

# Выбор суперпозиций моделей при прогнозировании временных рядов

Никита Денисович Уваров

Московский физико-технический институт  
Физтех-школа прикладной математики и информатики  
Факультет управления и прикладной математики  
Кафедра «Интеллектуальные системы»

Научный руководитель: д.ф.-м.н. В.В. Стрижов

Выпускная квалификационная работа бакалавра

Москва 2018

## Задача

Выбор оптимальной суперпозиции прогностических моделей для краткосрочного прогнозирования объемов железнодорожных грузовых перевозок.

## Цель

Предложить способ выбора прогностической модели, адекватно описывающей временные ряды с асимметрично распределённым шумом.

## Проблема

В зашумлённых временных рядах точность прогноза существенно зависит от распределения ошибки.

## Предложение

Построить суперпозиции моделей временного ряда и ошибки.

- Журавлев Ю. И., Рудаков К. В., Корчагин А. Д., Кузнецов М. П., Мотренко А. П., Стенина М. М., Стрижов В. В. Методы прогнозирования временных рядов на примере железнодорожных грузоперевозок // Вестник Российской академии наук. 2016.
- А.П. Мотренко, М.М. Стенина, К.В. Рудаков, М.П. Кузнецов. Выбор оптимальной модели прогнозирования объёмов грузовых железнодорожных перевозок. 2015.
- Н.Э. Голяндина. Метод «Гусеница»-SSA: прогноз временных рядов, 2004.

## Задача ретроспективного прогноза

По известному отрезку ряда  $\mathbf{x} = [x_1, \dots, x_T]^T$  построить прогноз  $\hat{x}_{T+1}$  ряда в момент времени  $T + 1$ .

$$\varepsilon_t = x_t - \hat{x}_t, \quad \varepsilon_t \sim \text{асимм. распр.}, \quad E(\varepsilon_t) \neq 0, \quad D(\varepsilon_t) = \sigma^2$$

## Прогноз суперпозицией моделей

$$\hat{x}_{T+1} = f \circ g(\hat{\mathbf{w}}, x_T, x_{T-1}, \dots, x_1),$$

$$f : x_t \rightarrow \hat{x}_{t+1}^f, \quad g : x_{t+1}^f \rightarrow \hat{x}_{t+1}^{f,g}.$$

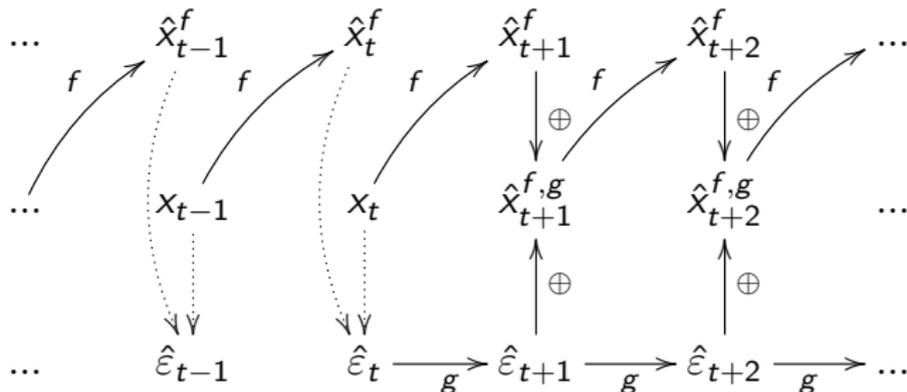
$f, g \in \mathfrak{F}$  – семейство моделей-кандидатов,  $\hat{\mathbf{w}} \in \mathbb{R}^n$ .

