Параметры сглаживания STL

- $n_l$  — сила сглаживания низкочастотного фильтра.
- $n_s$  — сила сглаживания сезонных подрядов.
- $n_t$  — сила сглаживания при выделении тренда на последнем шаге.

## Что настроить?

1. Обязательно указать периодичность  $n_p$ .
2. Возможно, поиграться с  $n_s$ .

# STL: итоги

- LOESS — локальная регрессия.

# STL: итоги

- LOESS — локальная регрессия.
- STL — хорошо проверенный временем алгоритм без модели.

# STL: итоги

- LOESS — локальная регрессия.
- STL — хорошо проверенный временем алгоритм без модели.
- При желании можно поиграться с силами сглаживания.

# Характеристики рядов

# Характеристики рядов: план

- Выборочная автокорреляция.



# Характеристики рядов: план

- Выборочная автокорреляция.
- Выборочная частная автокорреляция.

# Характеристики рядов: план

- Выборочная автокорреляция.
- Выборочная частная автокорреляция.
- STL-характеристики.

# Задачи для множества рядов

- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

# Задачи для множества рядов

- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

## Как решить?

1. Для каждого ряда сгенерировать **признаки**.

# Задачи для множества рядов

- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

## Как решить?

1. Для каждого ряда сгенерировать **признаки**.
2. К полученным признакам применить алгоритм для перекрестных данных.

# Задачи для множества рядов

- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

## Как решить?

1. Для каждого ряда сгенерировать **признаки**.
2. К полученным признакам применить алгоритм для перекрестных данных.

# Задачи для множества рядов

- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

## Как решить?

1. Для каждого ряда сгенерировать **признаки**.
2. К полученным признакам применить алгоритм для перекрестных данных.

Классифицировать: с помощью случайного леса.

Измерить расстояние с помощью метрики Махаланобиса.

Кластеризовать с помощью иерархической кластеризации.

# Создаём признаки

Два множества признаков:

- Выборочная ACF (автокорреляционная функция, AutoCorrelation Function).



# Создаём признаки

Два множества признаков:

- Выборочная ACF (автокорреляционная функция, AutoCorrelation Function).
- Выборочная PACF (частная автокорреляционная функция, Partial ACF).

# Создаём признаки

Два множества признаков:

- Выборочная ACF (автокорреляционная функция, AutoCorrelation Function).
- Выборочная PACF (частная автокорреляционная функция, Partial ACF).

# Создаём признаки

Два множества признаков:

- Выборочная ACF (автокорреляционная функция, AutoCorrelation Function).
- Выборочная PACF (частная автокорреляционная функция, Partial ACF).

Из одного ряд получим:

$ACF_1, ACF_2, ACF_3, \dots$

$PACF_1, PACF_2, PACF_3, \dots$

## Выборочная ACF

Оценим множество парных регрессий:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

## Выборочная ACF

Оценим множество парных регрессий:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-2}, \quad ACF_2 = \hat{\beta}_2;$$

## Выборочная ACF

Оценим множество парных регрессий:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-2}, \quad ACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-k}, \quad ACF_k = \hat{\beta}_2;$$

## Выборочная ACF

Оценим множество парных регрессий:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

