

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (государственный университет)
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»
ПРИ ВЫЧИСЛИТЕЛЬНОМ ЦЕНТРЕ ИМ. А. А. ДОРОДНИЦЫНА РАН

Уваров Никита Денисович

**Выбор суперпозиций моделей
при прогнозировании временных рядов**

03.03.01 — Прикладные математика и физика

БАКАЛАВРСКАЯ ДИССЕРТАЦИЯ

Научный руководитель:
д.ф.-м.н., профессор
Стрижов Вадим Викторович

Москва
2018 г.

Содержание

1	Введение	3
1.1	Задача построения оптимальной суперпозиции	5
2	Суперпозиция при ретроспективном прогнозе	6
2.1	Метод параллельного прогноза	7
2.2	Метод прогноза с накоплением	8
2.3	Алгоритм поточечного прогнозирования по запросу	9
2.4	Алгоритм блочного прогноза	9
3	Набор моделей-претендентов	10
3.1	Экспоненциальное сглаживание	10
3.2	Ядерное сглаживание	10
3.3	Метод Кростена	11
3.4	Метод анализа спектральных компонент Гусеница	11
3.5	Модель ARIMA(p, d, q)	12
3.6	Квантильная регрессия	12
4	Вычислительный эксперимент	13
4.1	Подбор структурных параметров моделей	14
4.2	Перебираемые структурные параметры	14
4.3	Результаты экспериментов	16
4.3.1	Результаты эксперимента: airline	17
4.3.2	Результаты эксперимента: synthetic	20
4.3.3	Результаты эксперимента: german	23
4.3.4	Результаты эксперимента: railroads	26
4.3.5	Горизонт прогнозирования в экспериментах	29
4.4	Эксперимент: зависимость качества от асимметричности распределения	32
	Заключение	33

Аннотация

Рассматривается задача выбора оптимальной прогностической системы моделей для краткосрочного прогнозирования объемов железнодорожных грузовых перевозок. Исторические данные представляют собой временные ряды, отражающие объемы перевозок различных типов грузов между станциями. Свойствами данных временных рядов являются их высокая волатильность, зашумленность и нестационарность. Для построения прогнозов предлагается создать систему, которая бы выбирала оптимальную суперпозицию прогностических моделей с учетом особенности исторических данных. В качестве возможных моделей для включения в суперпозицию рассматриваются модели скользящего среднего, экспоненциального и ядерного сглаживания, векторной авторегрессии, интегрированные модели авторегрессии и скользящего среднего, а также метод Кростона и LSTM.

Ключевые слова: *временные ряды, прогнозирование, суперпозиция, прогностические модели, асимметричное распределение, регрессионные остатки.*

1 Введение

Российские железные дороги играют ключевую роль в экономике страны. Железнодорожный транспорт занимает второе место, после трубопроводного, по грузообороту. На сегодняшний день около 80% своей прибыли РЖД получают именно от грузовых перевозок [1]. Поэтому перед РЖД ставится задача по привлечению грузов с конкурирующих видов транспорта. В связи с этим прогнозирование объемов перевозимых грузов имеет большое практическое значение для повышения экономической эффективности загрузки железнодорожных путей и их рационального использования. В статье [2] обосновывается необходимость в усовершенствовании и оптимизации работы, связанной с использованием железных путей, и говорится, что в текущей экономической ситуации дальнейшее использование экстенсивных методов развития производства и сферы услуг не приводит к желаемой экономической отдаче. Использование интенсивного подхода, связанного с привлечением современных научных методов анализа данных и оптимизации, должно улучшить показатели эффективности работы РЖД и увеличить величину добавленной стоимости услуг железнодорожных грузоперевозок.

Не все временные ряды подчиняются гипотезе о нормальном распределении шумовой компоненты [3]. Для повышения качества прогноза полезно использование отдельной модели. Так, в [4] для прогнозирования остатков строится марковская цепь, что улучшает точность прогноза электроэнергии, генерируемой ветряной фермой, на 10%. Распространён подход коррекции Фурье-спектра остатков во временных рядах туристической отрасли [5].

Целью работы является построение прогностической модели, способной описывать временные ряды с асимметрично распределённым шумом. Предлагается использовать идею суперпозиции двух моделей, изложенную в [6]. Одновременно используются две модели: одна из них предсказывает основное, обычно гладкое, поведение временного ряда, а вторая корректирует прогноз с целью минимизации ошибки от наличия шума. В качестве обучающей выборки для второй модели выступают ошибки прогнозирования первой модели. При прогнозировании на более чем один отсчёт в будущее, возникает возможность использования для построения прогноза первой модели как её собственного прогноза, так и скорректированного прогноза суперпозиции. В вычислительном эксперименте на исторических данных качество прогноза суперпозиции сравнивается с качеством прогноза базовых алгоритмов, а на синтетических данных исследуется зависимость качества прогноза от меры асимметрии шума.

Исследуемыми подходами к прогнозированию временных рядов являются модели авторегрессии и скользящего среднего ARIMA [7], анализа спектральных компонент SSA [8], а также подходы на основе нейронных сетей, особенно LSTM [9]. Из более простых подходов исследуются модели экспоненциального [10] и ядерного [11] сглаживания. Перечисленные алгоритмы составляют семейство базовых алгоритмов, используемых для построения суперпозиций в данной работе. Оптимальная комбинация алгоритмов зависит от решаемой задачи, и для её нахождения перебираются все возможные пары моделей: строится матрица качества прогноза суперпозиций.

Исторические данные, предлагаемые для исследования, представляют собой временные ряды объемов грузоперевозок по российским железным дорогам [12]. Имеется информация о дате погрузки товара, станции отправления, станции назначения, количестве вагонов, коде перевозимого груза, роде вагонов, в которых этот груз перевозился, суммарном весе груза в тоннах и признаке маршрутной отправки. Ряды отличаются высокой волатильностью, неравномерностью, а также зашумленностью. Для некоторых временных рядов свойственна нестационарность. Кроме того, между

некоторыми парами станций перевозки производятся крайне нерегулярно, и число нулевых измерений в таких парах превышает число ненулевых. Проверка временных рядов на стационарность выполняется с помощью теста Дики-Фуллера.

Данные о посуточной загруженности железнодорожных путей представлены таблицей 1, в которой числится информация о времени перевозки, станциях, между которыми выполняются перевозки, а также коде и объеме перевозимого груза. Представлена годовая история перевозок между 78 регионами и 4000 станциями по 43 наименованиям, среди которых нефть, руда, торф, автомобили, хлопок, сахар, зерно.

Помимо временных рядов РЖД, в эксперименте участвуют временные ряды пассажирских авиаперевозок [14], цены на электроэнергию в Германии [15], а также синтетическая выборка, представляющая собой синусоидальный временной ряд с аддитивным шумом, получаемым из распределения Вальда.

Таблица 1: Данные о посуточной загруженности железнодорожных путей

Дата погрузки	Код станции отправления	Код станции назначения	Код груза	Суммарный вес груза
2007-01-01	14605	831504	1	56
2007-01-01	135602	165504	18	63
2007-01-01	830304	814208	13	246

1.1 Задача построения оптимальной суперпозиции

Поставим задачу построения оптимальной прогностической модели в виде суперпозиции двух функций прогноза ряда и прогноза ошибки. Моделью $f \circ g$ назовём параметрическую функцию, которая по временному ряду $\mathbf{x} = [x_1, \dots, x_t]^T$ и фиксированным параметрам вычисляет значение \hat{x}_{t+1} ряда в момент времени $t + 1$.

Прогностической моделью называется суперпозиция, где прогноз

$$\hat{x}_{t+1} = f \circ g(\hat{\mathbf{w}}, x_t, x_{t-1}, \dots, x_1) \quad (1.1)$$

получен суперпозицией моделей f и g . Параметры $\hat{\mathbf{w}}$ оптимизированы согласно функции ошибки.

Прогностические модели выбираются из набора моделей-претендентов $f, g \in \mathfrak{F}$. Набор моделей описан в разделе 4. Невязка (регрессионный остаток) прогнозирования ε — разности $\varepsilon = x_t - \hat{x}_t$ между фактическими значениями ряда и прогнозом $x_{t+1} = f \circ g(x_t, x_{t-1}, \dots, x_1) + \varepsilon_{t+1}$. Предполагается, что математическое ожидание остатков не равно нулю, $E(\varepsilon) \neq 0$, и их дисперсия постоянна, $D(\varepsilon) = \sigma^2$.

Функция f прогнозирует тренд и сезонность, а функция g прогнозирует остаток. Конечный прогноз \hat{x}_{t+1} складывается с учетом прогноза обеих функций. Предполагается, что функция g повышает точность прогноза, полученного с помощью функции f :

$$\begin{aligned} f &: x_t \rightarrow \hat{x}_{t+1}^f, \\ g &: x_{t+1}^f \rightarrow \hat{x}_{t+1}^{f,g}. \end{aligned}$$

Требуется построить поточечный прогноз (1.1) и прогноз на r отсчётов вперед $[\hat{x}_{t+1}, \dots, \hat{x}_{t+r}]^\top$. Назовём *запросом на прогнозирование* число r , которое задаёт число точек, на которые необходимо сделать прогноз.

Назовем *горизонтом прогнозирования* число h , $h \leq r$, при котором расстояние между распределением двух случайных величин \hat{x} и x значимо после шага под номером h

$$\|\hat{x}_{t+h-1} - x_{t+h-1}\| \leq \mu, \text{ при условии } \|\hat{x}_{t+h} - x_{t+h}\| > \mu. \quad (1.2)$$

Для получения оценок параметров $\hat{\mathbf{w}}$ модели требуется минимизировать значение функции ошибки S

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{\tau=1}^t S(\mathbf{w}, x_\tau - \hat{x}_\tau). \quad (1.3)$$

Ошибка прогнозирования S оценивается с помощью MAPE (4.1).

2 Суперпозиция при ретроспективном прогнозе

В данной работе предлагаются два способа построения суперпозиции и алгоритм поточечного прогнозирования. Метод параллельного прогноза и метод прогноза с накоплением основаны на уточнении базового добавлением прогноза ошибок. В них используются две функции для прогноза: одна — для базового прогноза, вторая — для ошибок. Их отличие заключается в том, что в методе параллельного прогноза учитывается только предыдущее значение и ошибки, и базового прогноза, а в методе

прогноза с накоплением — несколько предыдущих. В алгоритме поточечного прогнозирования прогнозируется следующее значение ряда, которое затем используется как элемент истории.

2.1 Метод параллельного прогноза

Задан временной ряд \mathbf{x} и две функции f и g . Алгоритм (2.1) состоит в следующем.

1. С помощью функции f вычисляется n прогнозов конца истории $\hat{x}_t^f, \dots, \hat{x}_{t-n+1}^f$ по точкам.
2. Вычисляется n остатков $\hat{\varepsilon}_t, \dots, \hat{\varepsilon}_{t-n+1}$ в виде разницы

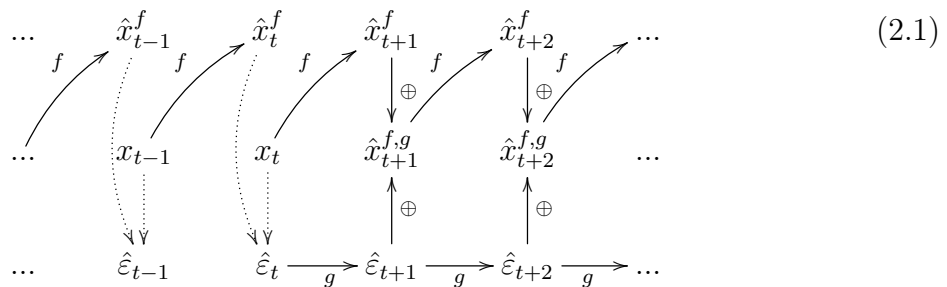
$$\hat{\varepsilon}_{t-k} = x_{t-k} - \hat{x}_{t-k}^f,$$

где n — длина исторических значений временного ряда.