## Оптимальная суперпозиция

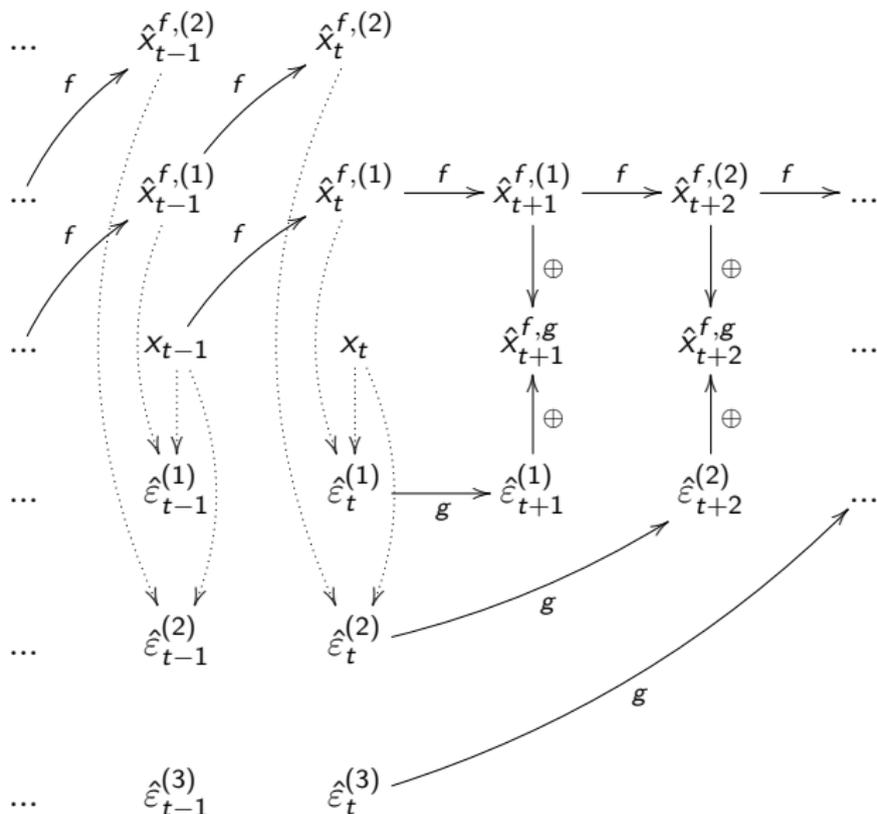
$$(f, g, \hat{\mathbf{w}}) = \arg \min_{f \in \mathfrak{F}, g \in \mathfrak{F}, \hat{\mathbf{w}}} \text{MAPE}(f \circ g(\hat{\mathbf{w}}, x_T, x_{T-1}, \dots, x_1))$$

# Метод параллельного прогноза

- 1 С помощью  $f$  вычисляются  $n$  прогнозов  $\hat{x}_t^f, \dots, \hat{x}_{t-n+1}^f$
- 2 Вычисляются  $t$  остатков  
 $\hat{\varepsilon}_{t-k} = x_{t-k} - \hat{x}_{t-k}^f, k = 0, \dots, t - 1.$
- 3 С помощью  $g$  прогнозируются остатки  $\hat{\varepsilon}_{t+i}$  на  $h$  отсчетов вперед.
- 4 Конечный прогноз  $\hat{x}_{t+i}^{f,g} = \hat{x}_{t+i}^f + \hat{\varepsilon}_{t+i}.$



# Метод прогноза с накоплением



- 1 С помощью базовой функции  $f$  вычисляется ретроспективный прогноз  $\hat{x}_{t+1}^{f,(1)}, \dots, \hat{x}_{t+i}^{f,(i)}$  с горизонтом прогнозирования  $i$ , каждый — на глубине  $i$ .
- 2 С помощью базовой функции  $f$  вычисляется  $r$  наборов прогнозов конца истории,  $\hat{x}_t^{f,(i)}, \dots, \hat{x}_{t-n+1}^{f,(i)}$ , каждый набор — на глубине  $i$ ,  $i = 1, \dots, r$ .
- 3 Вычисляется  $r$  наборов остатков  $\hat{\varepsilon}_t^{(i)}, \dots, \hat{\varepsilon}_{t-n+1}^{(i)}$