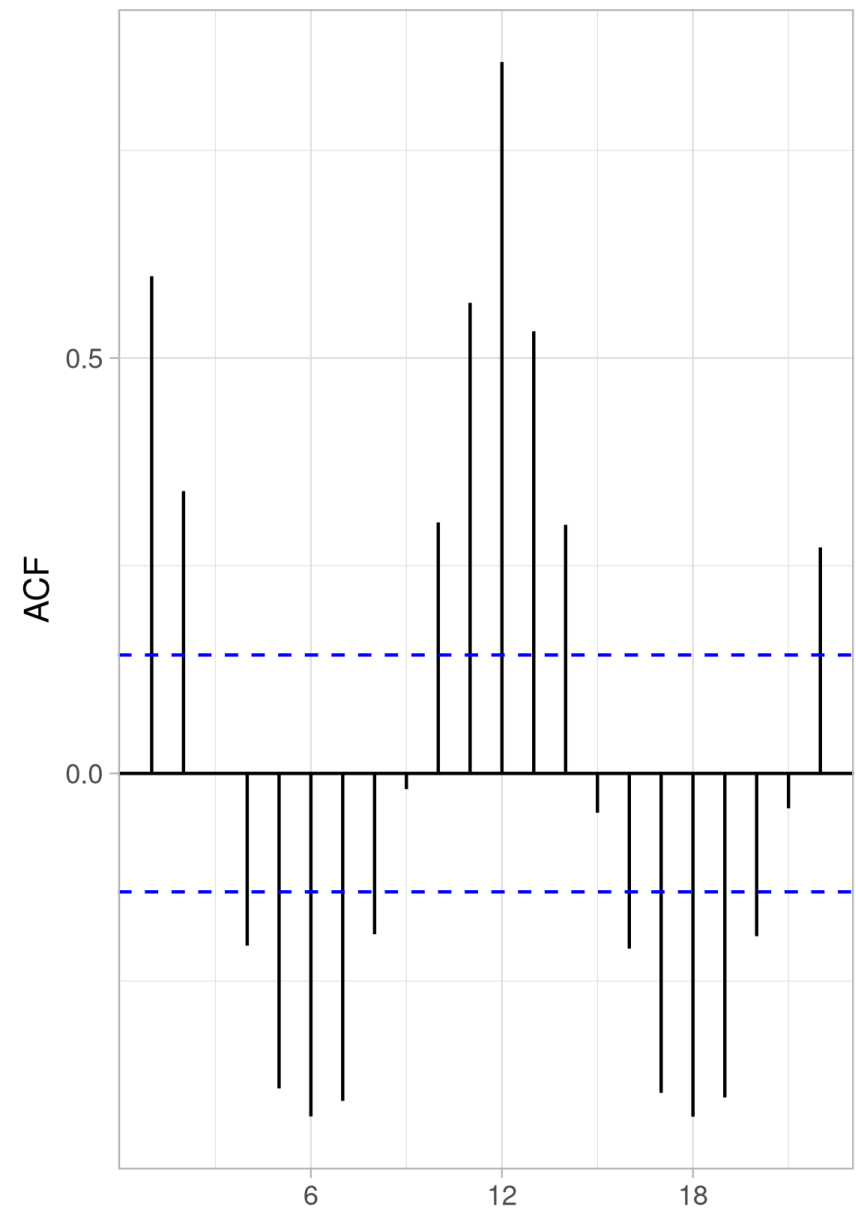
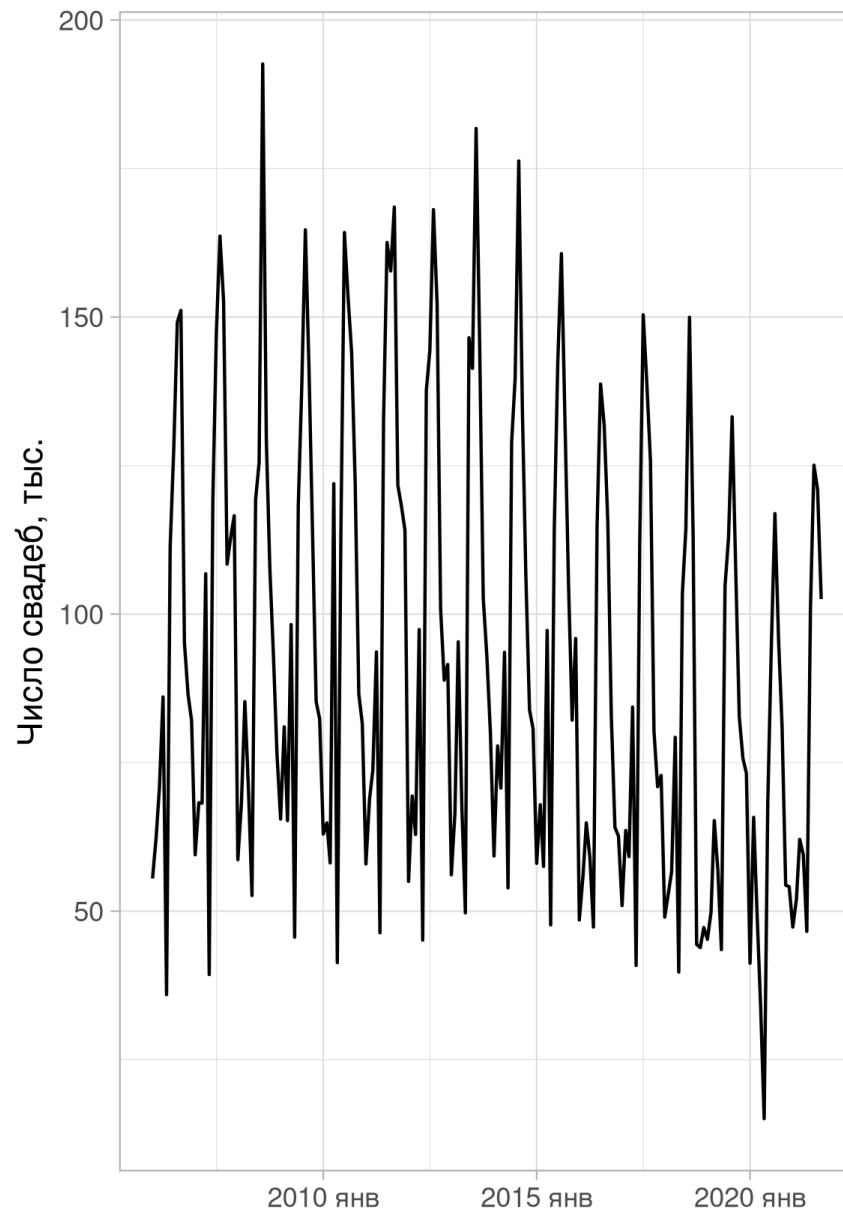
$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-2}, \quad ACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-k}, \quad ACF_k = \hat{\beta}_2;$$

**Смысл**  $ACF_2$ : на сколько единиц в среднем  $y_t$  выше среднего, если  $y_{t-2}$  выше среднего на одну единицу.

# Ряд и его ACF

График числа свадеб в России и ACF





# Почему АСФ — корреляция?

Классическое определение

## Выборочная АСФ

$ACF_k$  — выборочная корреляция между рядом  $y_t$  и рядом  $y_{t-k}$ .

