

**МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)**

Факультет управления и прикладной математики

Институт кибернетики имени В.М. Глушкова НАН Украины

Кафедра теоретической кибернетики и методов оптимального управления

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**“Построение адаптивных композиций алгоритмов
прогнозирования при несимметричной функции потерь”**

Заведующий кафедрой: академик НАНУ Сергиенко И.В.
Научный руководитель: академик НАНУ Сергиенко И.В.,
Научный консультант: н.с. ВЦ РАН, к.ф.-м.н. Воронцов К.В.
Рецензент: к.ф.-м.н. Китов В.В.
Студент: Коваль А.С.

Киев 2009

Содержание

1	ВВЕДЕНИЕ.....	3
1.1	Задача ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ	4
1.2	Задача ПРОГНОЗИРОВАНИЯ ПОТРЕБИТЕЛЬСКОГО СПРОСА.....	4
1.3	НЕСИММЕТРИЧНАЯ ФУНКЦИЯ ПОТЕРЬ.....	5
1.4	О НЕОБХОДИМОСТИ ПОСТРОЕНИЯ АЛГОРИТМИЧЕСКИХ КОМПОЗИЦИЙ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ПОТРЕБИТЕЛЬСКОГО СПРОСА	7
1.5	ПОСТАНОВКА ЗАДАЧИ	8
2	МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ПРОГНОЗИРОВАНИЯ.....	9
2.1	СИММЕТРИЧНАЯ ФУНКЦИЯ ПОТЕРЬ	9
2.1.1	АВТОРЕГРЕССИОННЫЕ МОДЕЛИ И МОДЕЛИ СКОЛЬЗЯЩЕГО СРЕДНЕГО	9
2.1.2	МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ.....	10
2.1.3	АДАПТИВНЫЕ МОДЕЛИ.....	13
2.1.4	НЕЙРОННЫЕ СЕТИ	15
2.2	НЕСИММЕТРИЧНАЯ ФУНКЦИЯ ПОТЕРЬ.....	19
2.2.1	КВАНТИЛЬНАЯ РЕГРЕССИЯ.....	19
2.2.2	ПОСТРОЕНИЕ И ОЦЕНИВАНИЕ ПРОГНОЗА ПЛОТНОСТИ	21
2.3	КОМПОЗИЦИИ АЛГОРИТМОВ ПРИ СИММЕТРИЧНОЙ ФУНКЦИИ ПОТЕРЬ	25
2.3.1	ПРОСТОЕ УСРЕДНЕНИЕ	26
2.3.2	ВЫБОР НАИЛУЧШЕГО АЛГОРИТМА СРЕДИ БАЗОВЫХ.....	26
2.3.3	МЕТОД НАИМЕНЬШИХ КВАДРАТОВ.....	26
2.3.4	МЕТОД НАИМЕНЬШИХ КВАДРАТОВ С РЕГУЛЯРИЗАЦИЕЙ	27
2.3.5	МЕТОД ЛОКАЛЬНОЙ АДАПТАЦИИ ВЕСОВ С РЕГУЛЯРИЗАЦИЕЙ.....	28
2.4	КОМПОЗИЦИИ АЛГОРИТМОВ ПРИ НЕСИММЕТРИЧНОЙ ФУНКЦИИ ПОТЕРЬ.....	29
2.4.1	МЕТОД НАИМЕНЬШИХ МОДУЛЕЙ	30
2.4.2	МЕТОД ЛОКАЛЬНОЙ АДАПТАЦИИ ВЕСОВ С РЕГУЛЯРИЗАЦИЕЙ.....	31
2.5	РАЗЛИЧНЫЕ ВАРИАНТЫ ПОСТАНОВКИ ЗАДАЧИ	33
3	ОБ ОДНОМ ЭФФЕКТИВНОМ МЕТОДЕ ПОСТРОЕНИЯ ПРОГНОЗА ПЛОТНОСТИ ...	35
3.1	ПОСТРОЕНИЕ ПРОГНОЗА ПЛОТНОСТИ	35
3.2	ЭКСПЕРИМЕНТ.....	36
3.3	РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ.....	37
4	СРАВНЕНИЕ МЕТОДА КВАНТИЛЬНОЙ РЕГРЕССИИ И МЕТОДА ПРОГНОЗИРОВАНИЯ ПЛОТНОСТИ	39
4.1	ВЫЧИСЛЕНИЕ ПРОГНОЗОВ ДЛЯ АВТОРЕГРЕССИОННОЙ МОДЕЛИ БЕЗ УЧЕТА ВЕСОВЫХ КОЭФФИЦИЕНТОВ.....	39
4.2	ВЫЧИСЛЕНИЕ ПРОГНОЗОВ ДЛЯ АВТОРЕГРЕССИОННОЙ МОДЕЛИ С УЧЕТОМ ВЕСОВЫХ КОЭФФИЦИЕНТОВ.....	40
4.3	РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ.....	41
4.4	ПРЕИМУЩЕСТВА И НЕДОСТАТКИ МЕТОДА ПРОГНОЗИРОВАНИЯ ПЛОТНОСТИ И МЕТОДА КВАНТИЛЬНОЙ РЕГРЕССИИ	42
5	ЗАКЛЮЧЕНИЕ	43
	ПРИЛОЖЕНИЯ	44
	СПИСОК ЛИТЕРАТУРЫ	58

Аннотация

Работа выполнена в рамках алгебраического подхода к проблеме распознавания, развиваемого научной школой академика РАН Ю. И. Журавлева. Рассматривается задача прогнозирования временных рядов, в частности, методы построения адаптивных композиций прогнозирования при несимметричной функции потерь.

В работе предлагается новый метод прогнозирования при несимметричной функции потерь, основанный на оценивании плотностей распределения ошибок по скользящему контролю. Приведены результаты экспериментального сравнения (на реальных данных объёмов продаж) двух методов прогнозирования: квантильной регрессии (quantile regression) и прогнозирования плотности распределения (density forecast). Эксперименты показали, что предложенный метод является более вычислительно эффективным, по сравнению со стандартным методом квантильной регрессии, и может использоваться при произвольной функции потерь, а не только для несимметричной L_1 -нормы.

1 Введение

В настоящей работе рассматривается задача прогнозирования временных рядов. Существует большое количество методов построения прогнозов, в том числе авторегрессионные, корреляционные, спектральные, нейросетевые, прогнозирование плотности (density forecast), квантильная регрессия (quantile regression) и так далее. Все эти алгоритмы основаны на той или иной модели временного ряда, учитывающей специфические особенности конкретной задачи прогнозирования или класса похожих задач. В данной работе особое внимание уделяется прогнозированию временных рядов с несимметричной функцией потерь. Во многих прикладных задачах прогнозирования и планирования завышение или занижение прогноза ведут за собой потери принципиально разного характера. Например, при планировании объема закупок товара для магазина завышенный прогноз влечет расходы на хранение и замораживание средств, а заниженный прогноз – потерю потенциальной прибыли. Для современных ритейлинговых сетей потери от занижения прогноза, как правило, существенно выше потери от завышения. В данной работе рассматривается несимметричная кусочно-линейная функция потерь, который отражает специфику таких задач более адекватно, чем стандартная квадратичная функция, см. Рис.1.1.

Стандартным методом для построения прогнозов при кусочно-линейной функции потерь, представляющей собой несимметричную L_1 -норму – это метод квантильной регрессии [4]. Главным её недостатком является вычислительная неэффективность, особенно на длинных временных рядах. В данной работе предлагается более вычислительно эффективный метод, основанный на оценивании плотности распределения ошибок по скользящему контролю.

Большое разнообразие методов прогнозирования на практике приводит к тому, что строится некоторое число различных алгоритмов, возможно, основанных на разных моделях, учитывающих различные факторы, но каждый из этих алгоритмов обладает не достаточно хорошим качеством прогнозирования. В таких случаях имеет смысл строить композиции алгоритмов прогнозирования, в которых погрешности различных базовых алгоритмов взаимно компенсируются с помощью корректирующей операции. Общая методология построения алгоритмических композиций в задачах распознавания и прогнозирования развита в алгебраическом подходе академика Ю. И. Журавлева [2]. Во многих практических задачах прогнозирования временные ряды нестационарны, и алгоритмы, демонстрирующие наилучшее качество прогнозов, с течением времени могут «устаревать», уступая место другим алгоритмам. Модель временного ряда может изменяться со временем. Поэтому алгоритмические композиции должны обладать свойством адаптивности [6]. Стандартный способ обеспечить адаптивность заключается в том, чтобы по мере поступления новых данных перенастраивать как базовые алгоритмы, входящие в композицию, так и корректирующую операцию.