3. С помощью функции g прогнозируются остатки $\hat{\varepsilon}_{t+i}$ на h отсчетов вперед.
4. Выполняется подсчет конечных прогнозов

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

с последовательным подсчетом прогноза базовой функцией f на одну точку \hat{x}_{t+i}^f .



2.2 Метод прогноза с накоплением

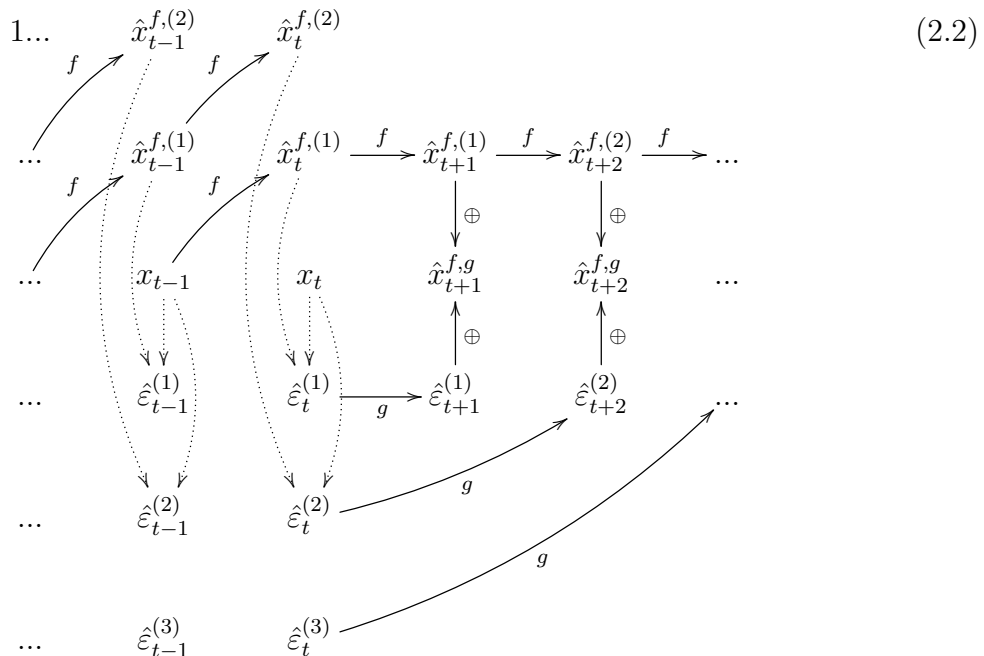
Задан временной ряд \mathbf{x} и две функции f и g , с помощью которых делается прогноз на глубине прогнозирования i . Глубина прогнозирования i — количество предшествующих элементов заданного временного ряда, на основе которых делается прогноз. Алгоритм (2.2) состоит в следующем.

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

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



2.3 Алгоритм поточечного прогнозирования по запросу

Зафиксируем время t_0 , считаем известными значения ряда $\mathbf{x}_{1 \div t_0} = [x_1, x_2, \dots, x_{t_0}]$. Вектор \mathbf{x}_{t_0+r} назовём прогнозом с запросом на горизонт прогнозирования r , $t_0+r < T$, где T — длина истории ряда.

1. С помощью базовой функции f вычисляется прогноз \hat{x}_{t+1} следующего элемента ряда.
2. Используя полученное значение \hat{x}_{t+1} и известную историю с помощью базовой функции f вычисляется прогноз \hat{x}_{t+2} .
3. Шаги повторяются до тех пор, пока не будет достигнут запрос на прогнозирование r .

2.4 Алгоритм блочного прогноза

Для учёта качества прогнозирования на запросах, превышающих единицу, предлагается строить блочный прогноз.

Для оценки качества прогноза временной ряд разбивается на обучающую $\mathbf{x}_{1 \div t_0}$ и проверочную $\mathbf{x}_{t_0+1 \div t_1}$ части. Для вычисления средней ошибки строится совокупность из $t_1 - t_0$ прогнозов, в которой i -ый прогноз ($0 \leq i < t_1 - t_0$) строится по известному ряду $\mathbf{x}_{1 \div t_0+i}$ на 1 отсчёт вперёд.

Зафиксируем длину блока r — максимальный запрос на прогнозирование, который не превышает горизонта прогнозирования используемых алгоритмов. Для сравнения с проверочной частью временного ряда составим следующий временной ряд-прогноз:

1. По известной обучающей части ряда $\mathbf{x}_{1 \div t_0}$ построим прогноз на r отсчётов вперёд алгоритмом поточечного прогнозирования.
2. Будем теперь считать первые r значений проверочной части ряда известными. По $\mathbf{x}_{1 \div t_0+r}$ 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.

В полученном блочном прогнозе одинаково часто встречаются значения, полученные в результате прогнозирования на $1, 2, \dots, r$ отсчётов вперёд. Таким образом, его качество является усреднённым качеством прогнозирования алгоритма на запросах от 1 до r .

3 Набор моделей-претендентов

Прогностическая модель строится как суперпозиция (1.1) пар моделей, описанных ниже.

3.1 Экспоненциальное сглаживание

Сложность прогнозирования обусловлена высокой волатильностью исторических данных по объёмам перевозок. Применение экспоненциального сглаживания к такому ряду позволяет выровнять ряд и определить его тенденцию. Экспоненциальное сглаживание представляется фильтром, на вход которого последовательно поступают члены исходного ряда, а на выходе формируются текущие значения экспоненциальной средней

$$z_t = \alpha x_t + (1 - \alpha)z_{t-1},$$

Чем меньше параметр сглаживания ряда $\alpha \in (0, 1)$, тем в большей степени фильтруются, подавляются изменения исходного ряда и шума.

Таким образом,

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t. \tag{3.1}$$

3.2 Ядерное сглаживание

С помощью функции ядра зададим веса объектов:

$$\omega_i(r) = K \left(\frac{T - i + r - 1}{h} \right),$$

где h — ширина окна сглаживания, r — запрос на прогнозирование, K — ядро, гладкая, невозрастающая, ограниченная функция $[0, +\infty) \rightarrow [0, +\infty)$.

По формуле Надарая-Ватсона

$$a(r, X^l) = \frac{\sum_{i=1}^l x_i \omega_i(r)}{\sum_{i=1}^l \omega_i(r)}.$$

В данной работе используется ядро Епанечникова: $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$.

Прогноз

$$\hat{x}_{t+1} = a_h(t + 1, X^l).$$

3.3 Метод Кростена

Для прогнозирования прерывистых рядов используют метод Кростена, в соответствии с которым исходный ряд разбивается на два ряда, один из которых описывает ненулевой объем перевозимых товаров, а второй — интервалы между ненулевыми перевозками (промежутки времени, в течение которых перевозки отсутствовали). Затем каждый ряд прогнозируется в отдельности экспоненциальным сглаживанием.

Вместе с временным рядом ненулевого спроса $\mathbf{x} = [x_1, \dots, x_T]^T$ известен временной ряд интервалов времени между ненулевым спросом $\mathbf{q} = [q_1, \dots, q_T]^T$. Экспоненциальное сглаживание обоих рядов имеет вид:

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

$$\hat{q}_t = \alpha q_t + (1 - \alpha)\hat{q}_{t-1},$$

$$\hat{x}_{t+1} = \frac{\hat{z}_t}{\hat{q}_t}.$$

Здесь $\alpha \in (0, 1)$ — параметр сглаживания, как и в экспоненциальном сглаживании (3.1).

В основе метода Кростена лежит предположение о том, что все перевозки являются случайной величиной с нормальным распределением, а факт того, что перевозка произошла, является случайной величиной с распределением Бернулли.

3.4 Метод анализа спектральных компонент Гусеница

Поставим в соответствие временному ряду $\mathbf{x} = [x_1, \dots, x_T]^T$ траекторную матрицу \mathbf{X} вида

$$\begin{bmatrix} x_1 & x_2 & \dots & x_l \\ x_2 & x_3 & \dots & x_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-l+1} & x_{T-l+2} & \dots & x_T \end{bmatrix}.$$

Это соответствие — взаимно однозначное, поскольку усреднением по антидиагоналям из матрицы \mathbf{X} получается исходный временной ряд. Построим сингулярное разложение матрицы \mathbf{X} :

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \quad \mathbf{S} = \text{diag}(s_1, \dots, s_l), \quad s_1 \geq s_2 \geq \dots \geq s_l,$$

где s_1, \dots, s_l — собственные числа матрицы $\mathbf{X}^\top\mathbf{X}$, соответствующие частотам спектра временного ряда \mathbf{x} . Возьмём компоненты s_1, \dots, s_m , $m \leq l$, соответствующие наибольшим собственным значениям $\mathbf{X}^\top\mathbf{X}$; получим приближение траекторной матрицы матрицей меньшего ранга $\mathbf{X}' = \mathbf{U}'\mathbf{S}'\mathbf{V}'^\top$. Прогноз получается дописыванием к траекторной матрице \mathbf{X}' столбца, не меняющего её сингулярного разложения, с последующим восстановлением временного ряда усреднением траекторной матрицы по антидиагоналям. Для \hat{x}_{t+1} получается линейное рекуррентное соотношение порядка m ; алгоритм вычисления коэффициентов авторегрессии приведён, например, в [8].

3.5 Модель ARIMA(p, d, q)

Интегрированная модель авторегрессии и скользящего среднего является расширением моделей ARMA для нестационарных временных рядов, которые можно преобразовать в стационарные взятием разностей некоторого порядка от исходного временного ряда. Модель ARIMA(p, d, q) означает, что разности временного ряда порядка d подчиняются модели ARMA(p, q).

Модель ARIMA с порядком авторегрессии p , порядком дифференцирования d и порядком скользящего среднего q для временного ряда x_t имеет вид

$$\nabla^d x_t = c + \sum_{i=1}^p a_i \nabla^d x_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t, \quad (3.2)$$

c, a_i, b_j — параметры модели, ∇^d — оператор конечной разности порядка d , то есть разность соседних отсчётов, взятая d раз, а ε_t — случайный нормально распределённый шум в момент времени t .

Более подробное описание модели, а также алгоритм обучения и построения прогноза можно найти в [14].

3.6 Квантильная регрессия

Метод квантильной регрессии предполагает, что все отсчёты временного ряда независимы и подчиняются неизвестному вероятностному распределению, а штраф за несоответствие прогнозируемого значения \hat{x}_{t+1} реальному значению x_t задаётся кусочно линейной, асимметричной функцией потерь $L(\hat{x}, x)$. В данной работе в эксперименте по построению суперпозиций используются квадратичный $L(\hat{x}, x) = (\hat{x} - x)^2$

и линейный $L(\hat{x}, x) = |\hat{x} - x|$ штрафы, а в эксперименте по исследованию асимметричного шума используется асимметричная кусочно-линейная функция потерь (4.3).

По значениям ряда строится гистограмма

$$\{(y_k, g_k)\}_{k=1}^K, \quad (3.3)$$

где K — число интервалов $[y_k^{\min}, y_k^{\max}]$ со средним значением y_k , на которые разбита ось значений ряда \mathbf{x} , g_k — высота столбца гистограммы на интервале y_k , равная взвешенной сумме количества точек ряда, попавших в этот интервал. Прогноз минимизирует эмпирический риск

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

4 Вычислительный эксперимент

Цель эксперимента – анализ качества и устойчивости прогноза с использованием суперпозиции прогностических моделей при условии асимметричного распределения регрессионных остатков. Используются следующие выборки: временные ряды грузоперевозок [12], потребление электроэнергии [16], цена на электричество [15], число пассажиров авиалиний [14], а также синтетическая (синус, зашумлённый распределением Вальда).