# Почему ACF — корреляция?

Классическое определение

## Выборочная ACF

$ACF_k$  — выборочная корреляция между рядом  $y_t$  и рядом  $y_{t-k}$ .

Различие между определениями **мало**.

## Выборочная PACF

Оценим множество множественных регрессий:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

## Выборочная PACF

Оценим множество множественных регрессий:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}, \quad PACF_2 = \hat{\beta}_2;$$

## Выборочная PACF

Оценим множество множественных регрессий:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}, \quad PACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \dots + \hat{\beta}_k y_{t-k}, \quad PACF_k = \hat{\beta}_k;$$

## Выборочная PACF

Оценим множество множественных регрессий:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

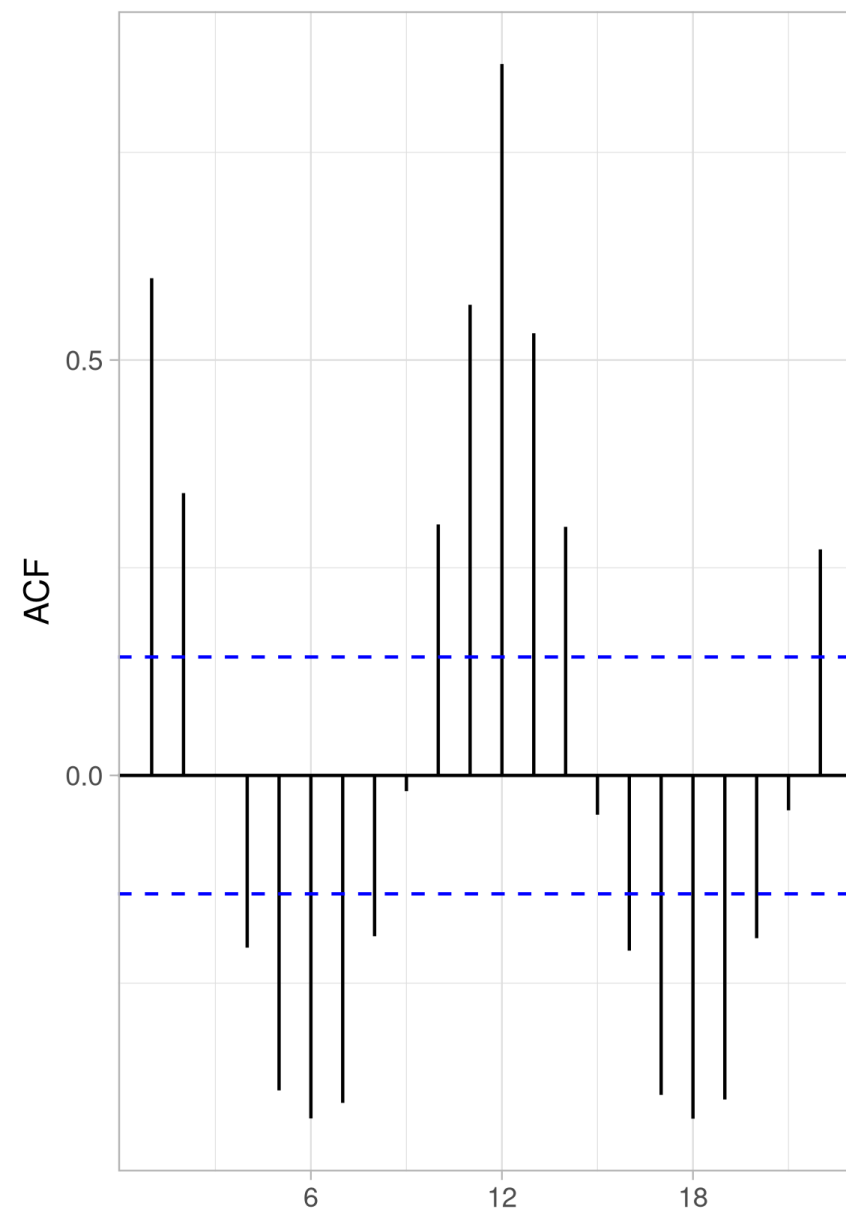
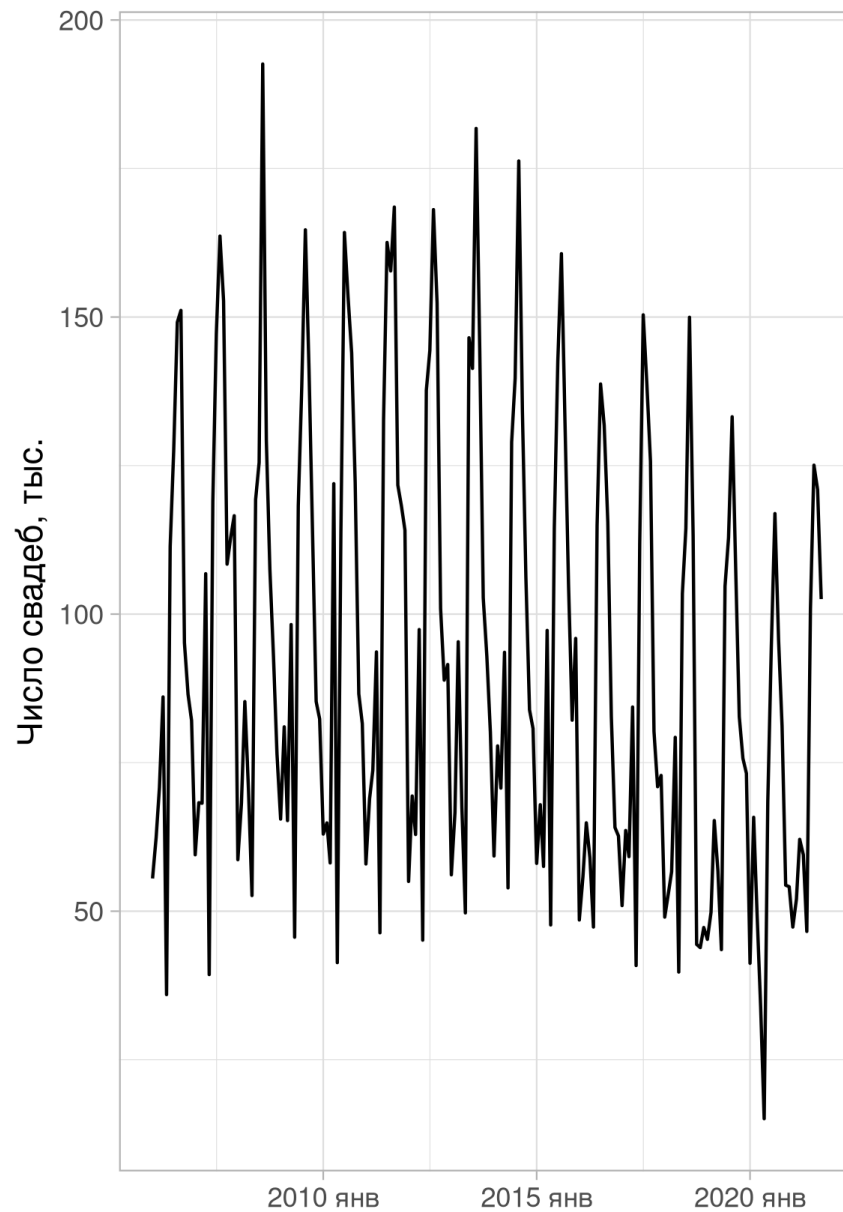
$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}, \quad PACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \dots + \hat{\beta}_k y_{t-k}, \quad PACF_k = \hat{\beta}_k;$$