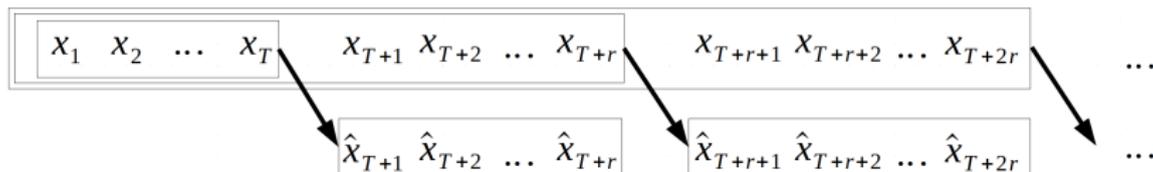
$$\hat{\varepsilon}_{t-k}^{(i)} = x_{t-k} - \hat{x}_{t-k}^{f,(i)}, \quad i = 1, \dots, r.$$

- 4 С помощью функции  $g$  прогнозируются остатки  $\hat{\varepsilon}_{t+i}^{(i)}$ , каждый прогноз выполняется на одну точку и использует вычисленную последовательность  $\hat{\varepsilon}_t^{(i)}, \dots, \hat{\varepsilon}_{t-n(g)+1}^{(i)}$ .
- 5 Выполняется подсчет конечных прогнозов

$$\hat{x}_{t+i}^{f,g} = \hat{x}_{t+i}^{f,(i)} + \hat{\varepsilon}_{t+i}^{(i)}.$$

# Алгоритм блочного прогноза

- 1 По известной обучающей части ряда  $x_{1 \div T}$  построим прогноз на  $r$  отсчётов вперёд алгоритмом поточечного прогнозирования.
- 2 Будем теперь считать первые  $r$  значений проверочной части ряда известными. По  $x_{1 \div T+r}$  построим прогноз на  $r$  отсчётов вперёд.
- 3 Повторим шаг 2, каждый раз считая очередные  $r$  проверочных значений известными, пока не дойдём до конца проверочной части.



## Теорема

Если распределение ошибки симметрично (гауссово, лапласово, ...), а функция потерь (критерий качества) модели, используемой для прогнозирования регрессионных остатков, предполагает симметричное распределение ошибки, использование суперпозиции с данной моделью не даст уменьшения функции потерь по сравнению с базовой моделью.

## Теорема

Если распределение ошибки асимметрично, существует функция потерь  $L(x)$  такая, что суперпозиция базовой модели  $f$  и квантильной регрессии  $g$  с данной функцией потерь  $L$  дают ошибку прогноза, не превосходящую ошибки прогноза базовой модели  $f$ .

# Семейство моделей-кандидатов $\mathfrak{F}(1)$

- 1 Экспоненциальное сглаживание (параметр сглаживания  $\alpha$ )

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t$$

- 2 Метод Кростена (параметр сглаживания  $\alpha$ ),

$$\hat{z}_t = \alpha x_t + (1 - \alpha)\hat{z}_{t-1}, \quad \hat{q}_t = \alpha q_t + (1 - \alpha)\hat{q}_{t-1}, \quad \hat{x}_{t+1} = \frac{\hat{z}_t}{\hat{q}_t}.$$

- 3 Ядерное сглаживание (ядро  $K$ , ширина окна  $h$ ),

$$\omega_i(r) = K\left(\frac{T-i+r-1}{h}\right), \quad a_h(r, X^t) = \frac{\sum_{i=1}^t x_i \omega_i(r)}{\sum_{i=1}^t \omega_i(r)}, \quad \hat{x}_{t+1} = a_h(t+1, X^t).$$

- 4 ARIMA (p,d,q),

$$\nabla^d x_t = c + \sum_{i=1}^p a_i \nabla^d x_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t.$$

- 5 Квантильная регрессия (функция штрафа  $L$ ),

$$\hat{x}_{t+1} = \arg \min_x \sum_{k=1}^K L(x, y_k) \cdot g_k.$$

- 6 Гусеница «SSA» (ширина окна  $l$ , число спектральных компонент  $m$ ),

$$\begin{bmatrix} x_1 & x_2 & \dots & x_l \\ x_2 & x_3 & \dots & x_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-l+1} & x_{T-l+2} & \dots & x_T \end{bmatrix}, \quad \mathbf{X} \approx \mathbf{USV}^T, \quad \mathbf{S} = \text{diag}(s_1, \dots, s_m),$$

$$\mathbf{p} = \frac{1}{1-\nu^2} \sum_{i=1}^m \pi_i \mathbf{u}_i,$$

$$\hat{x}_{t+1} = \sum_{i=1}^m x_{t+1-i} p_i.$$

- 7 LSTM (длина истории, эвристика).