Цель данной работы – предложить метод построения адаптивных композиций алгоритмов прогнозирования при несимметричной функции потерь.

В первом разделе данной работы рассматривается общая постановка задачи прогнозирования временных рядов, целесообразность использования несимметричной функции потерь при прогнозировании временных и рассматривается прикладная задача прогнозирования потребительского спроса. Формулируется постановка задачи.

Второй раздел содержит описание различных методов прогнозирования временных рядов: как при симметричной функции потерь, так и при несимметричной. Особое внимание уделяется адаптивным моделям, в частности, алгоритмам экспоненциального сглаживания с учетом тренда и сезонности, задаче прогнозирования плотности, а также

методу квантильной регрессии. Кроме этого, проводится краткий обзор построения композиций алгоритмов при симметричной функции потерь. Формулируются различные варианты решения задачи.

Третий раздел содержит эффективный метод построения прогноза плотности, основанный на стандартных алгоритмах точечного прогнозирования. Описывается эксперимент, подтверждающий возможность использования данного метода.

В четвертом разделе сравниваются два метода построения прогнозов при несимметричной функции потерь: метод квантильной регрессии и метод построения прогноза плотности. Для чистоты эксперимента в каждом методе рассматривается идентичная авторегрессионная функция. Проводятся результаты вычислительных экспериментов и сравнительный анализ указанных методов по скорости и качеству прогнозирования. В качестве прикладной задачи рассматривается задача прогнозирования потребительского спроса в супермаркетах. Предложенные методы протестированы на реальных данных объемов продаж.

1.1 Задача прогнозирования временных рядов.

Совокупность величин (y_1, y_2, \dots, y_t) , представляющая собой значения некоторого параметра, в моменты времени $1, 2, \dots, t$ называется временным рядом. Задача прогнозирования заключается в оценивании значения измеряемой величины $\hat{y}_{t+1}, \hat{y}_{t+2}, \dots$, в будущие моменты времени $t+1, t+2, \dots$.

Рассмотрим две постановки задачи прогнозирования временных рядов.

1) Необходимо определить прогноз некоторой величины в момент времени $t+d$, как функцию значений этой величины в предыдущие моменты времени $1, 2, \dots, t$ так, чтобы минимизировать квадрат разности прогнозируемого и реального значения:

$$\hat{y}_{t+d} = f(y_1, y_2, \dots, y_t), \quad d = \{1, 2, 3, \dots\} \quad (1.1)$$

$$Q(t) = (y_{t+d} - \hat{y}_{t+d})^2 \rightarrow \min$$

2) Необходимо также определить прогноз некоторой величины в момент времени $t+d$, как функцию значений этой величины в предыдущие моменты времени $1, 2, \dots, t$, но в этом случае минимизируется произвольная неквадратичная функция потерь (частный случай, кусочно-линейная функция потерь, представляющая собой несимметричную L_1 -норму):

$$\hat{y}_{t+d} = f(y_1, y_2, \dots, y_t), \quad d = \{1, 2, 3, \dots\} \quad (1.2)$$

$$Q(t) = F(y_{t+d}, \hat{y}_{t+d}) \rightarrow \min$$

1.2 Задача прогнозирования потребительского спроса.

В настоящей работе рассматривается задача прогнозирования потребительского спроса. Она заключается в том, чтобы предсказать объем продаж заданного товара в заданном магазине на заданный интервал времени вперед. Конечной целью прогнозирования является планирование закупок. Прогноз является основным фактором в принятии решения спроса.

Решения о сроках и объектах закупок принимаются на основании не только прогнозов потребительского спроса, но и календаря поставок, возможностей поставщика, остатка товара, выделяемого бюджета на заказ товара, максимальной загрузке склада данным товаром. Однако в данной работе рассматривается только прогнозирование спроса, как наиболее наукоемкая задача, решаемая в рамках планирования закупок.

Задачи прогнозирования потребительского спроса обладают специфическими особенностями. При прогнозировании необходимо учитывать ряд следующих факторов:

сезонность, тренд, праздники, рекламные акции, временное отсутствие товара на складе, неслучайное отсутствие спроса на товар, взаимное влияние продаж различных товаров, планы продаж, параметры выкладки товаров. Рассмотрим некоторые факторы подробнее.

Тренд. Трендом называется изменение, определяющее общее направление развития, основную тенденцию временного ряда. Часто рассматривают линейную модель тренда: $y_t = \alpha + \beta t + \varepsilon_t$, где α, β - коэффициенты тренда, ε_t - случайная величина, которая является стационарным временным рядом с нулевым математическим ожиданием. Другие примеры трендов: $y_t = \alpha + \beta t + \gamma t^2$ - квадратичный тренд, $y_t = \alpha e^{\beta t}$ - экспоненциальный тренд.

Сезонность. Временные ряды могут учитывать сезонную компоненту. Например, в квартальных данных может наблюдаться сезонная компонента с периодом n :

$$y_t = S(t) + \varepsilon_t$$

$$S(t+n) = S(t)$$

Здесь ряд y_t представлен в виде композиции детерминированной составляющей $S(t)$ (сезонная компонента) и случайной составляющей, которая является стационарным временным рядом с нулевым средним. Например, в задаче прогнозирования потребительского спроса годовой сезонностью обладает спрос на мороженое (летом мороженого продается больше, чем в остальные времена года).

Рекламные акции. К рекламным акциям относятся промо-акции, распродажи, реклама в СМИ. При проведении рекламных акций спрос на соответствующий товар резко возрастает. Например, явный рост продаж на напиток Pepsi-Cola наблюдался во время рекламной акции с 17 мая по 14 июня 2005 года. Однако не всегда рекламные акции дают однозначное увеличение спроса на соответствующий товар: рекламные акции могут привести к резкому уменьшению спроса на другие товары рекламируемой группы или товары-заменители.

Временное отсутствие товара в продаже. При временном отсутствии товара на складе спрос на него может оставаться ненулевым, однако продажи будут равны нулю. Интервалы времени, в течение которых товар отсутствует необходимо исключать из рассмотрения, так как они искажают информацию о реальном спросе.

Влияние продаж взаимозаменяемых и взаимодополняемых товаров. Примером заменяемости является уменьшение продаж холодного чая Lipton с появлением новых сортов холодного чая. Примером дополняемости является покупка чипсов при покупке пива. Задача поиска взаимозаменяемых и взаимодополняемых товаров, как правило, решается методами корреляционного анализа или векторной регрессии. Если товары обладают высокой корреляцией, то с высокой вероятностью можно утверждать, что они являются взаимодополняющими, а если товары обладают высокой антикорреляцией, то их можно отнести к взаимозаменяемым товарам. Задача учета влияния взаимозаменяемых и взаимодополняемых товаров является сложной, и в данной работе не рассматривается.

1.3 Несимметричная функция потерь

Большинство стандартных методов прогнозирования исходят из принципа минимизации квадратичной функции потерь:

$$F(y_t, \hat{y}_t) = (y_t - \hat{y}_t)^2, \quad (1.3)$$

где \hat{y}_t - прогноз, y_t - реальное значение.

Однако симметричная квадратичная функция потерь во многих прикладных задачах, в частности, в задаче прогнозирования потребительского спроса не вполне адекватно оценивает величину потерь, чем несимметричная кусочно-линейная функция, которая позволяет оценивать величину потерь в денежном выражении и задается следующим выражением:

$$F(y_t, \hat{y}_t) = \begin{cases} f_1(\hat{y}_t - y_t), & \text{при } \hat{y}_t \geq y_t; \\ f_2(y_t - \hat{y}_t), & \text{при } \hat{y}_t \leq y_t; \\ 0, & \text{при } \hat{y}_t = y_t; \end{cases} \quad (1.4)$$

где f_1, f_2 - потери, соответственно, при завышенном и заниженном прогнозах в расчете на единицу товара.