**Смысл**  $PACF_2$ : на сколько единиц в среднем  $y_t$  выше среднего, если  $y_{t-2}$  выше среднего на одну единицу, а  $y_{t-1}$  на среднем уровне.

# Ряд и его PACF

График числа свадеб в России и PACF



# Почему PACF — корреляция?

## Классическое определение

### Выборочная PACF

$PACF_4$  — выборочная корреляция между остатками  $a_t$  и остатками  $b_t$ .

$a_t$  — остатки из регрессии

$y_t$  на  $1, y_{t-1}, y_{t-2}, y_{t-3}$ .

$b_t$  — остатки из регрессии

$y_{t-4}$  на  $1, y_{t-1}, y_{t-2}, y_{t-3}$ .



# Почему PACF — корреляция?

## Классическое определение

### Выборочная PACF

$PACF_4$  — выборочная корреляция между остатками  $a_t$  и остатками  $b_t$ .

$a_t$  — остатки из регрессии

$y_t$  на  $1, y_{t-1}, y_{t-2}, y_{t-3}$ .

$b_t$  — остатки из регрессии

$y_{t-4}$  на  $1, y_{t-1}, y_{t-2}, y_{t-3}$ .

Различие между определениями **мало**.

# STL-характеристики

На выходе:

$$y_t = T_t + S_t + R_t.$$

# STL-характеристики

На выходе:

$$y_t = T_t + S_t + R_t.$$

Измерим:

- Выраженность тренда  $F_{trend}$ .
- Выраженность сезонности  $F_{seas}$ .