## Цель

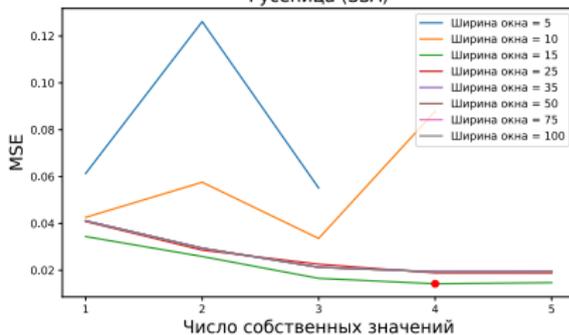
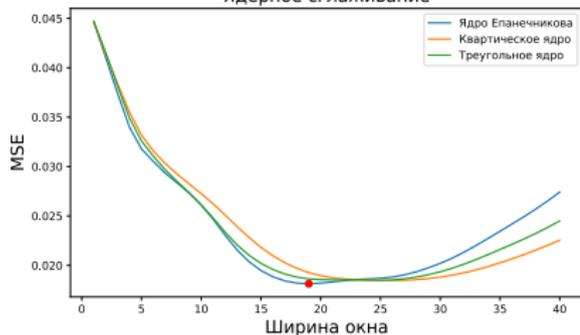
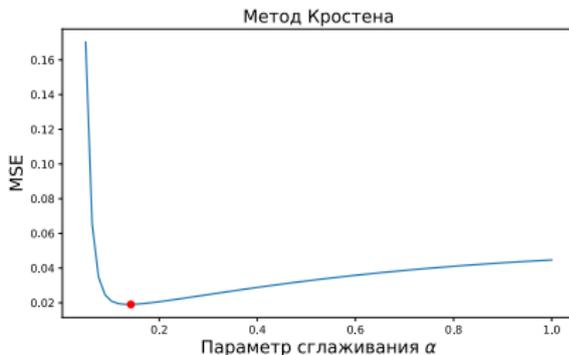
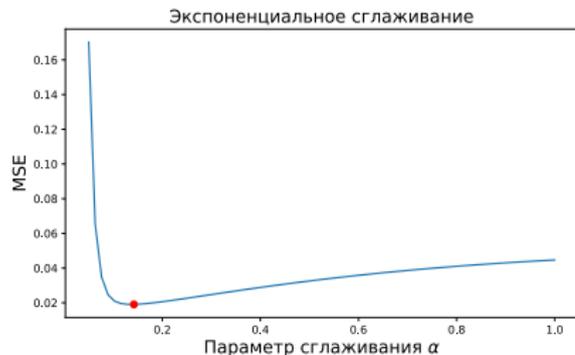
Сравнить качество прогноза суперпозиций прогностических моделей с качеством прогноза базовых моделей.

- На реальных данных
- На синтетических данных в зависимости от степени асимметричности распределения.