При тестировании суперпозиций исследуются прогнозы, построенные базовыми моделями и по их суперпозициям. Приводятся графические иллюстрации прогнозов, а также численные значения ошибки. В качестве функций ошибки взяты функция MAPE (4.1) и функция MSE (4.2):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\varepsilon_i}{x_i} \right|, \quad (4.1)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2. \quad (4.2)$$

Во всех экспериментах измеряется качество блочного прогноза (2.4).

4.1 Подбор структурных параметров моделей

Для подбора оптимальных структурных параметров моделей, включённых в семейство, используется схема валидации на отложенной части временного ряда:

1. Каждая модель обучается на 75% первых отсчётов обучающей части временного ряда со всевозможными осмысленными значениями совокупности структурных параметров (grid search). Используемые сетки структурных параметров приведены в 4.2.
2. Для выбора оптимального набора структурных параметров сравнивается качество предсказания по метрике MSE (4.2) на отложенных 25% последних отсчётов обучающей части временного ряда.

Заметим, что проверочная часть временного ряда (фиксированное количество последних измерений временного ряда) не используется при подборе оптимальных структурных параметров, чтобы исключить возможность влияния переобучения на результаты эксперимента.

При подборе оптимальных структурных параметров моделей-суперпозиций с целью сокращения числа вычислений сначала подбираются оптимальные структурные параметры основной модели, а затем – модели, используемой для прогнозирования остатков.

4.2 Перебираемые структурные параметры

1. Экспоненциальное сглаживание, метод Кростена
 - (а) α (коэффициент сглаживания) перебирается в диапазоне от 0 до 1; всего проверяется 44 значения, 20 из которых расположены в диапазоне от 0.05 до 0.3, который соответствует наиболее частому оптимальному коэффициенту сглаживания в наших экспериментах (при условии, что он отличен от единицы).
2. Ядерное сглаживание
 - (а) Ядро сглаживания: перебираются ядро Епанечникова, треугольное и квартическое ядра.
 - (б) Ширина окна перебирается от 1 до 40.

3. Гусеница (SSA)

- (a) Число используемых компонент спектра перебирается от 1 до 20; всегда используются компоненты с наибольшими по модулю коэффициентами в разложении по ним временного ряда.
- (b) Ширина окна перебирается по сетке: 5, 10, 15, 25, 35, 50, 75, 100.

4. ARIMA(p,d,q)

- (a) Число членов авторегрессии p перебирается от 1 до 10.
- (b) Порядок дифференцирования d перебирается по сетке 0, 1: допускается не более чем однократное дифференцирование временного ряда.
- (c) Число членов скользящего среднего q перебирается от 1 до 10.

5. Квантильная регрессия

Функция штрафа выбирается из двух симметричных функций-кандидатов: квадратичного штрафа $l(x) = x^2$ и линейного штрафа $l(x) = |x|$.

6. LSTM

Для предсказания следующего отсчёта временного ряда по w предыдущим отсчётам (w – длина доступной истории) используется сеть, состоящая из двух слоёв: одного LSTM-слоя из k нейронов и одного полносвязного.

Для повышения качества предсказания LSTM как основной модели на нестационарных датасетах (например, *airline*) с ярко выраженным трендом, дополнительно проверяется целесообразность использования одной из двух эвристик:

- (a) Предварительное однократное дифференцирование ряда – если эвристика была применена, соответствующий классификатор называется в результатах LSTM-diff.
- (b) Расширение обучающей выборки путём добавления к ней фрагментов ряда ширины w , к значениям которых прибавлена константа порядка среднего значения ряда (одновременно такая же константа прибавляется к целевому значению). Такое преобразование призвано улучшать качество предсказания в ситуации, когда LSTM переобучается предсказывать ряд в диапазоне

значений, который встречается в обучающей выборке, строит нестабильный прогноз вне него. Если эвристика была применена, соответствующий классификатор называется в результатах LSTM-trend.

С учётом вышесказанного, перебираются следующие структурные параметры:

- (a) Эвристика. Для модели f суперпозиции (1.1) перебирается целесообразность использования LSTM без модификаций, с применением эвристики LSTM-diff или LSTM-trend. Для моделей прогнозирования остатков перебираются только LSTM без модификаций и LSTM-diff, поскольку предполагается стационарность ряда остатков
- (b) Число нейронов LSTM-слоя k фиксировано и равно 20 во всех экспериментах.
- (c) Длина истории w перебирается по сетке: 10, 20, 40.

Число эпох обучения фиксировано и равно 60, кроме экспериментов с эвристикой LSTM-trend, в которой оно равно 120.

4.3 Результаты экспериментов

Для каждого из трёх временных рядов (airline, synthetic, german) приводятся:

1. Общий вид временного ряда и разбиение на обучающую и проверочную части.
2. Прогнозы всех основных моделей с оптимальными структурными параметрами. Размер блока во всех экспериментах равен 10. Все прогнозы строятся с помощью алгоритма 2.4.
3. Прогноз лучшей суперпозиции и основной модели этой суперпозиции.
4. Матрица качества прогноза суперпозиций (в логарифмическом масштабе). По строкам отложены основные модели, по столбцам – модели прогнозирования; первый столбец (None) соответствует основной модели, без суперпозиции. Ячейки, обведённые зелёными рамками, соответствуют наилучшим прогнозам для данной основной модели (то есть минимальному значению в строке). Глобальный минимум обозначен более широкой рамкой.