# Выраженность тренда и сезонности

Получили разложение:

$$y_t = trend_t + seas_t + remainder_t.$$

# Выраженность тренда и сезонности

Получили разложение:

$$y_t = trend_t + seas_t + remainder_t.$$

**Идея определения:** При идеальном разложении с некоррелированными компонентами:

$$F_{trend} = \frac{sVar(trend)}{sVar(trend) + sVar(remainder)},$$

# Выраженность тренда и сезонности

Получили разложение:

$$y_t = trend_t + seas_t + remainder_t.$$

**Идея определения:** При идеальном разложении с некоррелированными компонентами:

$$F_{trend} = \frac{sVar(trend)}{sVar(trend) + sVar(remainder)},$$

$$F_{seas} = \frac{sVar(seas)}{sVar(seas) + sVar(remainder)},$$

# Выраженность тренда и сезонности

Получили разложение:

$$y_t = trend_t + seas_t + remainder_t.$$

# Выраженность тренда и сезонности

Получили разложение:

$$y_t = trend_t + seas_t + remainder_t.$$

На практике:

- Выраженность тренда:

$$F_{trend} = \max \left\{ 1 - \frac{sVar(remainder)}{sVar(trend + remainder)}, 0 \right\}.$$



# Выраженность тренда и сезонности

Получили разложение:

$$y_t = trend_t + seas_t + remainder_t.$$

На практике:

- Выраженность тренда:

$$F_{trend} = \max \left\{ 1 - \frac{sVar(remainder)}{sVar(trend + remainder)}, 0 \right\}.$$

- Выраженность сезонности:

$$F_{seas} = \max \left\{ 1 - \frac{sVar(remainder)}{sVar(seas + remainder)}, 0 \right\}.$$

# Выраженность тренда и сезонности

Близость регионов России по динамике числа свадеб



# Характеристики рядов: итоги

- ACF — коэффициенты в **парных** регрессиях или корреляции.

# Характеристики рядов: итоги

- ACF — коэффициенты в **парных** регрессиях или корреляции.
- PACF — коэффициенты во **множественных** регрессиях или корреляции.

# Характеристики рядов: итоги

- ACF — коэффициенты в **парных** регрессиях или корреляции.
- PACF — коэффициенты во **множественных** регрессиях или корреляции.
- STL позволяет измерить **выраженность тренда и сезонности** по сравнению с остаточной компонентой.

# Простейшие модели

# Простейшие модели: план

- Белый шум.

# Простейшие модели: план

- Белый шум.
- Независимые наблюдения.



# Простейшие модели: план

- Белый шум.
- Независимые наблюдения.
- Случайное блуждание.

# Белый шум

## Белый шум

Временной ряд  $u_t$  — белый шум, если:

- $\mathbb{E}(u_t) = 0$ ;
- $\text{Var}(u_t) = \sigma^2$ ;
- $\text{Cov}(u_s, u_t) = 0$  при  $s \neq t$ .

# Белый шум

## Белый шум

Временной ряд  $u_t$  — белый шум, если:

- $\mathbb{E}(u_t) = 0$ ;
  - $\text{Var}(u_t) = \sigma^2$ ;
  - $\text{Cov}(u_s, u_t) = 0$  при  $s \neq t$ .
- 
- Составная часть всех моделей. Чаще всего белый шум — это, что отказались моделировать.

# Белый шум

## Белый шум

Временной ряд  $u_t$  — белый шум, если:

- $\mathbb{E}(u_t) = 0$ ;
  - $\text{Var}(u_t) = \sigma^2$ ;
  - $\text{Cov}(u_s, u_t) = 0$  при  $s \neq t$ .
- 
- Составная часть всех моделей. Чаще всего белый шум — это, что отказались моделировать.
  - Часто дополнительно предполагают **независимость** и **нормальность**.

# Белый шум

## Белый шум

Временной ряд  $u_t$  — белый шум, если:

- $\mathbb{E}(u_t) = 0$ ;
- $\text{Var}(u_t) = \sigma^2$ ;
- $\text{Cov}(u_s, u_t) = 0$  при  $s \neq t$ .

- Составная часть всех моделей. Чаще всего белый шум — это, что отказались моделировать.
- Часто дополнительно предполагают **независимость** и **нормальность**.
- В белом шуме **черти водятся**.

ARCH, GARCH модели волатильности основаны на том, что  $u_t$  и  $u_s$  могут быть зависимы!

# Независимые наблюдения

## Модель

$$y_t = \mu + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ .

# Независимые наблюдения

## Модель

$$y_t = \mu + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ .

Оценки:

$$\hat{\mu}_{ML} = \bar{y}, \quad \hat{\sigma}_{ML}^2 = \frac{\sum (y_i - \bar{y})^2}{T}.$$

# Независимые наблюдения

## Модель

$$y_t = \mu + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ .

Оценки:

$$\hat{\mu}_{ML} = \bar{y}, \quad \hat{\sigma}_{ML}^2 = \frac{\sum (y_i - \bar{y})^2}{T}.$$

Интервальный прогноз на  $h$  шагов вперёд:

$$[\bar{y} - 1.96\hat{\sigma}; \bar{y} + 1.96\hat{\sigma}]$$



# Случайное блуждание

## Наивная модель

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

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , задано стартовое  $y_1$ .