## Данные

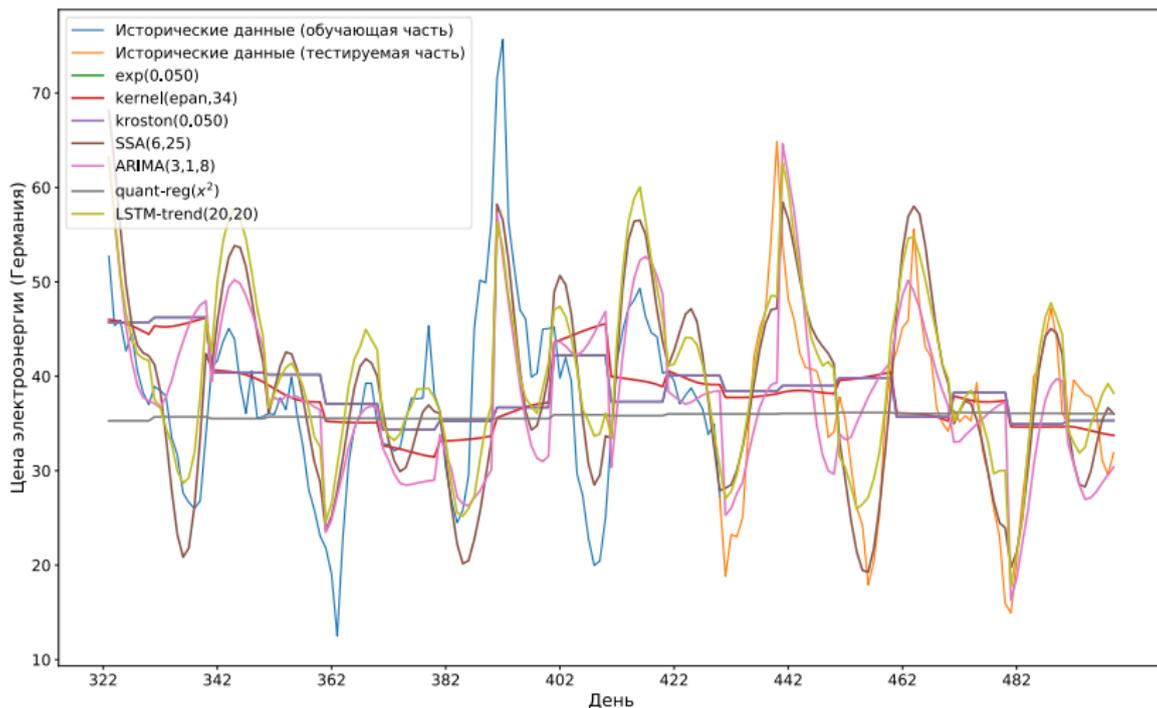
- 1 Пассажирские авиаперевозки по месяцам,
- 2 Объём железнодорожных перевозок нефти в Омской области по дням,
- 3 Цена на электричество в Германии по часам,
- 4 Потребление электроэнергии в Польше по часам.

# Зависимость качества от структурных параметров



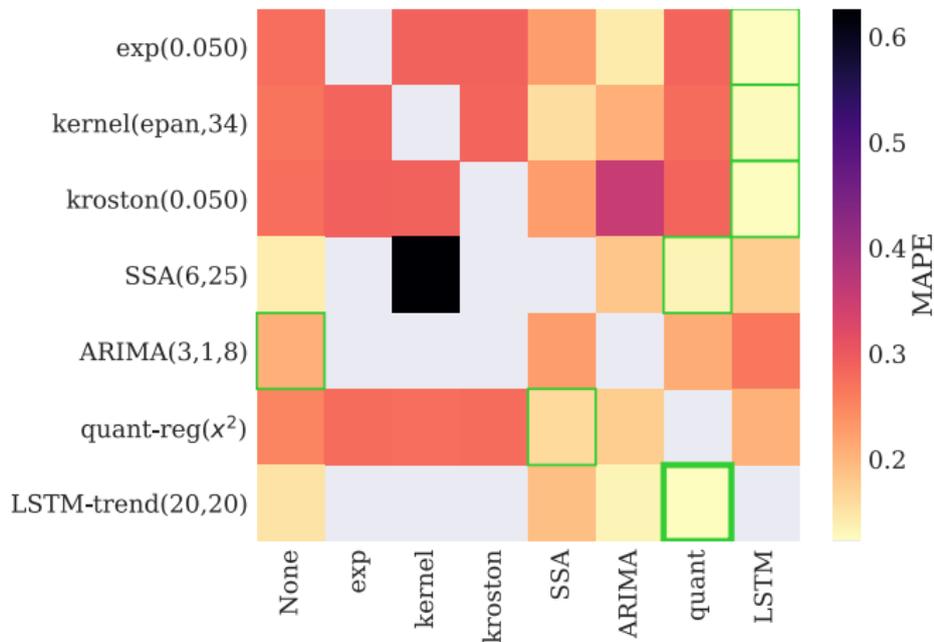
Выпуклый характер зависимости показывает, что используемые сетки структурных параметров позволяют в эксперименте определить значения, близкие к оптимальным.