Еще более детальный анализ потерь приводит к кусочно-нелинейной несимметричной функции потерь. На рис.1.1 показан вид такой функции для задачи прогнозирования потребительского спроса [14].

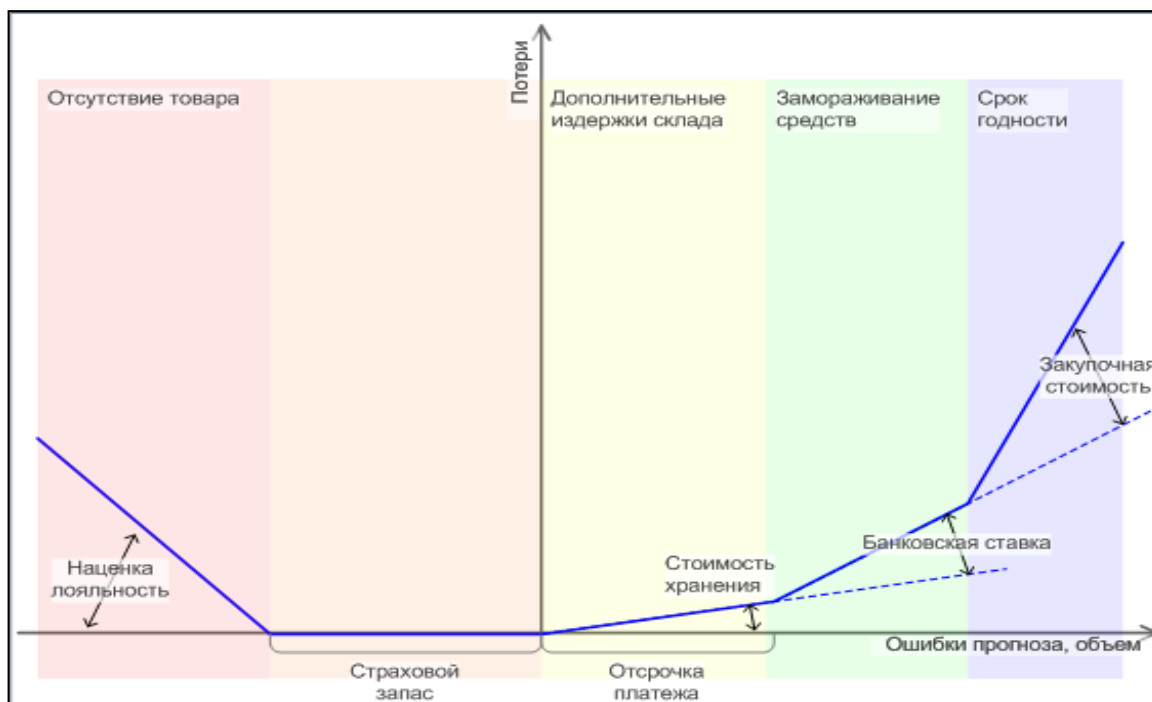


Рис.1.1 Общий вид несимметричной функции потерь.

Рассмотрим подробнее вид данной функции.

- 1) Если прогноз больше, чем продажи, то потери возникают по следующим причинам:
 - дополнительные издержки склада – необходимо учитывать стоимость хранения товара, который не был продан;
 - замораживание средств – деньги, потраченные на покупку товара, который не был продан. Более выгодным было бы внести эти деньги в банк. Потери от замораживания средств будут большими, чем от издержек склада;
 - срок годности. Если товар испортился, то мы теряем сумму закупочной стоимости данного товара, соответственно, потери в данном случае будут наибольшими.
- 2) Если прогноз меньше, чем продажи, то:
 - существует некоторый страховой запас товара, который был продан. В этом случае потери равны нулю;
 - после продажи страхового запаса товар будет отсутствовать, несмотря на имеющийся спрос.

Очевидно, во всех перечисленных случаях величина потерь линейно зависит от ошибки прогноза $\hat{y}_t - y_t$.

Но при решении задачи прогнозирования потребительского спроса довольно сложно учитывать вышеописанную функцию, поэтому в данной работе мы будем использовать упрощенную функцию потерь, вид которой приведен на рис. 1.2 и которая задается следующим выражением:

$$F(y_t, \hat{y}_t) = \begin{cases} a(y_t - \hat{y}_t), & y_t \geq \hat{y}_t, \\ b(\hat{y}_t - y_t), & y_t \leq \hat{y}_t \end{cases} \quad (1.5)$$

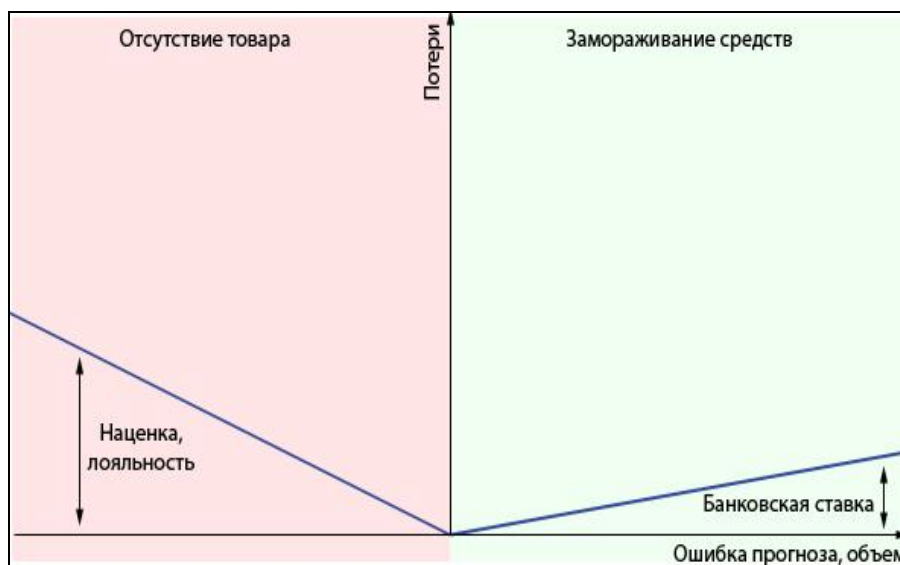


Рис.1.2. Вид несимметричной функции потерь, используемой в данной работе.

1.4 О необходимости построения алгоритмических композиций в задаче прогнозирования потребительского спроса.

Задача прогнозирования потребительского спроса обладает и другими специфическими особенностями. Большое количество временных рядов, равное произведению количества магазинов на число товаров в них (для некоторых предприятий порядок может быть близок к 10^5) накладывает жесткие ограничения на время вычисления прогнозов и настройки (адаптации) алгоритмов прогнозирования. Следующей особенностью является высокое отношение шум/сигнал, нестационарность (в частности, непостоянство дисперсии), что затрудняет применение многих стандартных методов прогнозирования.

Вследствие этих особенностей практически невозможно моделировать временные ряды в рамках какой-либо одной модели алгоритмов, что приводит к необходимости построения адаптивных композиций. Кроме того необходимо иметь эффективные методы настройки как базовых алгоритмов, так и корректирующей операции.

Наиболее общее определение алгоритмической композиции даётся в алгебраическом подходе Журавлёва. Вводится множество R , называемое пространством оценок, и рассматриваются алгоритмы, имеющие вид суперпозиции $A(x) = C(B(x))$, где функция $B: X \rightarrow R$ называется алгоритмическим оператором, а функция $C: R \rightarrow Y$ — решающим правилом. В задачах восстановления регрессии и прогнозирования решающие правила обычно не используют, полагая $R = Y$ и $C(B) = B$.

Алгоритмической композицией, составленной из алгоритмических операторов $B_j: X \rightarrow R$, $j = 1, \dots, p$, корректирующей операции $F: R^p \rightarrow R$ и решающего правила $C: R \rightarrow Y$ называется алгоритм $A: X \rightarrow Y$ вида

$$A(x) = C(F(B_1(x), \dots, B_p(x))), x \in X. \quad (1.6)$$

В данной работе будут предложены, описаны и протестированы методы настройки адаптивных композиций алгоритмов прогнозирования при несимметричной функции потерь. Также будем предполагать, что каждый алгоритм B_j и корректирующая операция F дают в качестве результата прогноз потребительского спроса за прогнозируемый период в будущем, и все алгоритмы B_j являются алгоритмами прогнозирования спроса при симметричной или несимметричной функции потерь.