# Случайное блуждание

## Наивная модель

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

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , задано стартовое  $y_1$ .

Переформулируем:  $y_t - y_{t-1} = \Delta y_t = u_t$ .

# Случайное блуждание

## Наивная модель

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

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , задано стартовое  $y_1$ .

Переформулируем:  $y_t - y_{t-1} = \Delta y_t = u_t$ . **Оценки:**

$$\hat{\sigma}_{ML}^2 = \frac{\sum (\Delta y_i - \overline{\Delta y})^2}{T - 1}.$$

# Случайное блуждание

## Наивная модель

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

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , задано стартовое  $y_1$ .

Переформулируем:  $y_t - y_{t-1} = \Delta y_t = u_t$ . **Оценки:**

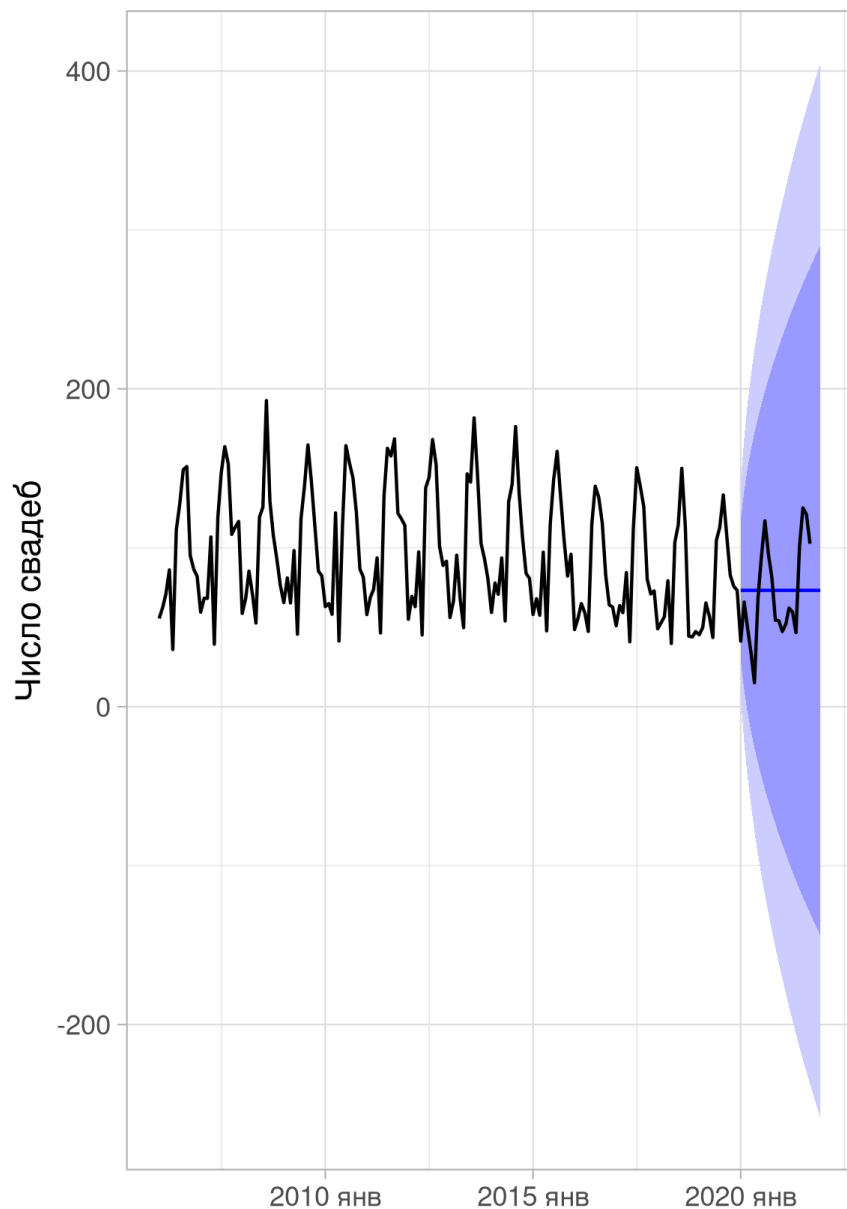
$$\hat{\sigma}_{ML}^2 = \frac{\sum (\Delta y_i - \overline{\Delta y})^2}{T - 1}.$$