4.3.1 Результаты эксперимента: airline

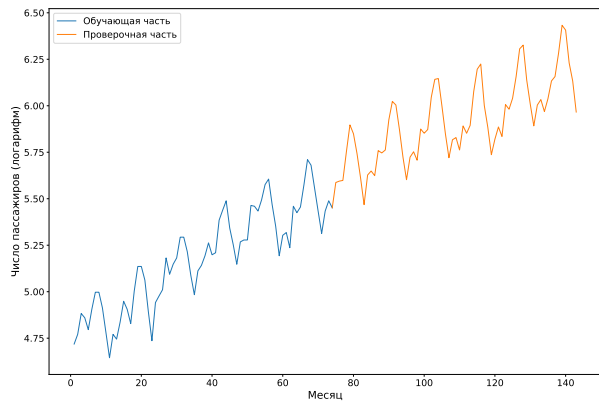


Рис. 1: Временной ряд airline

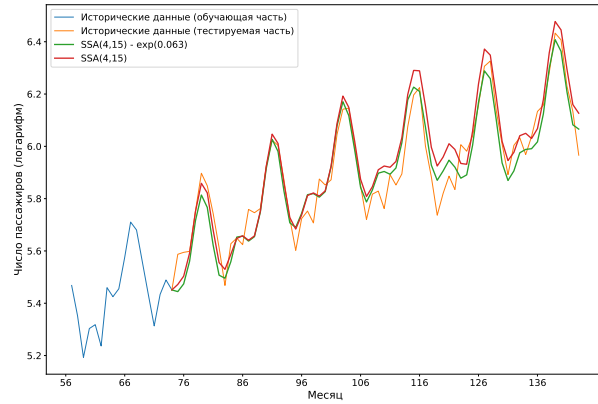


Рис. 2: Прогноз лучшей суперпозиции против прогноза основной модели

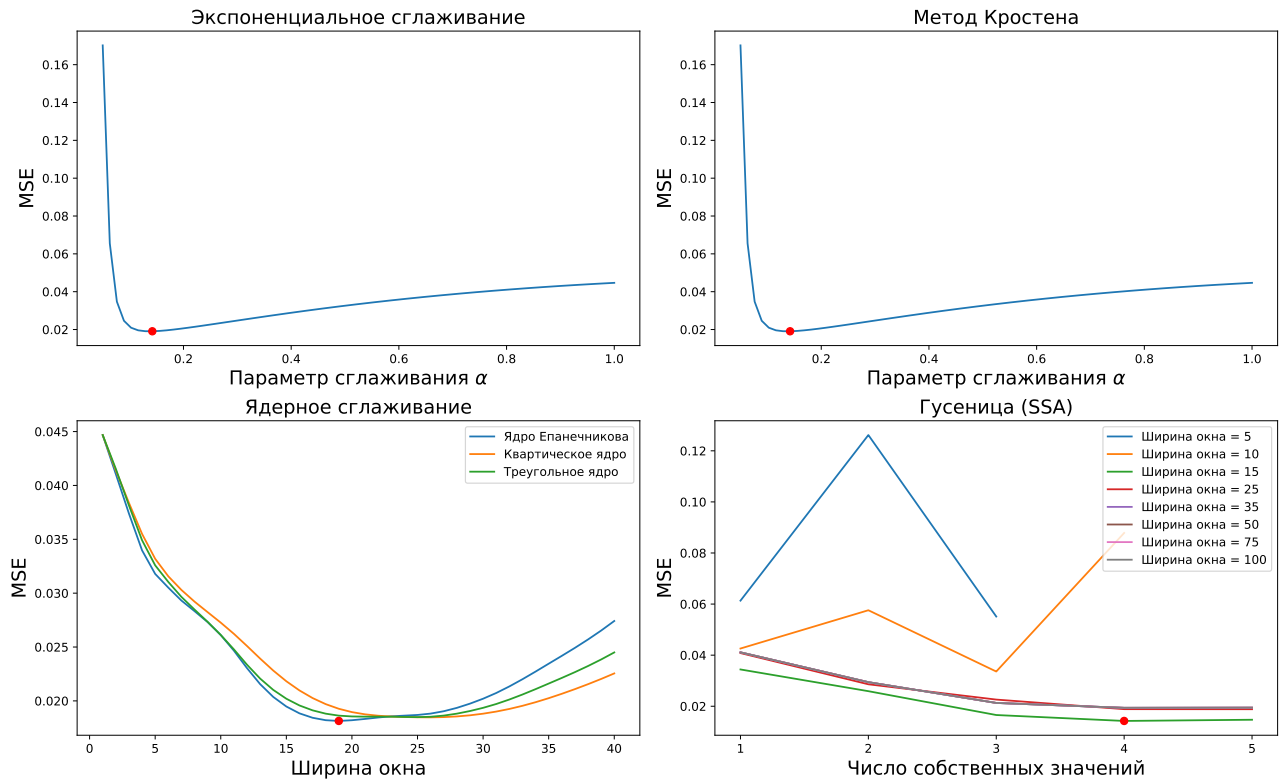


Рис. 3: Зависимость качества прогноза от структурных параметров

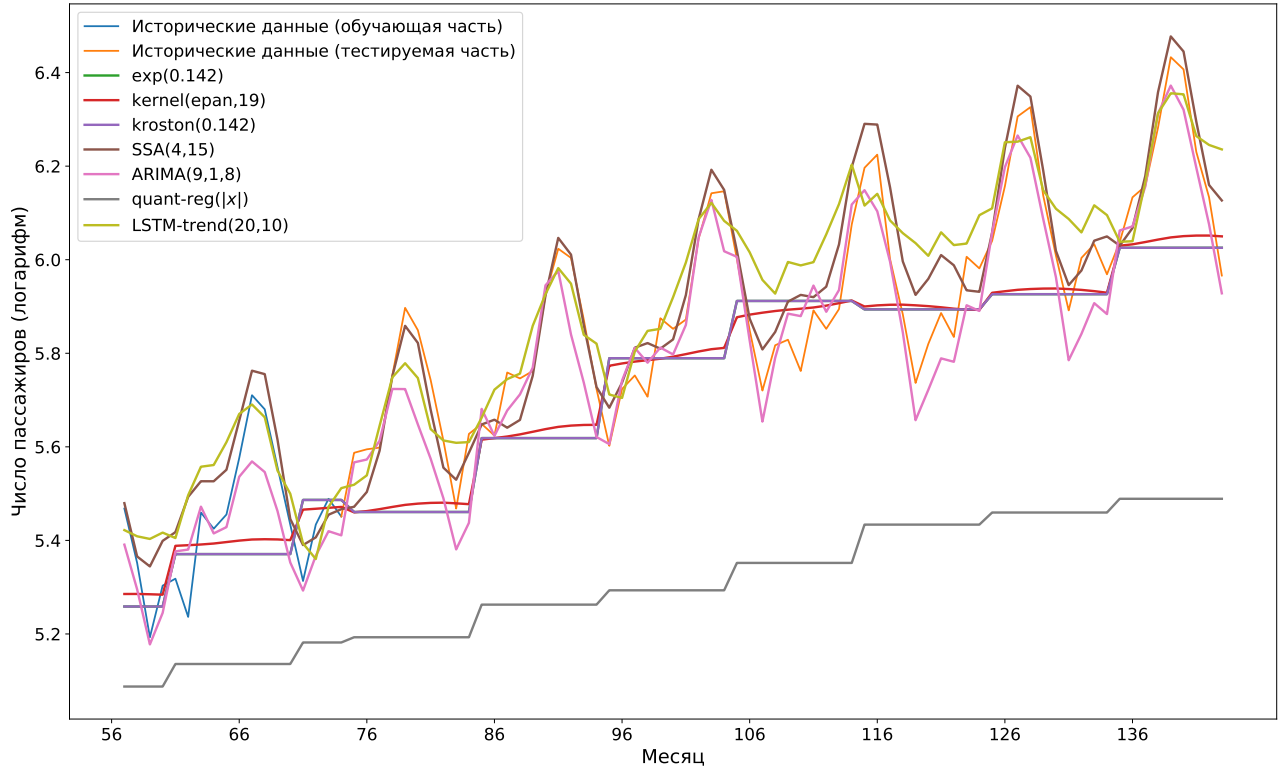


Рис. 4: Прогнозы основных моделей

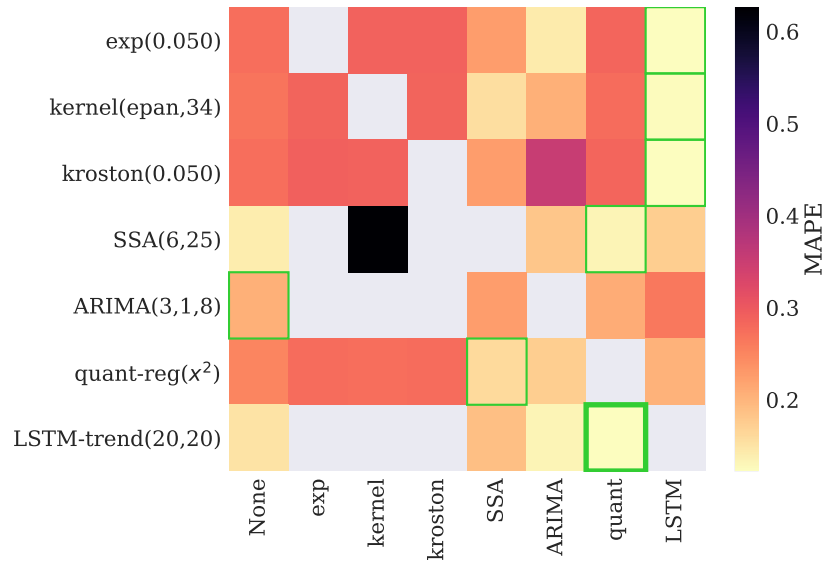


Рис. 5: Матрица качества прогноза суперпозиций (log MSE)

	None	exp	kernel	kroston	SSA	ARIMA	quant	LSTM
exp(0.142)	-3.197	0.000	-3.443	-3.457	-3.855	-3.992	-3.595	-4.487
kernel(epan,19)	-3.305	-3.565	0.000	-3.544	-4.078	-4.557	-3.690	-4.485
kroston(0.142)	-3.197	-3.382	-3.443	0.000	-3.940	-4.643	-3.594	-4.499
SSA(4,15)	-5.146	-5.336	-5.259	-4.832	0.000	-5.204	-5.331	-5.307
ARIMA(9,1,8)	-5.013	-4.937	-4.991	-4.870	-4.053	0.000	-5.125	-4.607
quant-reg(x)	-1.061	-2.880	-3.465	-2.880	-1.976	-4.057	0.000	-3.121
LSTM-trend(20,10)	-4.323	-4.620	-4.615	-4.694	-4.657	-4.311	-4.687	0.000

Рис. 6: Таблица качества прогноза суперпозиций (log MSE)

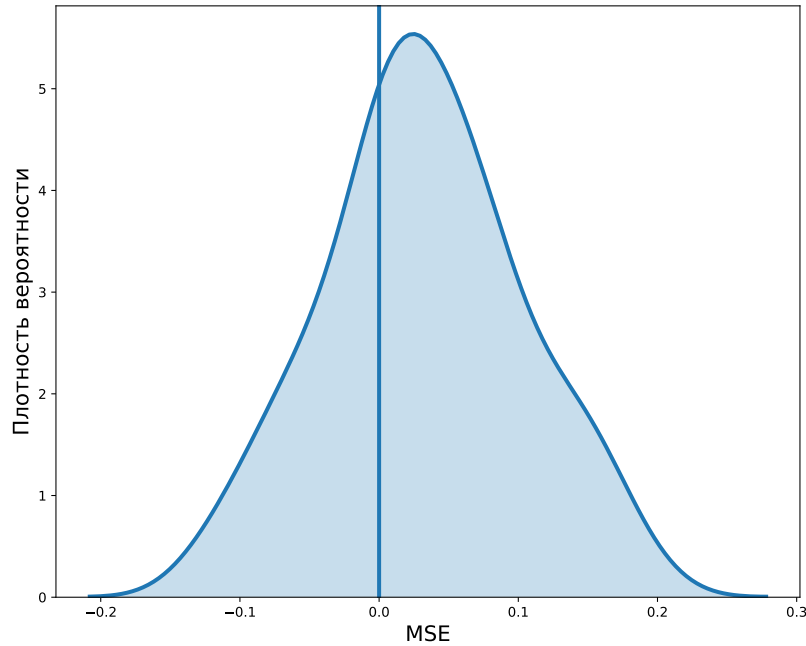


Рис. 7: Эмпирическая функция распределения ошибок лучшей основной модели (SSA)

На рис. 7 видно, что распределение ошибок лучшей основной модели асимметрично. Для такого распределения ошибок квантильная регрессия ожидаемо даёт один из лучших результатов.