1.5 Постановка задачи.

1) Описать методы решения задач прогнозирования при симметричной и несимметричной функциях потерь;

2) Предложить методы построения адаптивных композиций алгоритмов прогнозирования так, чтобы эти модели были настроены по несимметричной функции потерь, а данные методы были эффективны (так как, например, в задаче прогнозирования потребительского спроса мы имеем дело с огромным количеством временных рядов с большой предысторией);

3) Реализовать эти методы в среде MATLAB и провести вычислительные эксперименты на реальных задачах потребительского спроса;

4) Провести сравнительный анализ алгоритмов на реальных данных по критериям точности и времени работы.

2 Методы решения задач прогнозирования.

Выбор метода построения прогноза по временному ряду зависит от стоящей задачи и, как правило, определяется тем, какой прогноз требуется получить - долговременный или кратковременный. Построение долговременного прогноза является (как это и можно ожидать) задачей значительно более сложной, не всегда имеющей решение (причем не только по техническим причинам - в ряде случаев существует так называемый «горизонт прогноза», то есть максимальное время, на которое принципиально возможно дать прогноз поведения) и обычно требующей привлечения дополнительной информации о системе. В данном разделе рассматриваются различные методы решения задач прогнозирования при симметричной и несимметричной функции потерь, а также приводится обзор адаптивных моделей композиций алгоритмов.

2.1 Симметричная функция потерь.

Задача прогнозирования при симметричной функции потерь имеет огромное применение. В современной литературе известно множество методов для решения такой задачи. Выделяют следующие основные методы: авторегрессионные модели и модели скользящего среднего (алгоритмы ARIMA, GARCH, ARMA), нейросетевые, методы множественной линейной регрессии, а также адаптивные модели.

2.1.1 Авторегрессионные модели и модели скользящего среднего.

2.1.1.1 Модели ARMA и ARIMA.

Рассмотрим класс моделей стационарных временных рядов:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_p \varepsilon_{t-p}, \quad \varepsilon_t \approx iid(0, \sigma^2) \quad (2.1)$$

или в более короткой записи

$$\Phi(L) \cdot y_t = \delta + \Theta(L) \cdot \varepsilon_t, \quad \varepsilon_t \approx iid(0, \sigma^2) \quad (2.2)$$

где $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ и $\Theta(L) = 1 - \theta_1 L - \dots - \theta_p L^p$ – полиномы от оператора сдвига. Такая модель называется моделью авторегрессии и скользящего среднего (autoregressive moving average) или $ARMA(p, q)$.

Например, процесс ARMA (1,0)

$$y_t = \delta + \phi_1 \cdot y_{t-1} + \varepsilon_t \quad \varepsilon_t \approx iid(0, \sigma^2)$$

является $AR(1)$ процессом.

Моделью скользящего среднего (moving average) порядка q называется модель $ARMA(0, q)$

$$y_t = \delta + \Theta(L)\varepsilon_t = \delta + \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \dots - \theta_q \cdot \varepsilon_{t-q}, \quad \varepsilon_t \approx iid(0, \sigma^2), \quad (2.3)$$

которая обозначается $MA(q)$. Очевидно, что процесс $MA(q)$ стационарен при любом q и любых θ_i .

Существует условие обратимости процесса, то есть возможности его представления в виде AR процесса.

Предположим, что временной ряд y_t после того, как к нему применили d раз оператор последовательной разности, стал стационарным рядом $\Delta^d y_t$ удовлетворяющим $ARMA(p, q)$ модели

$$y_t - \phi_1 \cdot y_{t-1} - \dots - \phi_p \cdot y_{t-p} = \delta + \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \dots - \theta_p \cdot \varepsilon_{t-p}, \quad \varepsilon_t \approx iid(0, \sigma^2). \quad (2.4)$$

Тогда процесс y_t называется интегрированным процессом авторегрессии и скользящего среднего (integrated autoregression and moving average), $ARIMA(p, q, d)$.

2.1.1.2 Методология Бокса-Дженкинса.

Для прогнозирования временных рядов применяется методология Бокса-Дженкинса подбора $ARIMA$ модели, которая состоит из следующих трех этапов:

I. Идентификация модели

- I.1. Первый шаг – получение стационарного ряда. Ряд тестируется на стационарность с помощью визуального анализа графиков, визуального анализа автокорреляционной функции ACF и частной автокорреляционной функции PACF. Если ряд нестационарен, то необходимо применить оператор взятия последовательной разности (как правило, не больше двух раз) и повторить тестирование.
- I.2. После получения стационарного временного ряда, строятся его выборочные ACF и PACF, которые позволяют сформулировать несколько гипотез о возможных порядках авторегрессии (p) и скользящего среднего (q). Обычно рекомендуется использовать модели возможно более низкого порядка, как правило, с $p + q \leq 3$ (если нет сезонной компоненты).

Выборочные ACF и PACF, конечно, не обязаны в точности следовать теоретическим аналогам, но должны быть «достаточно близки» к ним.

II. Оценивание модели и проверка адекватности модели

- II.1. Для каждой из выбранных на первом этапе моделей оцениваются их параметры и вычисляются остатки.
- II.2. Каждая из моделей проверяется, насколько она соответствует данным. Из моделей, адекватных данным, выбирается самая простая модель, т. е. модель с наименьшим количеством параметров.

III. Прогнозирование

После того как на втором этапе выбрана модель, можно строить прогноз на один или несколько шагов по времени и оценивать доверительные границы прогнозных значений.

2.1.2 Множественная линейная регрессия.

2.1.2.1 Общие положения

В регрессионном анализе изучается связь и определяется количественная зависимость между зависимой переменной и одной или несколькими независимыми переменными. Множественная линейная регрессия предполагает, что результатом или искомой переменной является линейная функция, зависящая от входных и независимых переменных и ошибки.

$$Y_i = a_0 + a_1 \cdot x_{i1} + a_2 \cdot x_{i2} + \dots + a_k \cdot x_{ik} + \varepsilon_i,$$

где Y_i – зависимая переменная, a_0, a_1, \dots, a_k – линейные коэффициенты, $x_{i1}, x_{i2}, \dots, x_{ik}$ – независимые переменные, ε_i – ошибка.

Для проведения регрессионного анализа необходимо, чтобы случайные ошибки наблюдений имели нормальный закон распределения:

$$\varepsilon_i \approx N(0, \sigma), M(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma^2 = const,$$

также необходимым является взаимная некоррелированность (отсутствие автокорреляции) между ошибками наблюдений, т.е. последовательные значения ε_i не зависят друг от друга

$$M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 & \text{при } i = j \\ 0 & \text{при } i \neq j \end{cases}.$$

2.1.2.2 Метод наименьших квадратов.

Для нахождения линейных коэффициентов модели используется метод наименьших квадратов (МНК). Пусть проведено n независимых наблюдений случайной величины Y_i при соответствующих значениях $x_{i0}, x_{i1}, \dots, x_{ik}$, совместный закон распределения которых неизвестен. Необходимо оценить эмпирическую функцию регрессии

$$\tilde{Y}_i = \tilde{a}_0 + \tilde{a}_1 \cdot x_{i1} + \dots + \tilde{a}_k \cdot x_{ik}.$$

Согласно МНК, параметры подбираются таким образом, чтобы минимизировать сумму квадратов отклонений наблюдаемых значений от расчетных по модели значений

$$F = \sum_{i=1}^n (Y_i - \tilde{Y}_i)^2 = \sum_{i=1}^n (Y_i - \tilde{a}_0 - \tilde{a}_1 \cdot x_{i1} - \dots - \tilde{a}_k \cdot x_{ik})^2 \rightarrow \min, \quad (2.5)$$

где Y_i – наблюдаемые значения выходной переменной; \tilde{Y}_i – значения выходной переменной, рассчитанные по модели.