**Интервальный прогноз** на  $h$  шагов вперёд:

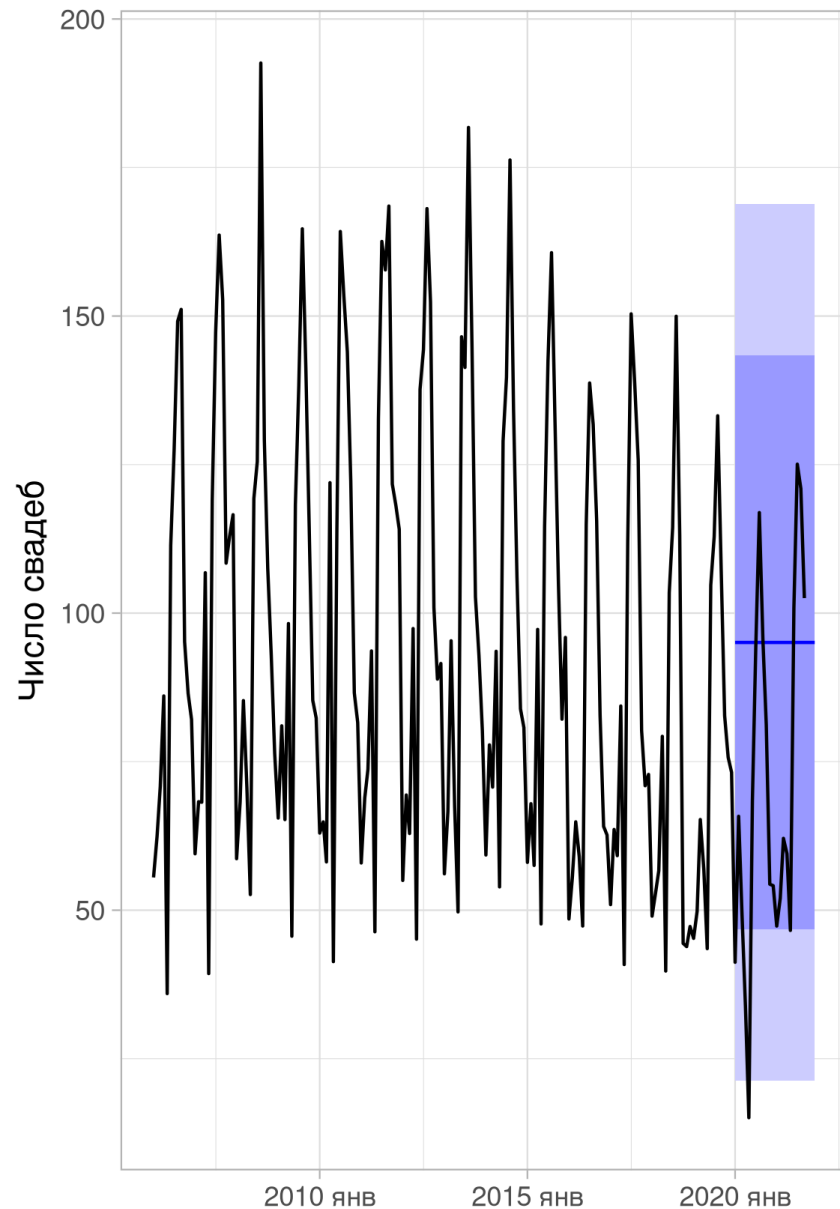
$$[y_T - 1.96\hat{\sigma}\sqrt{h}; y_T + 1.96\hat{\sigma}\sqrt{h}]$$

# Первые прогнозы!

Случайное блуждание



Модель независимых наблюдений



# Сезонное случайное блуждание

## Сезонная наивная модель

$$y_t = y_{t-12} + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , заданы  $y_1, \dots, y_{11}$ .

# Сезонное случайное блуждание

## Сезонная наивная модель

$$y_t = y_{t-12} + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , заданы  $y_1, \dots, y_{11}$ .

Переформулируем:  $y_t - y_{t-12} = \Delta_{12}y_t = u_t$ .

# Сезонное случайное блуждание

## Сезонная наивная модель

$$y_t = y_{t-12} + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , заданы  $y_1, \dots, y_{11}$ .

Переформулируем:  $y_t - y_{t-12} = \Delta_{12}y_t = u_t$ . **Оценки:**

$$\hat{\sigma}_{ML}^2 = \frac{\sum (\Delta_{12}y_i - \overline{\Delta_{12}y})^2}{T - 12}.$$



# Сезонное случайное блуждание

## Сезонная наивная модель

$$y_t = y_{t-12} + u_t,$$

где  $u_t$  — белый шум,  $u_t \sim \mathcal{N}(0; \sigma^2)$ , заданы  $y_1, \dots, y_{11}$ .