4.3.2 Результаты эксперимента: synthetic

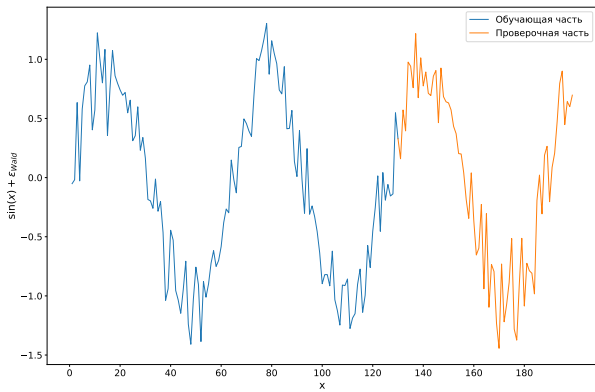


Рис. 8: Временной ряд synthetic

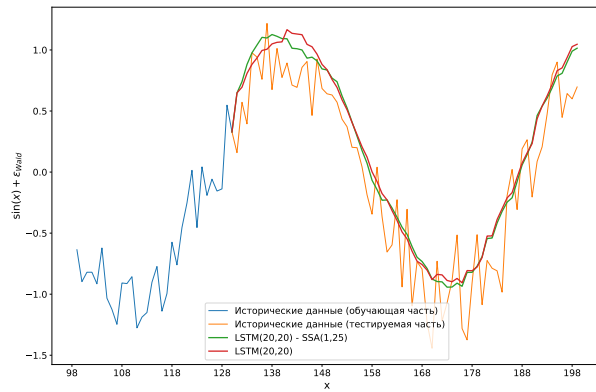


Рис. 9: Прогноз лучшей суперпозиции против прогноза основной модели

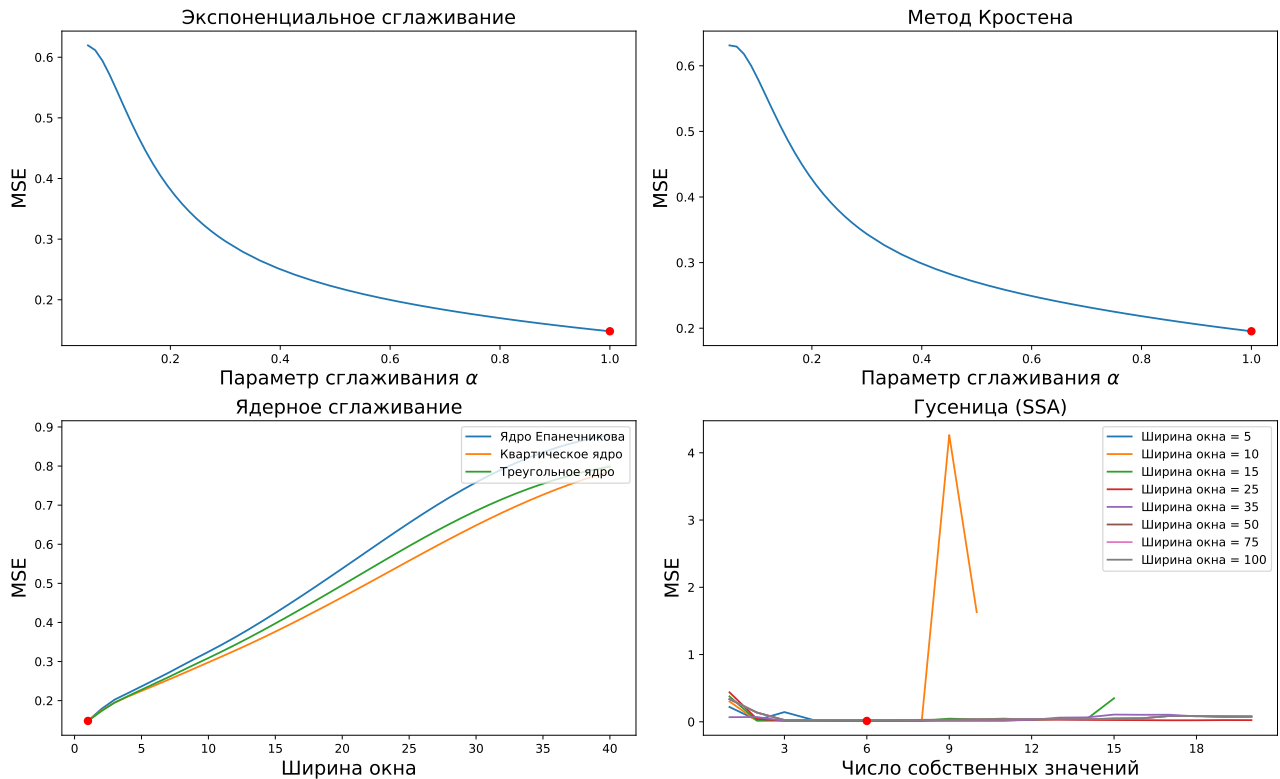


Рис. 10: Зависимость качества прогноза от структурных параметров

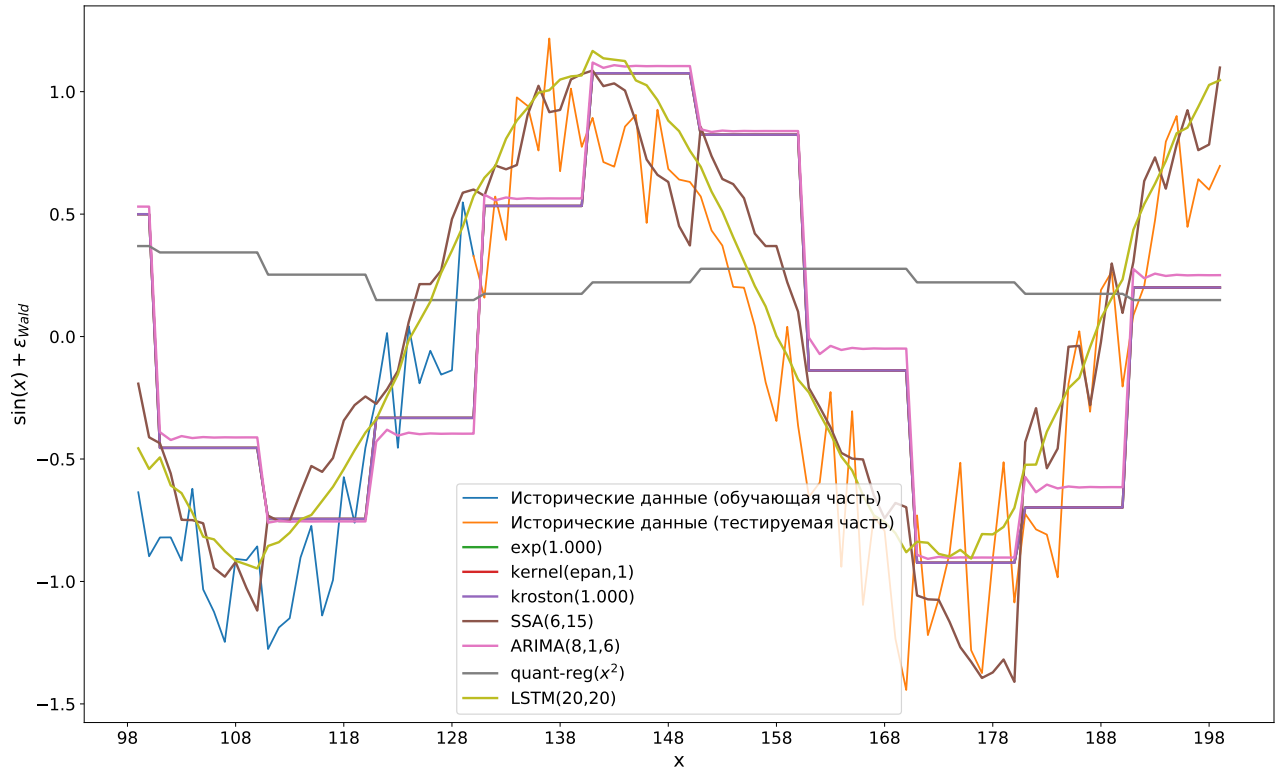


Рис. 11: Прогнозы основных моделей

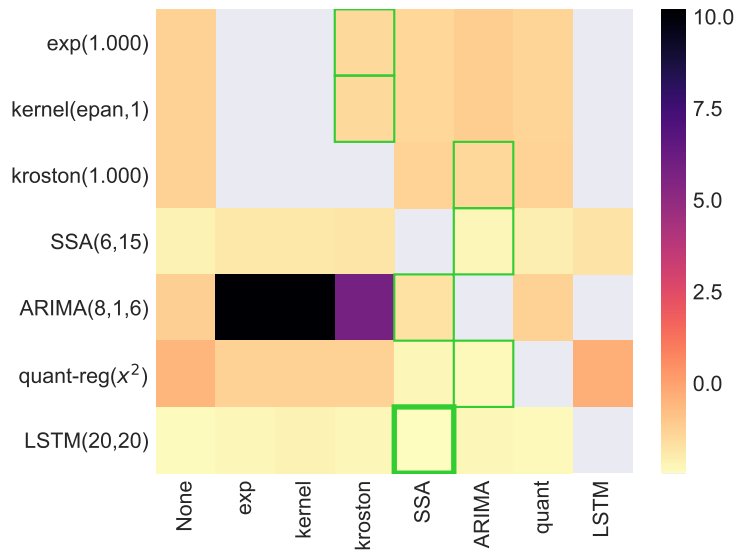


Рис. 12: Матрица качества прогноза суперпозиций (log MSE)

	None	exp	kernel	kroston	SSA	ARIMA	quant	LSTM
exp(1.000)	-1.447	0.000	24.058	-1.915	-1.887	-1.543	-1.498	20.476
kernel(epan,1)	-1.447	24.058	0.000	-1.915	-1.887	-1.532	-1.498	20.477
kroston(1.000)	-1.447	23.253	23.253	0.000	-1.653	-1.838	-1.389	21.906
SSA(6,15)	-2.566	-2.432	-2.315	-2.194	0.000	-2.597	-2.799	-2.032
ARIMA(8,1,6)	-1.492	10.201	10.201	5.886	-2.263	0.000	-1.510	16.610
quant-reg(x^2)	-0.655	-1.434	-1.434	-1.451	-2.894	-3.920	0.000	-0.754
LSTM(20,20)	-4.195	-4.123	-3.872	-4.322	-4.074	-4.240	-4.284	0.000

Рис. 13: Таблица качества прогноза суперпозиций (log MSE)

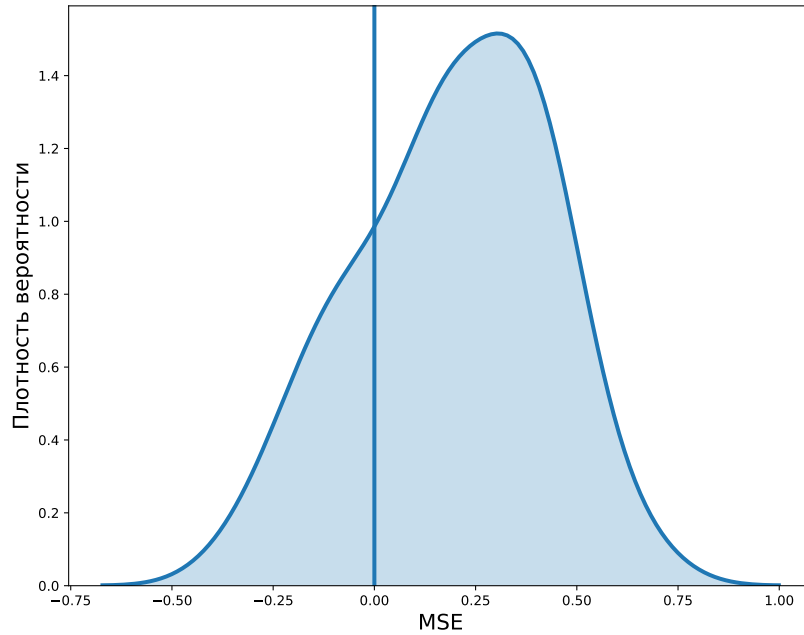


Рис. 14: Эмпирическая функция распределения ошибок лучшей основной модели (LSTM)

На рис. 14 видно, что распределение ошибок лучшей основной модели асимметрично. Для такого распределения ошибок квантильная регрессия ожидаемо даёт один из лучших результатов.

4.3.3 Результаты эксперимента: german

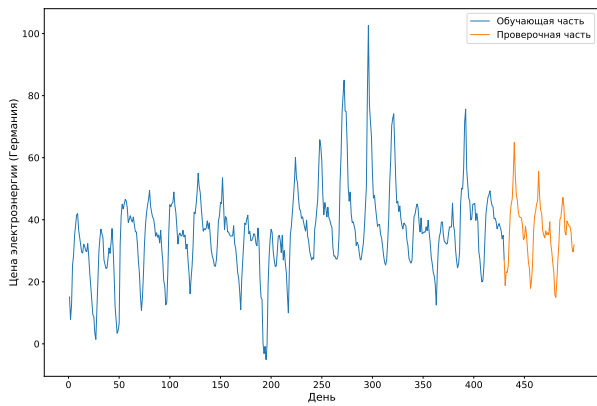


Рис. 15: Временной ряд герман

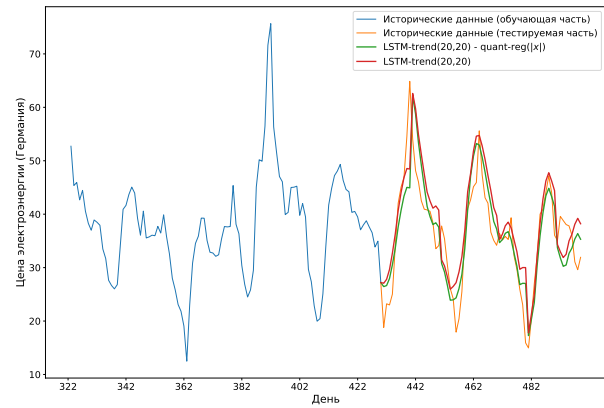


Рис. 16: Прогноз лучшей суперпозиции против прогноза основной модели

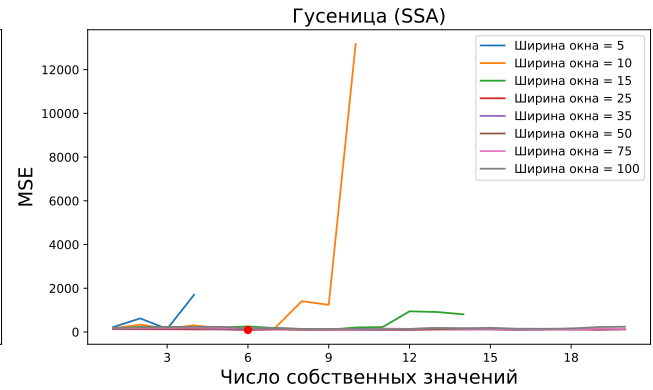
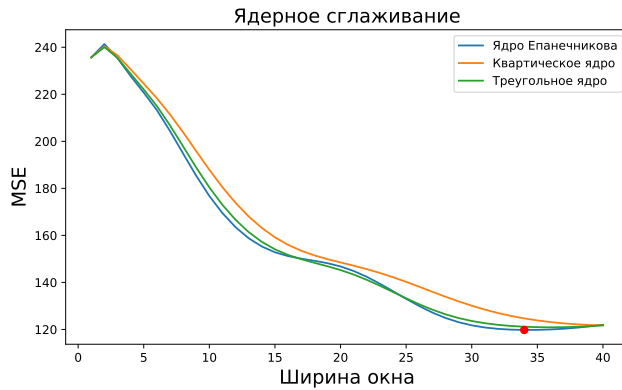
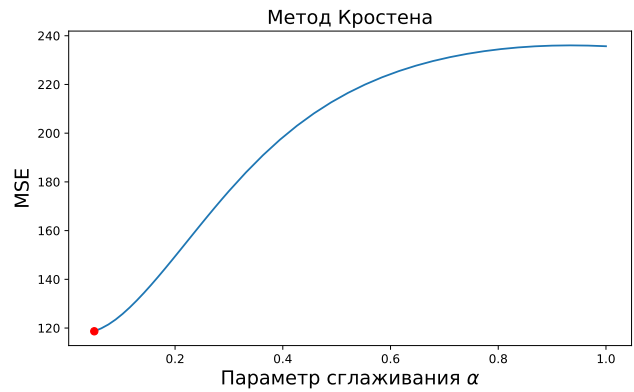
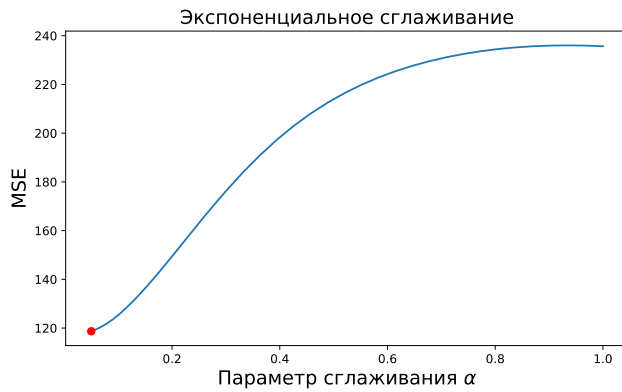