Введем матричные обозначения:

$y = (y_1 \ y_2 \ \dots \ y_k)^T$ – вектор наблюдений,

$a = (a_0 \ a_1 \ \dots \ a_k)^T$ – вектор параметров,

$$A = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \dots & & & \ddots & \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} \text{ – регрессионная матрица размерности } (nk + 1).$$

Получаем новую запись для F :

$$F = \|y - Aa\|^2 \rightarrow \min$$

Из необходимых условий минимума

$$\frac{\partial F}{\partial a} = 0,$$

или

$$A^T \cdot (y - Aa) = 0$$

Система нормальных уравнений имеет вид $(A^T \cdot A) \cdot a = A^T \cdot y$. При условии, что $A^T A$ – невырожденная матрица, решение системы можно записать в виде

$$a = (A^T \cdot A)^{-1} \cdot A^T \cdot y$$

2.1.2.3 Проверка предпосылок регрессионного анализа

Оценка адекватности модели.

Для оценки адекватности рассчитанной регрессионной модели вычисляется коэффициент детерминации

$$R^2 = \frac{\sum_{n=1}^k (\hat{Y}_n - \bar{Y})^2}{\sum_{n=1}^k (Y_n - \bar{Y})^2},$$

где \hat{Y}_n – прогнозные значения, \bar{Y} – среднее значение для Y , $\bar{Y} = \frac{1}{k} \cdot \sum_{n=1}^k Y_n$. Коэффициент

детерминации показывает долю правильно спрогнозированных величин. Если $R^2 = 0.75$ это значит, что модель работает на 75%, а 25% приходится на ошибку или неучтенные в модели факторы

Далее вычисляется величина $R = \sqrt{R^2}$, которая является оценкой множественного коэффициента корреляции между результатами наблюдений и вычисленными значениями \hat{Y}_n .

Проверка значимости параметров модели.

Для проверки значимости l -го параметра уравнения регрессии используют t -критерий:

$$t_l = \frac{a_l}{S(a_l)},$$

подчиняющуюся закону распределения Стьюдента с $(k - m - 1)$ степенями свободы (где m – количество параметров x_{ik}). Знаменатель этого выражения – среднее квадратичное отклонение коэффициента уравнения регрессии a_m :

$$S(a_m) = \sqrt{(X^T \cdot X)^{-1}_{jm} \cdot S_{ocm}^2},$$

где S_{ocm} – среднее квадратичное отклонение для остатков:

$$S_{ocm}^2 = \frac{1}{k - m - 1} \cdot \sum_{n=1}^k (Y_n - \tilde{Y}_n)^2.$$

Когда расчетное значение t -критерия превосходит его табличное значение при заданном уровне значимости, коэффициент регрессии считается значимым.

Проверка гипотезы о равенстве нуля всех коэффициентов.

Для проверки гипотезы о равенстве нулю всех коэффициентов рассчитывается F -критерий:

$$F = \frac{1}{k-1} \cdot \frac{\sum_{n=1}^k (Y_n - \bar{Y}_n)^2}{S_{ocm}^2},$$

где \bar{Y}_n – среднее значение для Y , а F подчиняется распределению Фишера с $(k-1)$ и $(k-m-1)$ степенями свободы. Если расчетное значение F -критерия больше его табличного значения при заданном уровне значимости, то уравнение регрессии считается значимым.

Проверка на автокорреляцию случайных ошибок.

Критерий Дарбина-Уотсона позволяет определить наличие автокорреляции остатков в модели:

$$DU = \frac{\sum_{n=1}^k (d_n - d_{n-1})^2}{\sum_{n=1}^k d_n^2},$$

где $d_n = Y_n - \tilde{Y}_n$ – разность прогнозируемого и фактического значения (остаток) на интервале в k точек, для которых Y_n известно. Если значение DU близко к 2, то автокорреляция остатков отсутствует. Предельное отклонение от 2, при котором гипотеза о некоррелированности остатков принимается – табличная величина, зависящая от N .

2.1.3 Адаптивные модели.

2.1.3.1 Общие положения.

Во многих практических задачах прогнозирования временные ряды нестационарны, поэтому алгоритмы, демонстрирующие наилучшее качество прогнозов могут устаревать, уступая место другим алгоритмам. Модель временного ряда может изменяться со временем. Поэтому алгоритмические модели должны быть адаптивны. Стандартный способ обеспечить адаптивность заключается в том, чтобы по мере поступления новых данных перенастраивать алгоритмическую модель.

Одной из самых распространенных адаптивных моделей является метод экспоненциального сглаживания. Суть метода состоит в следующем. Каждое значение y_i берется с определенным весом: последнее значение берется с весом 1, предпоследнее с весом α , следующее с весом α^2 и так далее. Экспоненциальное сглаживание осуществляется по рекуррентной формуле:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1}, \quad (2.6)$$

где S_t – значение экспоненциальной средней в момент t , α – параметр сглаживания.

Если последовательно применить рекуррентное соотношение (2.6), то экспоненциальную среднюю S_t можно выразить через значения временного ряда y :

$$S_t = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i y_{t-i} + (1-\alpha)^t S_0,$$

где N – количество членов ряда, S_0 – некоторая величина, характеризующая начальные условия для первого применения формулы (2.6) при $t = 1$.

При $N \rightarrow \infty$ $(1 - \alpha)^N \rightarrow 0$, а сумма коэффициентов $\alpha \sum_{i=0}^{N-1} (1 - \alpha)^i \rightarrow 1$. Тогда

$$S_t = \alpha \sum_{i=0}^{N-1} (1 - \alpha)^i y_{t-i}.$$

Таким образом, величина S_t оказывается взвешенной суммой всех членов ряда. Причем веса падают экспоненциально в зависимости от «возраста» наблюдений. Такая модель называется адаптивной экспоненциального типа, а величину α называют параметром адаптации.

Докажем важную лемму.

Лемма 1. Пусть прогноз в момент времени $t + 1$ строится исходя из минимизации следующего функционала потерь: $\sum_{i=1}^t w^{t-i} (\hat{y}_{t+1} - y_i)^2$, $w \in (0,1)$. Тогда для вычисления прогноза может быть применена стандартная формула для экспоненциального сглаживания.

Доказательство:

Минимизируем данный функционал потерь:

$$\sum_{i=1}^t w^{t-i} (\hat{y}_{t+1} - y_i)^2 \rightarrow \min, \Rightarrow \sum_{i=1}^t 2w^{t-i} (\hat{y}_{t+1} - y_i) = 0, \Rightarrow \hat{y}_{t+1} = \frac{\sum_{i=1}^t w^{t-i} y_i}{\sum_{i=1}^t w^{t-i}}.$$

$$\hat{y}_{t+1} = \frac{\sum_{i=1}^t w^{t-i} y_i}{\sum_{i=1}^t w^{t-i}} = \frac{w^{t-1} y_1 + w^{t-2} y_2 + \dots + w^0 y_t}{1 + w + \dots + w^{t-1}} = \frac{w(w^{t-2} y_1 + w^{t-3} y_2 + \dots + y_{t-1})}{1 + w + \dots + w^{t-1}} + \frac{y_t}{1 + w + \dots + w^{t-1}}.$$

Используя формулу для подсчета суммы бесконечно убывающей геометрической последовательности, получим:

$$\hat{y}_{t+1} = \frac{w \sum_{i=1}^{t-1} w^{t-i-1} y_i}{1 + w + \dots + w^{t-1}} + y_t (1 - w) = w \hat{y}_t + y_t (1 - w).$$

Пусть $w = 1 - \alpha$, $\alpha \in (0,1)$, тогда получим: $\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t$, что и требовалось доказать.

2.1.3.2 Адаптивные модели для различных типов тенденций и циклических явлений аддитивного и мультипликативного типа.

В экономике многие явления характеризуются периодически повторяющимися сезонными эффектами. Соответственно, временные ряды, их отражающие содержат периодические сезонные колебания. Эти ряды и их колебания можно представить как генерируемые моделями двух основных типов: моделями с мультипликативными и с аддитивными коэффициентами сезонности. Кроме этого, многие временные ряды имеют тренд. Эти ряды можно представить моделями с аддитивным и с экспоненциальным ростом.