Переформулируем:  $y_t - y_{t-12} = \Delta_{12}y_t = u_t$ . **Оценки:**

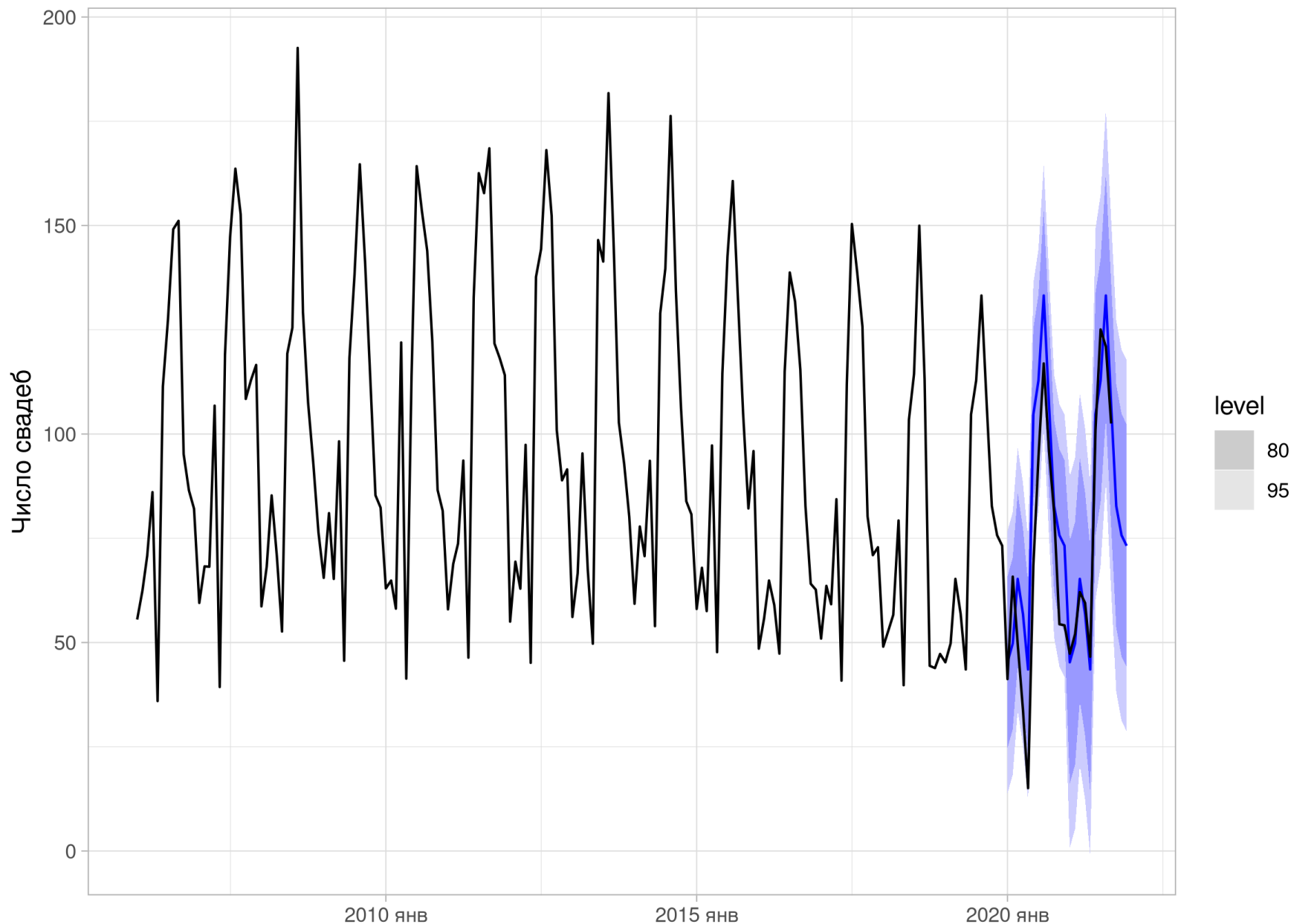
$$\hat{\sigma}_{ML}^2 = \frac{\sum (\Delta_{12}y_i - \overline{\Delta_{12}y})^2}{T - 12}.$$

**Интервальный прогноз** на  $h$  **сезонов** вперёд:

$$[y_T - 1.96\hat{\sigma}\sqrt{h}; y_T + 1.96\hat{\sigma}\sqrt{h}]$$

# Уже неплохо!

Сезонный наивный прогноз числа свадеб на 2 года



# Зачем нужны наивные модели?

- Идеи для сложных моделей.

Модели стационарных рядов похожи на модель независимых наблюдений.

Модели нестационарных рядов похожи на случайное блуждание.

# Зачем нужны наивные модели?

- Идеи для сложных моделей.

Модели стационарных рядов похожи на модель независимых наблюдений.

Модели нестационарных рядов похожи на случайное блуждание.

- База для сравнения.

При оценке сложной модели очень важно иметь базу сравнения.

# Зачем нужны наивные модели?

- **Идеи** для сложных моделей.

Модели **стационарных рядов** похожи на модель независимых наблюдений.

Модели **нестационарных рядов** похожи на случайное блуждание.

- **База для сравнения.**

При оценке сложной модели очень важно иметь базу сравнения.

- **Помощники** других моделей.

Можно **усреднить прогнозы** сложной модели и наивной сезонной!

# Наивные модели: итоги

- Белый шум — то, что не охота моделировать.

# Наивные модели: итоги

- Белый шум — то, что не охота моделировать.
- Независимые наблюдения и случайное блуждание.

# Наивные модели: итоги

- Белый шум — то, что не охота моделировать.
- Независимые наблюдения и случайное блуждание.
- Идеи, составные части и помощники других моделей.



# Наивные модели: итоги

- Белый шум — то, что не охота моделировать.
- Независимые наблюдения и случайное блуждание.
- Идеи, составные части и помощники других моделей.
- База для сравнения.