Рис. 17: Зависимость качества прогноза от структурных параметров

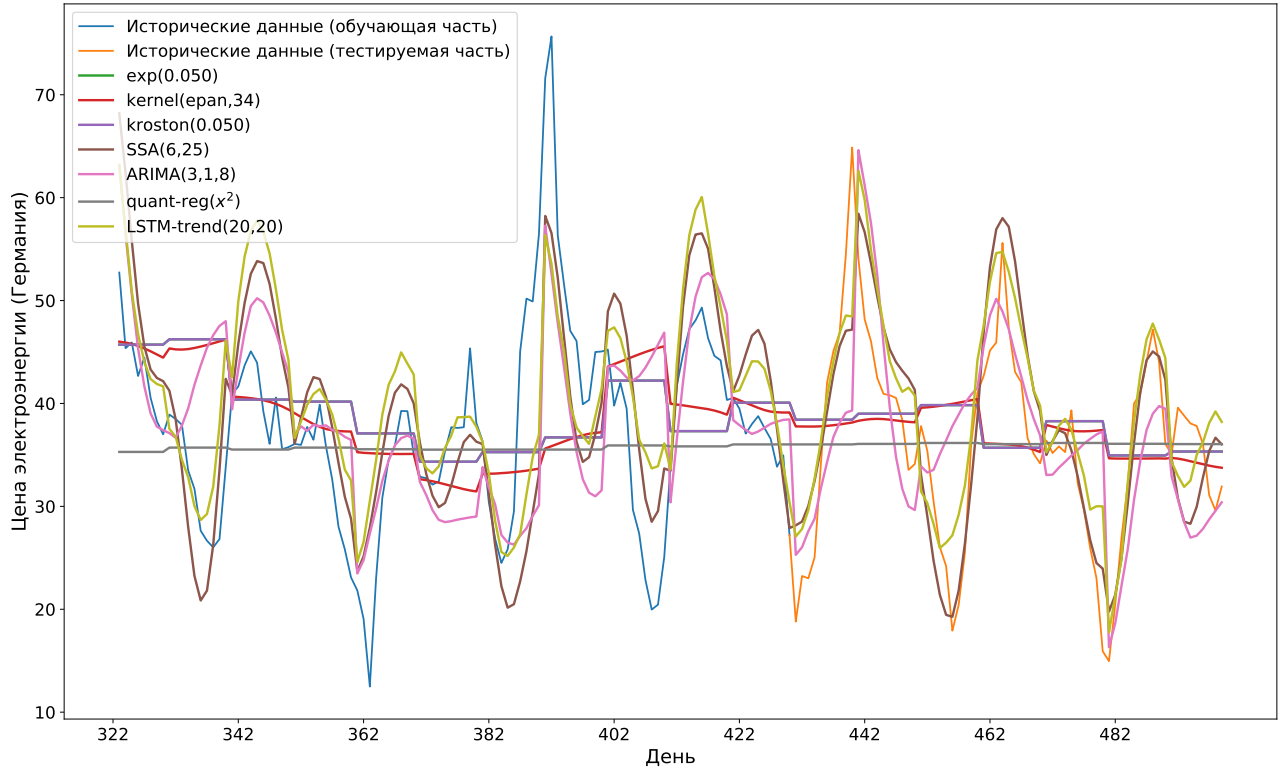


Рис. 18: Прогнозы основных моделей

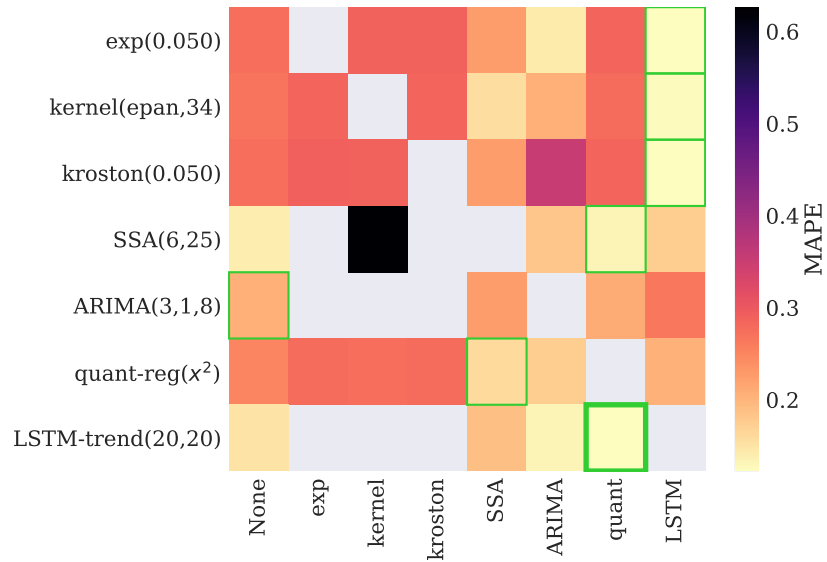


Рис. 19: Матрица качества прогноза суперпозиций (log MSE)

	None	exp	kernel	kroston	SSA	ARIMA	quant	LSTM
exp(1.000)	-1.447	0.000	24.058	-1.915	-1.887	-1.543	-1.498	20.476
kernel(epan,1)	-1.447	24.058	0.000	-1.915	-1.887	-1.532	-1.498	20.477
kroston(1.000)	-1.447	23.253	23.253	0.000	-1.653	-1.838	-1.389	21.906
SSA(6,15)	-2.566	-2.432	-2.315	-2.194	0.000	-2.597	-2.799	-2.032
ARIMA(8,1,6)	-1.492	10.201	10.201	5.886	-2.263	0.000	-1.510	16.610
quant-reg(x^2)	-0.655	-1.434	-1.434	-1.451	-2.894	-3.920	0.000	-0.754
LSTM(20,20)	-4.195	-4.123	-3.872	-4.322	-4.074	-4.240	-4.284	0.000

Рис. 20: Таблица качества прогноза суперпозиций (log MSE)

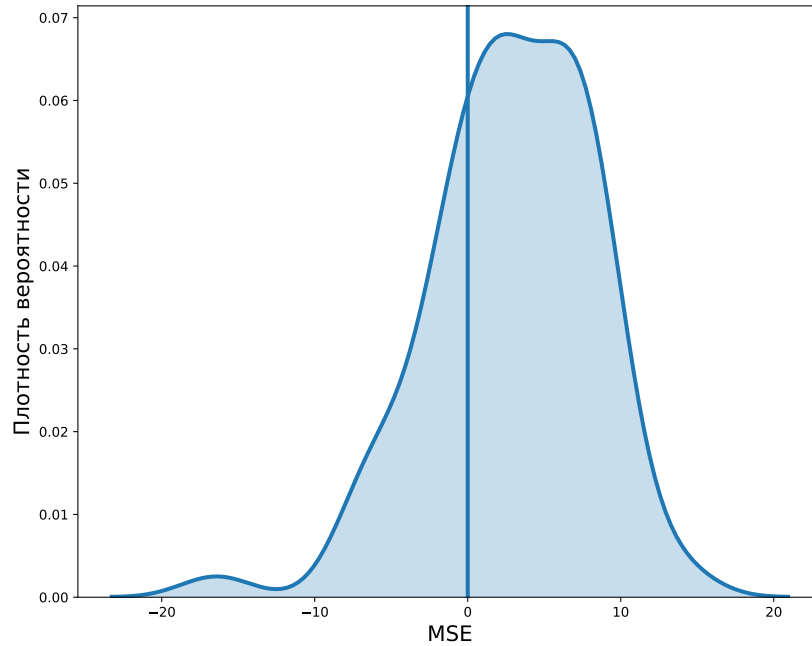


Рис. 21: Эмпирическая функция распределения ошибок лучшей основной модели (LSTM)

На рис. 21 видно, что распределение ошибок лучшей основной модели асимметрично. Для такого распределения ошибок квантильная регрессия ожидаемо даёт один из лучших результатов.

4.3.4 Результаты эксперимента: railroads

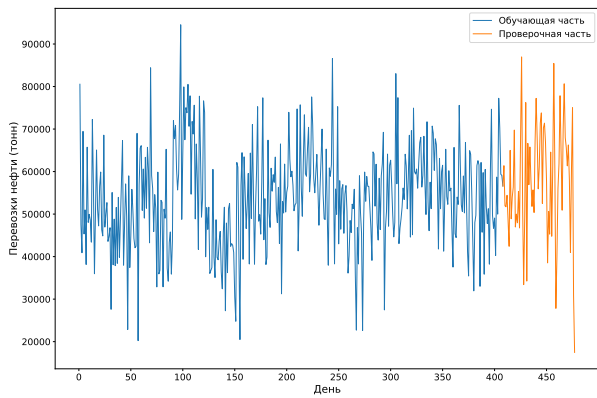


Рис. 22: Временной ряд railroads

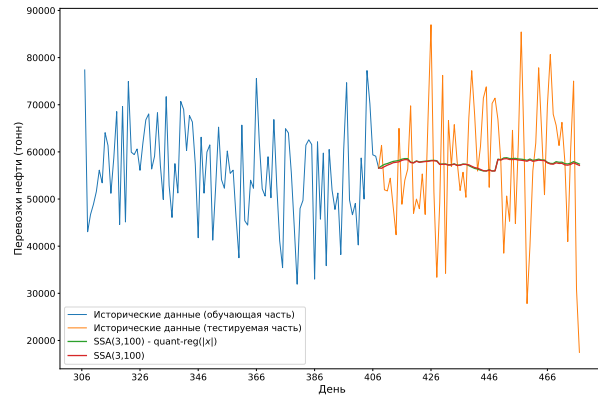


Рис. 23: Прогноз лучшей суперпозиции против прогноза основной модели

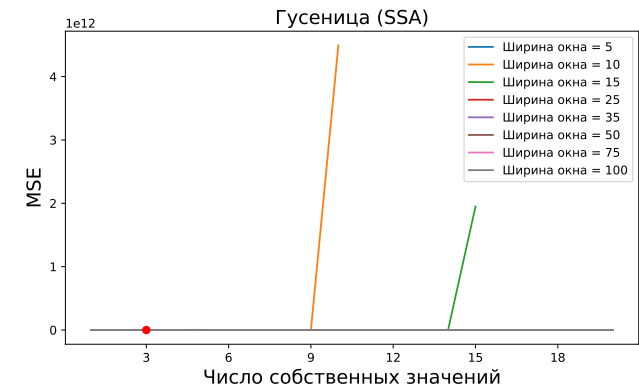
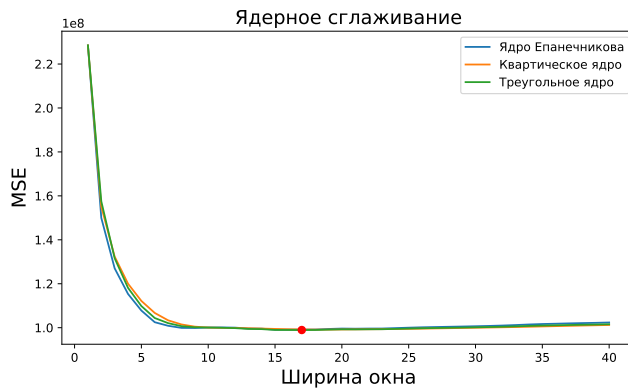
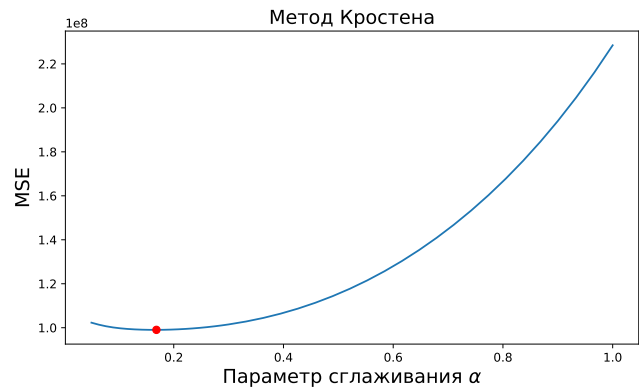
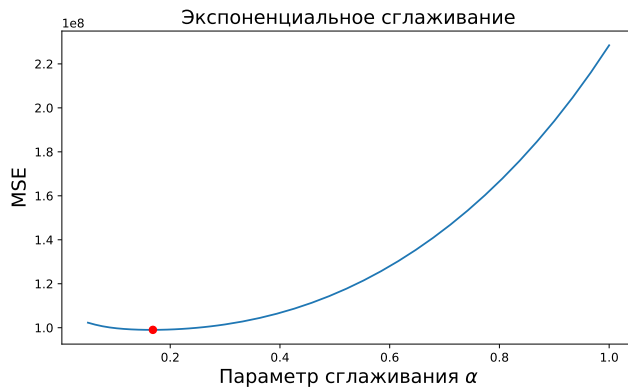