# Прогнозы основных моделей



Предсказания всех моделей после подбора оптимальных структурных параметров адекватно описывают периодическую компоненту временного ряда (компоненту  $f$ ).

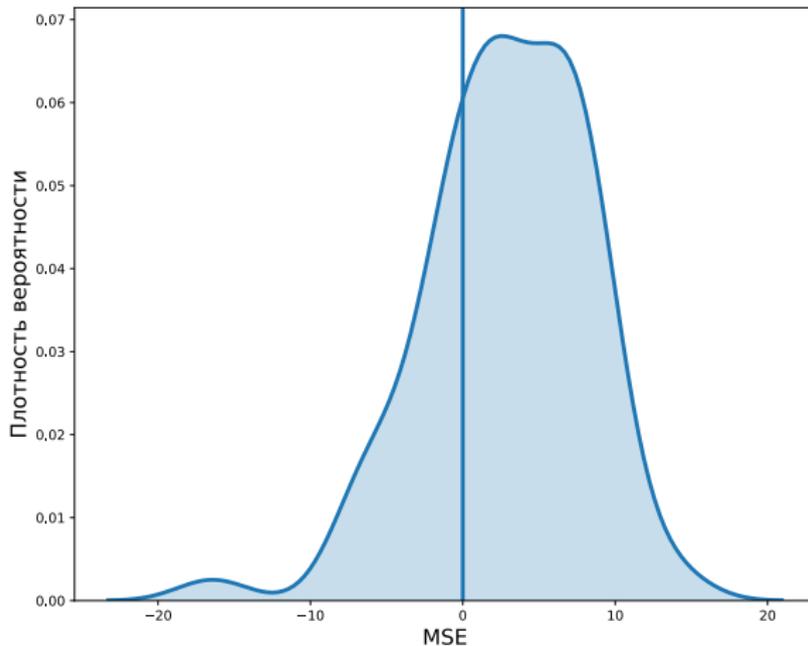
# Матрица качества суперпозиций



Лучшие базовые модели – SSA и LSTM.

Лучшие модели остатков – квантильная регрессия и LSTM.

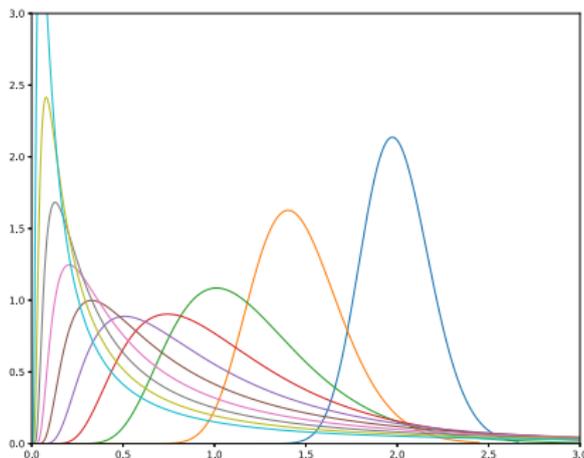
# Эмпирическая функция распределения ошибки



Наблюдаемое распределение ошибки имеет ненулевое матожидание и асимметрично.

Однопараметрическое (параметр  $l$ ) семейство распределения шума: распределение Вальда с параметрами