С помощью адаптивных моделей можно построить множество комбинаций различных типов тенденций и циклических явлений аддитивного и мультипликативного вида. Ниже предлагаются девять всевозможных моделей, которые обобщенно выражены одной формулой:

$$\hat{a}_{1,t} = \alpha_1 d_1 + (1 - \alpha_1) d_2, \quad (2.7)$$

где $\hat{a}_{1,t}$ - текущий уровень ряда,

α_1 - параметр сглаживания;

значения d_1, d_2 даны в таблице 2.1, каждая клетка характеризует ту или иную модель.

Таблица 2.1 Значения множителей d_1, d_2 для комбинаций различных типов тенденций и циклических явлений аддитивного и мультипликативного вида.

Тенденция роста/Характер сезонности		Отсутствие сезонного эффекта	Аддитивный сезонный эффект	Мультипликативный сезонный эффект
Отсутствие тенденции роста	d_1	y_t	$y_t - \hat{g}_{t-l}$	y_t / \hat{f}_{t-l}
	d_2	y_{t-1}	y_{t-1}	y_{t-1}
Аддитивный рост	d_1	y_t	$y_t - \hat{g}_{t-l}$	y_t / \hat{f}_{t-l}
	d_2	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$	$\hat{a}_{1,t-1} + \hat{a}_{2,t-1}$
Экспоненциальный рост	d_1	y_t	$y_t - \hat{g}_{t-l}$	y_t / \hat{f}_{t-l}
	d_2	$\hat{a}_{1,t-1} \hat{r}_{t-1}$	$\hat{a}_{1,t-1} \hat{r}_{t-1}$	$\hat{a}_{1,t-1} \hat{r}_{t-1}$

$\hat{a}_{2,t}$ - коэффициент линейного роста, оцениваемый по формуле:

$$\hat{a}_{2,t} = \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1}, \quad 0 < \alpha_2 < 1;$$

\hat{r}_t - коэффициент экспоненциального роста, получаемый по формуле:

$$\hat{r}_t = \alpha_r \frac{\hat{a}_{1,t}}{\hat{a}_{1,t-1}} + (1 - \alpha_r) \hat{r}_{t-1}, \quad 0 < \alpha_r < 1;$$

\hat{g}_t - аддитивная сезонная компонента, вычисляется по формуле:

$$\hat{g}_t = \alpha_g (y_t - \hat{a}_{1,t}) + (1 - \alpha_g) \hat{g}_{t-l}, \quad 0 < \alpha_g < 1;$$

l - периодичность сезонного явления;

\hat{f}_t - мультипликативный коэффициент сезонности, подсчитывается по формуле:

$$\hat{f}_t = \alpha_f \frac{y_t}{\hat{a}_{1,t}} + (1 - \alpha_f) \hat{f}_{t-l}, \quad 0 < \alpha_f < 1.$$

Прогноз фактических значений $\hat{y}_{t+\tau}$, ожидаемых в момент времени $t + \tau$, подсчитывается для каждой модели способом, указанным в таблице 2.2.

Таблица 2.2 Формулы для подсчета вычисления прогноза для комбинаций различных типов тенденций и циклических явлений аддитивного и мультипликативного вида

Тенденция роста/Характер сезонности	Отсутствие сезонного эффекта	Аддитивный сезонный эффект	Мультипликативный сезонный эффект
Отсутствие тенденции роста	$\hat{a}_{1,t}$	$\hat{a}_{1,t} + \hat{g}_{t-l+\tau}$	$\hat{a}_{1,t} \hat{f}_{t-l+\tau}$
Аддитивный рост	$\hat{a}_{1,t} + \hat{a}_{2,t}$	$\hat{a}_{1,t} + \hat{a}_{2,t} + \hat{g}_{t-l+\tau}$	$(\hat{a}_{1,t} + \hat{a}_{2,t}) \hat{f}_{t-l+\tau}$
Экспоненциальный рост	$\hat{a}_{1,t} \hat{r}_t^\tau$	$\hat{a}_{1,t} \hat{r}_t^\tau + \hat{g}_{t-l+\tau}$	$\hat{a}_{1,t} \hat{r}_t^\tau \hat{f}_{t-l+\tau}$

2.1.4 Нейронные сети.

2.1.4.1 Общие положения.

Нейронные сети были разработаны на основе функционирования нервной клетки. Основой метода явилось то, что нейрон достаточно легко представить в машинном виде, а всю сложность нервной клетки и другие качества ее работы можно определить связями. Это обуславливает применимость нейронных сетей для различных видов задач обучения по прецедентам, в том числе распознавания, классификации, регрессии и прогнозирования.

Нейронная сеть состоит из простейших элементов – нейронов. Каждый нейрон обладает группой синапсов – однонаправленных входных связей, соединенных с выходами других нейронов, а также имеет аксон – выходную связь данного нейрона, с которой сигнал поступает на синапсы следующих нейронов. Каждый синапс характеризуется величиной синаптической связи или ее весом w_i . Общий вид нейрона приведен на рисунке 2.1.

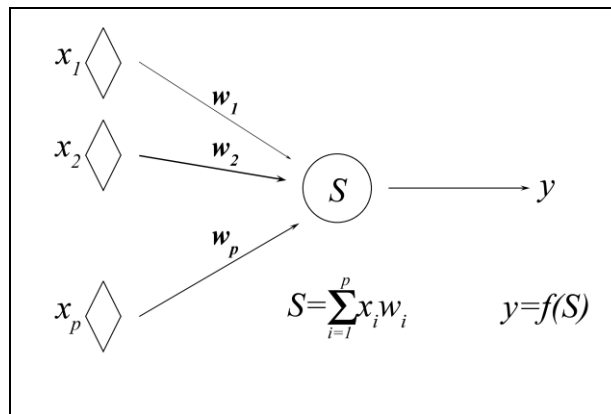


Рис. 2.1 Общий вид нейрона.

Текущее состояние нейрона определяется, как взвешенная сумма его входов:

$$s = \sum_{i=1}^p x_i \cdot w_i$$

Выход нейрона есть функция его состояния:

$$y = f(s)$$

Одной из наиболее распространенных является нелинейная функция с насыщением (так называемая логистическая функция или сигмоид (т.е. функция S-образного вида)):

$$f(x) = \frac{1}{1 + e^{-\alpha x}}$$

Одно из ценных свойств сигмоидной функции – простое выражение для ее производной:

$$f'(x) = \alpha \cdot f(x) \cdot (1 - f(x))$$

Любая нейронная сеть состоит из входного и выходного слоя и может иметь один или более скрытых слоев. На входной слой подается вектор исходных данных $X = \{x_1, \dots, x_p\}$, затем входные данные преобразуются нейронами сети и выдают результат на выходной слой, где производится сравнение результата с пороговой величиной и выдается результат – вектор $Y = \{y_1, \dots, y_k\}$.

Помимо входного и выходного слоев в многослойной сети существуют так называемые скрытые слои. Они представляют собой нейроны, которые не имеют непосредственных входов исходных данных, а связаны только с выходами входного слоя и с входом выходного слоя. Таким образом, скрытые слои дополнительно преобразуют информацию и добавляют нелинейности в модели.

Общий вид нейронной сети показан на рисунке 2.2.

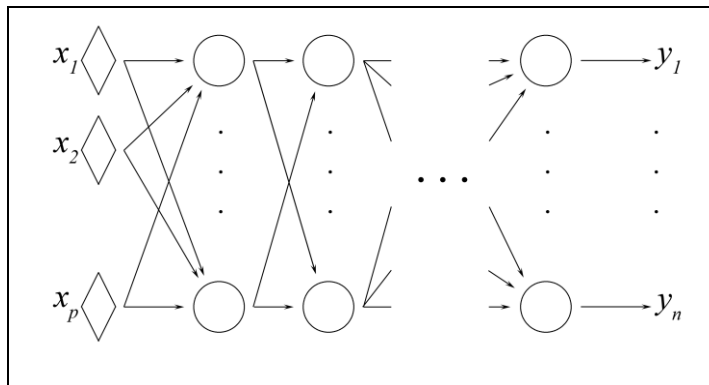


Рис. 2.2 Общий вид нейронной сети.

2.1.4.2 Классификация нейронных сетей.

По архитектуре связей нейронные сети могут быть сгруппированы в два класса: сети прямого распространения, в которых связи не имеют петель, и сети рекуррентного типа, в которых возможны обратные связи.