Рис. 24: Зависимость качества прогноза от структурных параметров

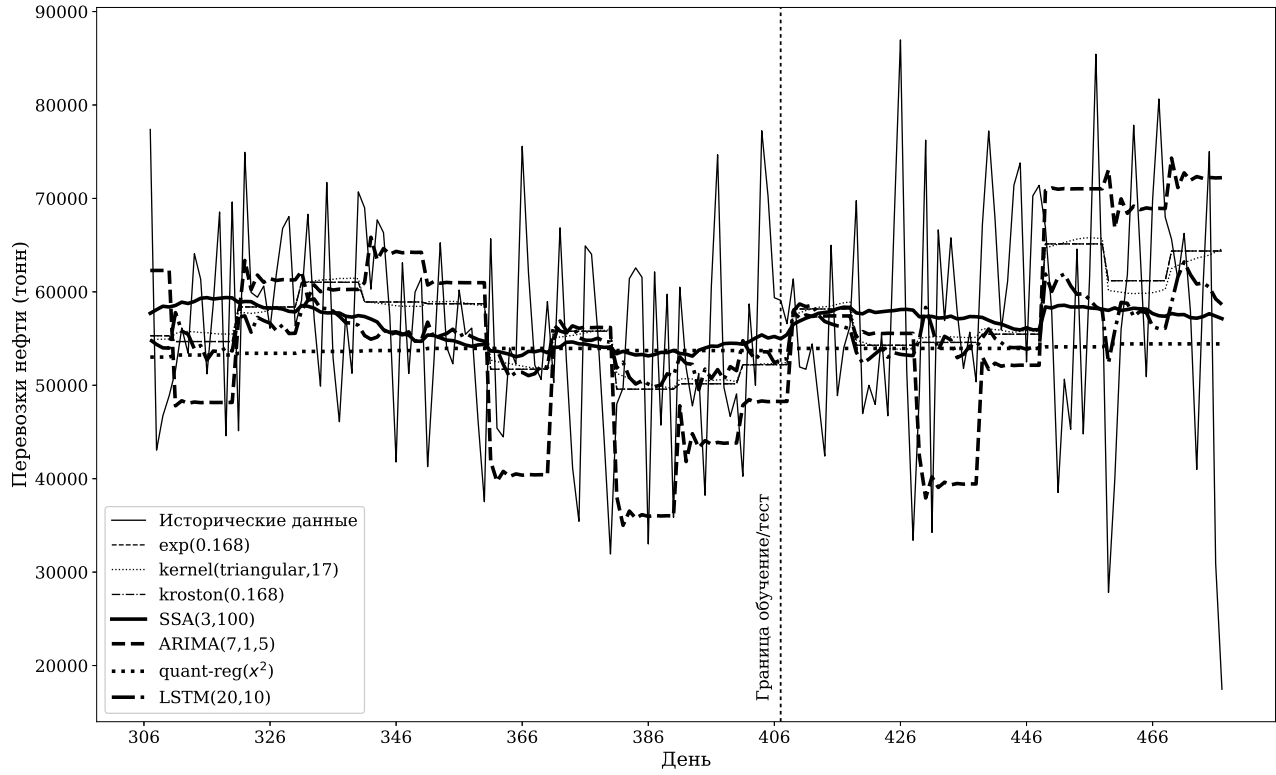


Рис. 25: Прогнозы основных моделей

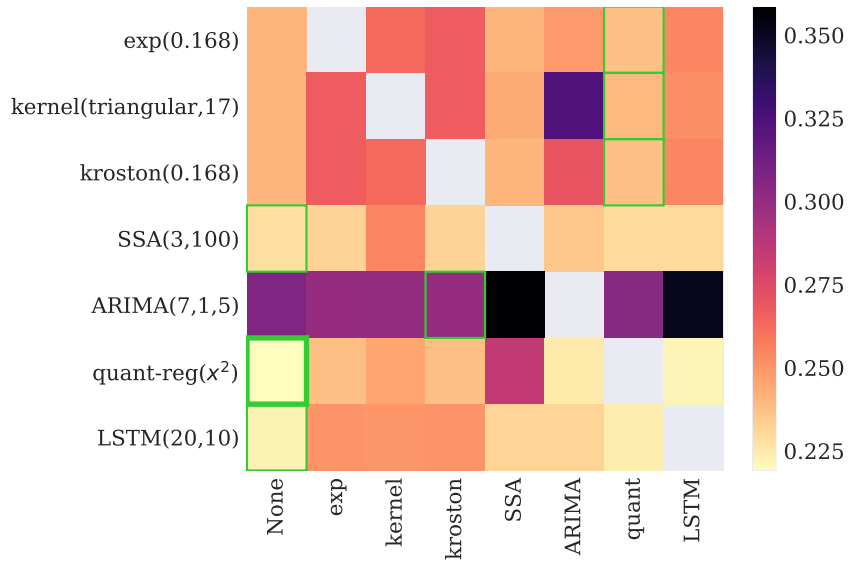


Рис. 26: Матрица качества прогноза суперпозиций (log MSE)

	None	exp	kernel	kroston	SSA	ARIMA	quant	LSTM
exp(0.168)	19.167	0.000	19.304	19.344	19.161	19.236	19.169	19.269
kernel(triangular,17)	19.162	19.350	0.000	19.350	19.159	19.749	19.167	19.260
kroston(0.168)	19.167	19.344	19.304	0.000	19.161	19.345	19.166	19.269
SSA(3,100)	19.038	19.095	19.257	19.095	0.000	19.126	19.044	19.092
ARIMA(7,1,5)	19.616	19.600	19.588	19.600	20.015	0.000	19.769	19.943
quant-reg(x^2)	19.084	19.149	19.197	19.149	19.462	19.100	0.000	19.069
LSTM(20,10)	19.076	19.223	19.211	19.223	19.090	19.093	19.077	0.000

Рис. 27: Таблица качества прогноза суперпозиций (log MSE)

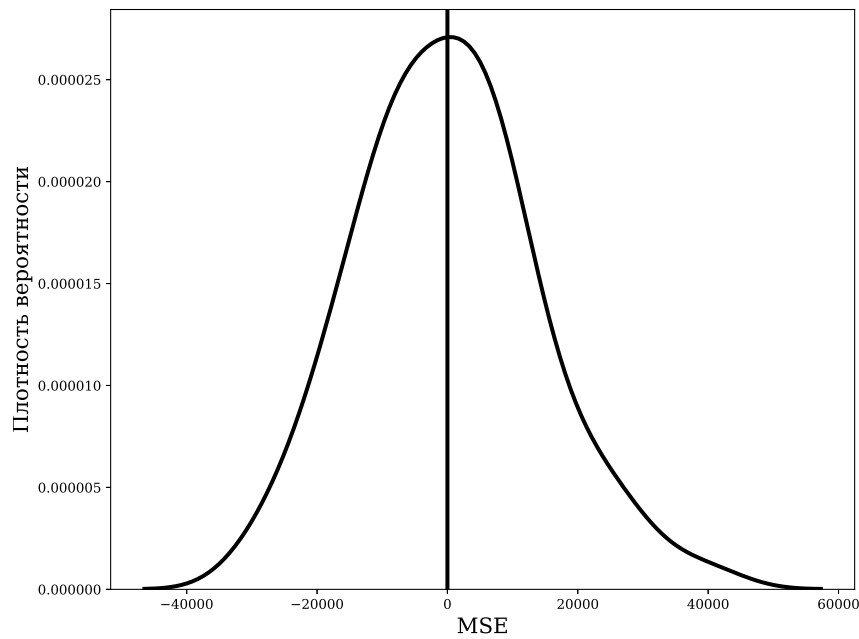


Рис. 28: Эмпирическая функция распределения ошибок лучшей основной модели (LSTM)

На рис. 28 видно, что распределение ошибок близко к нормальному, поэтому использование суперпозиции не улучшает качество прогноза.

4.3.5 Горизонт прогнозирования в экспериментах

Для определения горизонта прогнозирования (1.2) построим зависимость ошибки MAPE (4.1), а также её выборочного стандартного отклонения (представлена на графике в виде доверительной верхней границы), от запроса прогнозирования в проведённых экспериментах. Горизонт прогнозирования будем определять графически по правилу сломанной трости (рис. 29).

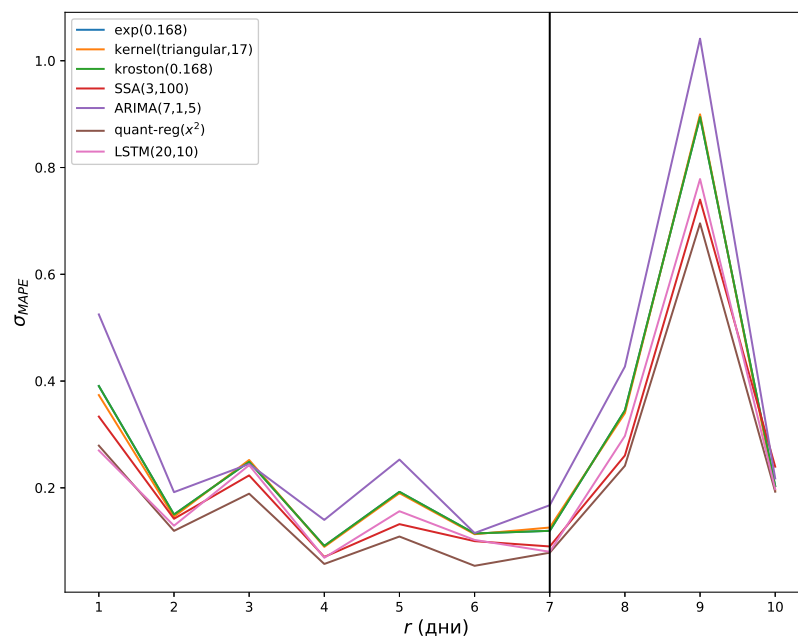


Рис. 29: Правило сломанной трости: резкий рост выборочного стандартного отклонения ошибки свидетельствует о достижении горизонта прогнозирования