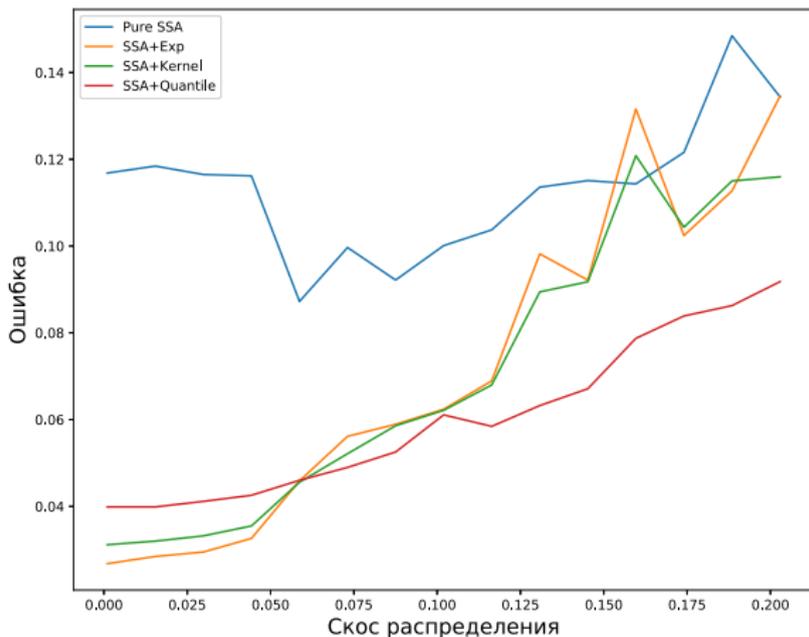
$$\mu = 1 + \frac{l}{30}, \quad \lambda = \frac{l^2}{4}, \quad f(x, \mu, \lambda) = \left[ \frac{\lambda}{2\pi x^3} \right]^{1/2} \exp \left\{ \frac{-\lambda(x - \mu)^2}{2\mu^2 x} \right\}.$$



Мера скоса распределения

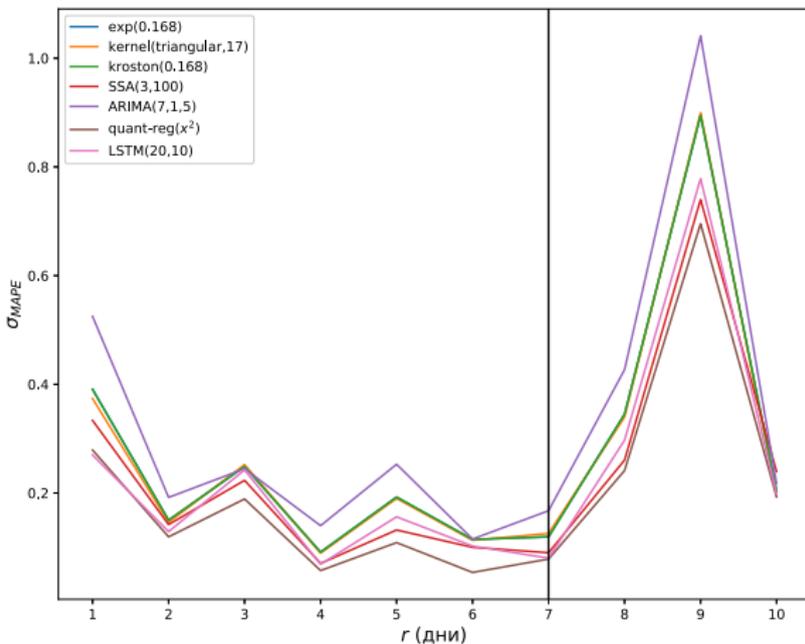
$$S = |\mathbb{P}(x > x_{\text{н.в.}}) - \mathbb{P}(x < x_{\text{н.в.}})|.$$

# Зависимость ошибки от скоса распределения



Суперпозиция с квантильной регрессией даёт меньшую ошибку по сравнению с базовой моделью при прогнозировании ряда с асимметричным шумом.

# Зависимость ошибки от горизонта прогнозирования



Горизонт прогнозирования определяется по «правилу сломанной трости»; например, на рисунке  $h = 7$ .

- Предложено два способа построения суперпозиций базовых алгоритмов (экспоненциальное сглаживание, метод Кростена, SSA, ARIMA, квантильная регрессия, LSTM).
- Исследованы свойства суперпозиций прогностических моделей при условии симметричного и асимметричного распределения регрессионных остатков.
- Построены матрицы качества суперпозиций, распределение остатков модели, определён горизонт прогнозирования.
- Показано, что использование суперпозиций может повышать качество прогноза.
- Исследована зависимость ошибки от степени асимметричности распределения шумовых остатков.
- Статья по теме дипломной работы подана в журнал «Вестник московского университета, сер.15. Вычислительная математика и кибернетика»