Сети прямого распространения подразделяются на однослойные и многослойные персептроны.

Класс рекуррентных нейронных сетей гораздо обширнее, сами сети сложнее по своему устройству. Сеть организована так, что каждый нейрон получает входную информацию от других нейронов и, возможно, нейронов того же слоя и от самого себя. Этот тип сетей имеет важное значение, так как с их помощью можно моделировать нелинейные динамические системы.

Среди рекуррентных сетей можно выделить сети Хопфилда и сети Кохонена. С помощью сетей Хопфилда можно обрабатывать неупорядоченные (рукописные буквы), упорядоченные во времени (временные ряды) или пространстве (графики) образцы.

2.1.4.3 Обучение нейронной сети.

На этапе обучения происходит вычисление синаптических коэффициентов w_i в процессе решения нейронной сетью задач, в которых нужный ответ определяется не по правилам, а с помощью примеров, сгруппированных в обучающие множества. Кроме параметра качества подбора весов важную роль играет время обучения. Как правило, эти два параметра связаны обратной зависимостью и их приходится выбирать на основе компромисса.

Одним из самых распространенных алгоритмов обучения нейронных сетей с учителем является алгоритм обратного распространения ошибки (Back Propagation Algorithm). Идея алгоритма заключается в том, что нейронная сеть за один проход «просматривается» дважды: на прямом проходе вычисляется ошибка нейронной сети на работе всех прецедентных векторов, на обратном проходе производится подсчет ошибок.

Согласно методу наименьших квадратов, минимизируемой целевой функцией ошибки нейронной сети является величина:

$$\mathfrak{J}(w) = \sum_{i=1}^n \sum_{j=1}^k (y_{ij}^N - \hat{y}_{ij})^2 \rightarrow \min$$

где $y_{ij}^{(N)}$ – реальное выходное состояние нейрона j последнего (выходного) слоя N нейронной сети при подаче на ее входы i -го объекта; \hat{y}_{ij} – идеальное (желаемое) выходное состояние этого нейрона.

Суммирование ведется по всем нейронам выходного слоя и по всем обрабатываемым сетью объектам. Минимизация ведется методом градиентного спуска, что означает подстройку весовых коэффициентов следующим образом:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \frac{\partial \mathfrak{J}}{\partial w_{ij}} \quad (\text{НС 1})$$

Здесь w_{ij} – весовой коэффициент синаптической связи, соединяющей i -ый нейрон слоя $n-1$ с j -ым нейроном слоя n , η – шаг градиентного спуска, $0 < \eta < 1$.

Далее

$$\frac{\partial \mathfrak{J}}{\partial w_{ij}} = \frac{\partial \mathfrak{J}}{\partial y_j} \cdot \frac{dy_j}{dv_j} \cdot \frac{\partial v_j}{\partial w_{ij}} \quad (\text{НС 2})$$

Здесь под y_j , как и раньше, подразумевается выход нейрона j , а под v_j – взвешенная сумма его входных сигналов, то есть аргумент активационной функции. Так как множитель $\frac{dy_j}{dv_j}$ является производной этой функции по ее аргументу, из этого следует, что производная активационной функция должна быть определена на всей оси абсцисс. В связи с этим функция единичного скачка и прочие активационные функции с неоднородностями не подходят для рассматриваемых нейронных сетей. В них применяются такие гладкие функции, как гиперболический тангенс или классический сигмоид с экспонентой.

Третий множитель $\frac{\partial v_j}{\partial w_{ij}}$, очевидно, равен выходу нейрона предыдущего слоя $y_i^{(n-1)}$.

Что касается первого множителя в (2), он легко раскладывается следующим образом:

$$\frac{\partial \mathfrak{J}}{\partial y_j} = \sum_k \frac{\partial \mathfrak{J}}{\partial y_k} \cdot \frac{dy_k}{dv_k} \cdot \frac{\partial v_k}{\partial y_j} = \sum_k \frac{\partial \mathfrak{J}}{\partial y_k} \cdot \frac{dy_k}{dv_k} \cdot w_{jk}^{(n+1)}$$

Здесь суммирование по k выполняется среди нейронов слоя $n+1$.

Введя новую переменную

$$\delta_j^{(n)} = \frac{\partial \mathfrak{J}}{\partial y_j} \cdot \frac{dy_j}{dv_j}$$

мы получим рекурсивную формулу для расчетов величин $\delta_j^{(n)}$ слоя n из величин $\delta_k^{(n+1)}$ более старшего слоя $n+1$.

$$\delta_j^{(n)} = \left[\sum_k \delta_k^{(n+1)} \cdot w_{jk}^{(n+1)} \right] \cdot \frac{dy_j}{ds_j}$$

Для выходного же слоя

$$\delta_l^{(N)} = (y_l^{(N)} - d_l) \cdot \frac{dy_l}{ds_l}$$

Теперь мы можем записать (2.1) в раскрытом виде:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \delta_j^{(n)} \cdot y_i^{(n-1)}$$

И скорректировать все веса в нейронной сети по формуле:

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t).$$

2.2 Несимметричная функция потерь.

Задача прогнозирования при несимметричной функции потерь ставилась многими исследователями, но, по сравнению с прогнозированием при симметричной функции потерь, занимает скромное место в литературе по прогнозированию временных рядов. Главной проблемой является построение вычислительно эффективных алгоритмов прогнозирования. К тому же, необходимо строить такие алгоритмы, которые бы могли строить прогнозы при любой несимметричной функции, а не только для конкретной несимметричной кусочно-линейной функции. Основным методом прогнозирования при несимметричной функции потерь является метод квантильной регрессии. Кроме этого, в настоящее время активно развиваются различные методы прогнозирования плотности, которые в качестве прогноза выдают не число, а конкретное распределение, что является очень важным при прогнозировании потребительского спроса, курсов валют и ценных бумаг, прогнозировании инфляции и т.д.

2.2.1 Квантильная регрессия.

Метод квантильной регрессии позволяет получать не только точные прогнозы, но и предсказывать границы колебаний показателей и строить точные прогнозы при несимметричной функции потерь. К сожалению, в России данный метод почти неизвестен, так как в научной литературе квантильная регрессия до настоящего времени практически не рассматривалась.

Метод квантильной регрессии является устойчивым к отклонениям от предположений классической модели. Одним из важных свойств является то, что модель устойчива к «выбросам», которые часто встречаются на практике, в особенности в случае показателей финансового рынка. Поэтому при прогнозе финансовых показателей использование модели квантильной регрессии более предпочтительно, чем использование классических моделей.

Кроме этого, на практике, в экономике часто некоторые значения зависимой величины могут быть цензурированы. В этом случае оценивание классическими методами затруднено. При использовании же моделей квантильной регрессии не нужно знать истинную величину зависимой переменной.

Общеупотребительные тесты, основанные на оценках методами наименьших квадратов и абсолютных модулей, свободны от распределения лишь асимптотически.

По существу, единственной серьезной претензией к этому методу является невысокая асимптотическая относительная эффективность для функций распределения со стандартными, типа гауссовских, «хвостами».

2.2.1.1 Общая модель квантильной регрессии.

Рассмотрим первоначальную модель квантильной регрессии, представленную Кенкером и Бассетом в 1978 году.