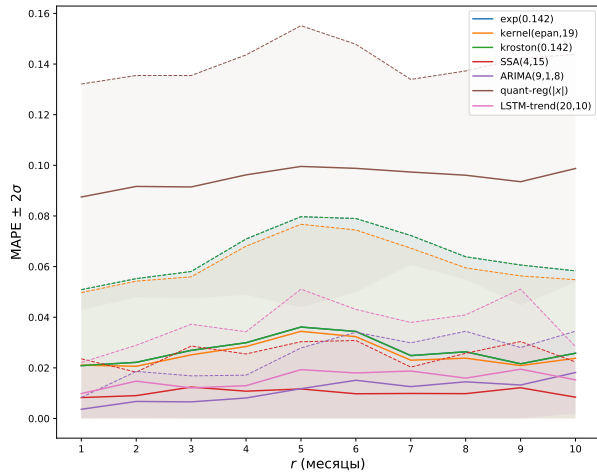


Рис. 30: Временной ряд airline

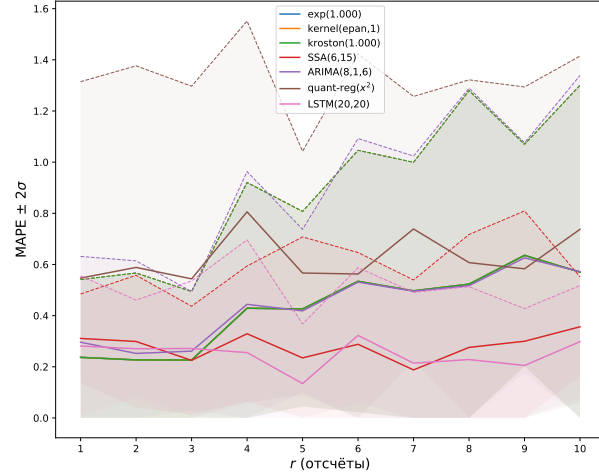


Рис. 31: Временной ряд synthetic

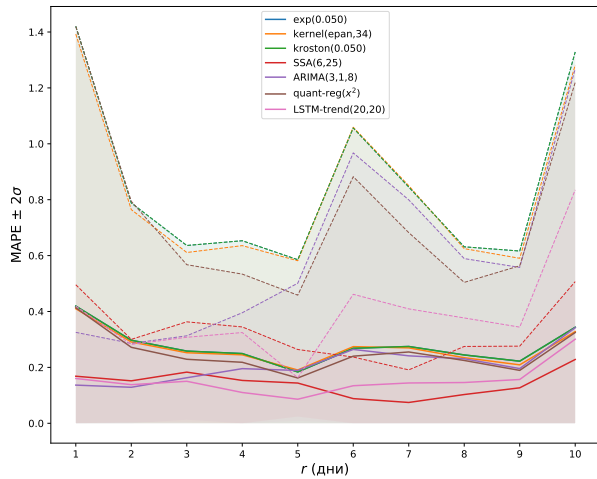


Рис. 32: Временной ряд герман

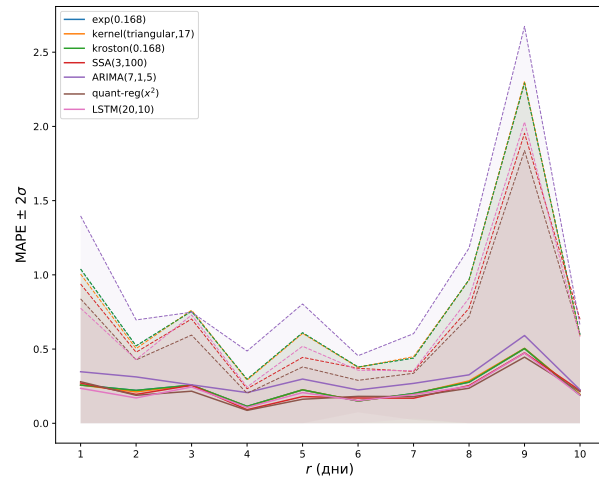


Рис. 33: Временной ряд railroads

Зависимость ошибки MAPE (сплошные линии) и её 95%-доверительной верхней границы MAPE+2σ (пунктир) от запроса прогнозирования

Видно, что на датасетах *airline* и *synthetic* верхний доверительный интервал и дисперсия ошибки практически не возрастают с увеличением запроса прогнозирования (на *airline* – для всех методов, на *synthetic* – для методов, которым удалось восстановить синусоидальный сигнал, таким как LSTM и SSA). Горизонт прогнозирования для этих рядов значительно превышает использованную длину блока 10.

На датасетах *german* и *railroads* наблюдается другой характер зависимости: существует величина запроса, на которой дисперсия ошибки значительно возрастает. Наиболее наглядно горизонт прогнозирования, равный 8, виден на графике ошибки для датасета *railroads*. Для датасета *german* видно, что на величине запроса 6 возрастает дисперсия ошибки для всех методов, кроме SSA.

В целом, использованная длина блока либо не превышает (airline, synthetic), либо незначительно превышает (railroads, german) горизонт прогнозирования.

4.4 Эксперимент: зависимость качества от асимметричности распределения

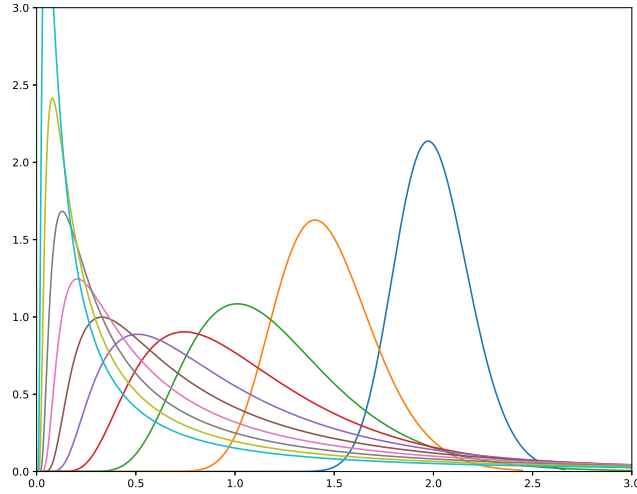


Рис. 34: Плотности вероятности распределений из используемого однопараметрического семейства

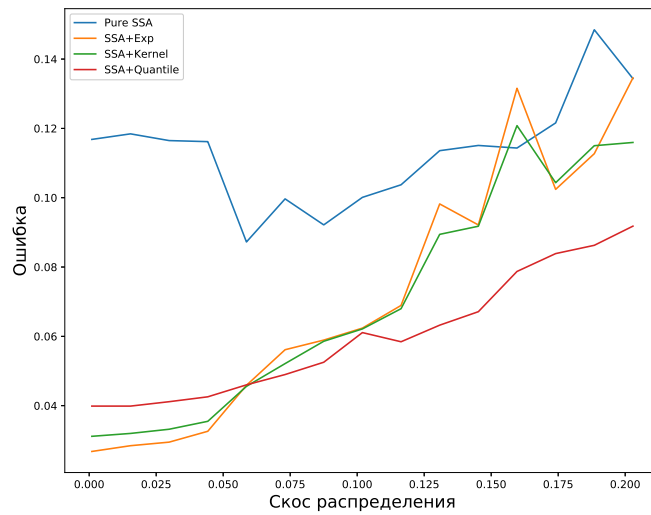


Рис. 35: Зависимость качества прогноза от “скоса” распределения

Для распределения шума временного ряда используется однопараметрическое подмножество распределений Вальда, задаваемое параметром l , по которому параметры распределения Вальда определяются как $\mu = 1 + \frac{l}{30}$, $\lambda = \frac{l^2}{4}$. Для наглядной интерпретации, в качестве меры скоса распределения введём

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

где $x_{н.в.}$ — наиболее вероятное значение случайной величины, соответствующее максимуму плотности вероятности и положению пика на рис. 35. Очевидно, $S \in [0, \frac{1}{2})$, причём значение 0 соответствует распределению, вероятностная масса которого расположена симметрично от наиболее вероятного значения, например, нормальному, а значения, близкие к $\frac{1}{2}$, соответствуют сильно скошенным распределениям. Использована кусочно-линейная асимметричная функция ошибки

$$l(x) = \begin{cases} x, & x > 0 \\ -\frac{1}{5}x, & x < 0 \end{cases}. \quad (4.3)$$

Заключение

Рассмотрена задача построения суперпозиции моделей при прогнозировании временных рядов. В качестве базовых алгоритмов рассмотрены экспоненциальное сглаживание, ядерное сглаживание, метод Кростена, SSA, ARIMA, квантильная регрессия и LSTM. Для рассматриваемых выборок: железнодорожных грузовых перевозок, пассажирских авиаперевозок, цен на электроэнергию и синтетической — построены матрица качества суперпозиции и распределение остатков модели, а также определён горизонт прогнозирования. Вычислительный эксперимент показал, что использование суперпозиции может повысить качество прогноза относительно качества базовых моделей. Отдельно исследована зависимость ошибки от степени асимметричности распределения шумовых остатков и показано, что с ростом асимметричности суперпозиции достигают меньшей ошибки по сравнению с базовыми моделями. Исходный код эксперимента размещён в [17].

Список литературы

- [1] Буряк Е.В., Кульпина В.П., Голяшев А.В., Лобанова А.А. Динамика грузоперевозок в России // Бюллетень социально-экономического кризиса в России. 2015. 1. № 4. С. 12-15.
- [2] Коришева О. В. Управление экономической устойчивостью транспортных компаний в сфере грузовых железнодорожных перевозок : дис. – М. : Коришева Ольга Викторовна, 2014.

- [3] Lijuan L., Liu H., Wu J. et al. A novel model for wind power forecasting based on Markov residual correction // Renewable Energy Congress. 2015. N 6. P. 1-5.
- [4] Cohen D. A., Lys T. Z. A note on analysts' earnings forecast errors distribution // Journal of Accounting and Economics. 2003. **36**. N 1-3. P. 147-164.
- [5] Kan M. L., Lee Y. B., Chen W. C. Apply grey prediction in the number of Tourist // Genetic and Evolutionary Computing. 2010. N 4. P. 481-484.
- [6] 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. **86**. № 2. С. 186–188.
- [7] Guha B., Bandyopadhyay G. Gold price forecasting using ARIMA model // Journal of Advanced Management Science Vol. 2016. **4**. N 2. P. 117-121.
- [8] Голяндина Н. Э. Метод “Тусеница”-SSA: прогноз временных рядов. М.: Издательский центр “Академия”, 2004.
- [9] Laptev N., Yosinski J., Li L. E. et al. Time-series extreme event forecasting with neural networks at Uber // International Conference on Machine Learning. 2017. N 34. P. 1-5.
- [10] Kalekar P. S. Time series forecasting using holt-winters exponential smoothing // Kanwal Rekhi School of Information Technology. 2004. P. 1-13.
- [11] Ralaivola L., D’Alch-Buc F. Time series filtering, smoothing and learning using the kernel Kalman filter // Neural Networks, proceedings of IEEE International Joint Conference. 2005. **3**. P. 1449-1454.
- [12] Временные ряды грузоперевозок. URL: <http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/data> (дата обращения: 27.03.2018).
- [13] Уваров Н.Д., Кузнецов М.П., Малькова А.С., Рудаков К.В., Стрижов В.В. Приложение к работе “Выбор суперпозиции моделей при прогнозировании грузовых железнодорожных перевозок”. 2018. URL: <http://svn.code.sf.net/>

p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/doc/addition.pdf (дата обращения: 27.03.2018).

- [14] Box G. E., Jenkins G. M., Reinsel G. C. et al. Time series analysis, forecasting and control. New York: John Wiley & Sons, 1976.
- [15] Цена на электричество в Германии. URL: http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/data/electricity_price_german/GermanSpotPrice.csv (дата обращения: 27.03.2018).
- [16] Потребление электроэнергии (Польша). URL: <http://gdudek.el.pcz.pl/files/PL.xls> (дата обращения: 27.03.2018).
- [17] Исходный код эксперимента к работе “Выбор суперпозиции моделей при прогнозировании грузовых железнодорожных перевозок” (модуль на Python и Jupyter-ноутбук). URL: <http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/code/> (дата обращения: 27.03.2018).