Пусть (y_i, x_i) , $i = 1, \dots, n$ - некоторые переменные, где x_i - $K \times 1$ - независимых переменных в уравнении регрессии. Допускается, что

$$y_i = x_i' \beta_\theta + u_{\theta i}, \quad \text{Quant}_\theta(y_i, x_i) = x_i' \beta_\theta, \quad (2.9)$$

где $\text{Quant}_\theta(y_i, x_i)$ обозначает условную квантиль y_i на векторе регрессора x_i . Следует отметить, что x_i и y_i наблюдаются без ошибки и уравнение (2.9) определено правильно. Если (2.9) определено некорректно (то есть, нелинейно), тогда модель можно рассматривать как лучший линейный прогноз для условной квантили. В данной модели распределение остаточного члена $u_{\theta i}$ остается неопределенным. Как показывает соотношение (2.9) единственным предположением является лишь то, что $u_{\theta i}$ удовлетворяет условию ограничения на квантиль:

$$\text{Quant}_\theta(u_{\theta i}, x_i) = 0. \quad (2.10)$$

Вообще, θ -ая простая квантиль ($0 < \theta < 1$) переменной, является решением следующей задачи:

$$\left(\sum_{i: y_i \geq b} \theta |y_i - b| + \sum_{i: y_i \leq b} (1 - \theta) |y_i - b| \right) \rightarrow \min_b \quad (2.11)$$

Аналог линейной модели для θ -ого квантиля определяется таким же образом. То есть $\hat{\beta}_\theta$, оценка для β_θ в соотношении (4) – называемом θ -ой квантильной регрессией – решает задачу:

$$\frac{1}{n} \left(\sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i \leq x_i' \beta} (1 - \theta) |y_i - x_i' \beta| \right) \rightarrow \min_\beta \quad (2.12)$$

Очевидно, при $\theta = 1/2$ получаем классическую задачу наименьших расстояний (LAD):

$$\frac{1}{n} \left(\sum_i \frac{1}{2} |y_i - x_i' \beta| \right) \rightarrow \min_\beta.$$

2.2.1.2 Представление квантильной регрессии в виде задачи линейного программирования.

Задача (5), то есть

$$\frac{1}{n} \left(\sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i \leq x_i' \beta} (1 - \theta) |y_i - x_i' \beta| \right) \rightarrow \min_\beta \quad (2.13)$$

может рассматриваться как задача линейного программирования (ЛП). Пару двойственных задач линейного программирования можно представить следующих образом:

$$\left| \begin{array}{l} \theta 1' u^+ + (1 - \theta) 1' u^- \rightarrow \min \\ X \beta + u^+ - u^- = y \\ u^+ \geq 0 \\ u^- \geq 0 \end{array} \right. , \quad \left| \begin{array}{l} y' \varphi \rightarrow \max \\ X' \varphi = 0 \\ \varphi \leq \theta \cdot 1 \\ -\varphi \leq (1 - \theta) \cdot 1 \end{array} \right. .$$

При этом:

$$\varphi_i \approx \theta - \frac{1}{2} + \frac{1}{2} \text{sgn}(y_i - x_i \beta),$$

$$u_i^+ = \begin{cases} y_i - x_i \beta, & y_i \geq x_i \beta \\ 0, & \text{иначе} \end{cases},$$

$$u_i^- = \begin{cases} 0, & \text{иначе} \\ \hat{y}_t - x_i \beta, & y_i \leq x_i \beta \end{cases}$$

Из теоремы двойственности следует решение задачи (2.9), так как обе задачи имеют допустимое решение, если матрица X имеет полный ранг. Из теории линейного программирования можно вывести и другие полезные свойства задачи (2.9).

2.2.1.3 Практическое применение метода квантильной регрессии.

В данной работе метод квантильной регрессии используется для построения прогнозов при несимметричной функции потерь, которая имеет следующий вид:

$$F(y_t, \hat{y}_t) = \begin{cases} a(y_t - \hat{y}_t), & y_t \geq \hat{y}_t, \\ b(\hat{y}_t - y_t), & y_t \leq \hat{y}_t \end{cases}$$

Очевидно, что для того чтобы получить прогноз \hat{y}_{t+1} в момент времени $t+1$, на основании предыстории ряда y_1, y_2, \dots, y_t необходимо минимизировать следующую функцию:

$$\left(\sum_{i: y_i \geq \hat{y}_{t+1}} \theta |y_i - \hat{y}_{t+1}| + \sum_{i: y_i \leq \hat{y}_{t+1}} (1-\theta) |y_i - \hat{y}_{t+1}| \right) \rightarrow \min_{\hat{y}_{t+1}}, \text{ где}$$

$$\theta = a/(a+b).$$

Данная задача с помощью следующих замен:

$$u_i^+ = \begin{cases} y_i - \hat{y}_{t+1}, & y_i \geq \hat{y}_{t+1} \\ 0, & y_i \leq \hat{y}_{t+1} \end{cases} \text{ и } u_i^- = \begin{cases} 0, & y_i \geq \hat{y}_{t+1} \\ \hat{y}_{t+1} - y_i, & y_i \leq \hat{y}_{t+1} \end{cases}$$

легко сводится к следующей задаче линейного программирования:

$$\begin{cases} \theta \mathbf{1}' u^+ + (1-\theta) \mathbf{1}' u^- \rightarrow \min_{u^+, u^-, \hat{y}_{t+1}} \\ u^+ \mathbf{1}' - u^- \mathbf{1}' + \hat{y}_{t+1} = y, \\ u^+, u^- \geq 0 \end{cases}$$

для которой существуют встроенные в MATLAB методы решения.

Если ввести коэффициент устаревания данных $w \in (0,1)$ и потери брать с определенным весом: потери по последнему значения – 1, по предпоследнему – w , по следующему w^2 и так далее, то для получения прогноза \hat{y}_{t+1} необходимо минимизировать следующую функцию:

$$\left(\sum_{i: y_i \geq \hat{y}_{t+1}} w^{t-i} \theta |y_i - \hat{y}_{t+1}| + \sum_{i: y_i \leq \hat{y}_{t+1}} w^{t-i} (1-\theta) |y_i - \hat{y}_{t+1}| \right) \rightarrow \min_{\hat{y}_{t+1}}.$$

Такая задача с помощью замен сводится к следующей задаче линейного программирования:

$$\begin{cases} \theta w^{t-i} u^+ + (1-\theta) w^{t-i} u^- \rightarrow \min_{u^+, u^-, \hat{y}_{t+1}} \\ u^+ \mathbf{1}' - u^- \mathbf{1}' + \hat{y}_{t+1} = y, \\ u^+, u^- \geq 0 \end{cases}$$

2.2.2 Построение и оценивание прогноза плотности.

Построение и оценивание прогноза плотности является довольно сложной задачей. Данная задача менее развита, чем построение точечных прогнозов при различных

функциях потерь. В настоящее время известно множество способов точечного прогнозирования, в то время неясно какие методы применять при прогнозировании плотности распределения. Прогнозирование плотности широко используется в макроэкономике, финансах, задаче прогнозирования потребительского спроса.

2.2.2.1 Построение прогноза плотности

Впервые долгосрочные прогнозы плотности в макроэкономике датируются 1968 годом, когда Статистическая секция бизнеса и экономики Американской статистической ассоциации и Национальное бюро экономических исследований совместно провели квартальный обзор макроэкономических прогнозов в США.[10] Позднее Федеральный банк развития Филадельфии стал использовать долгосрочные прогнозы плотности. В Великобритании построением прогнозов плотности стали заниматься гораздо позднее. В феврале 1996 года Банк Англии начал публиковать обзоры посвященные прогнозам плотности инфляции. Плотность прогноза инфляции представлялось аналитически ненормальным распределением, которое представляет два кусочных нормальных распределения с параметрами μ, σ_1, σ_2 :

$$f(y) = \begin{cases} A \exp(-(y - \mu)^2 / 2\sigma_1^2), & y \leq \mu \\ A \exp(-(y - \mu)^2 / 2\sigma_2^2), & y \geq \mu \end{cases}$$

где $A = (\sqrt{2\pi}(\sigma_1 + \sigma_2)/2)^{-1}$. Такое распределение формируется взятием двух половинок нормального распределения с параметрами (μ, σ_1) и (μ, σ_2) , и их наслаиванием, чтобы получить общее распределение $f(y)$ (Рис.2.3).

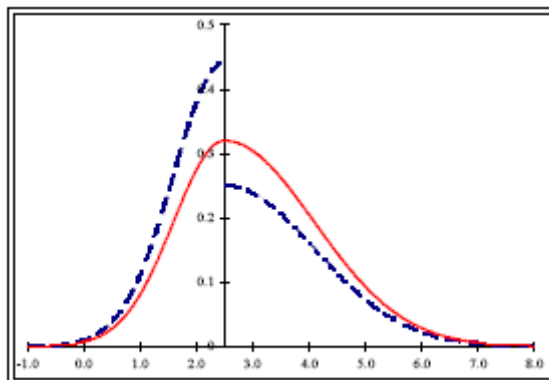


Рис. 2.3 Пример представления прогноза плотности инфляции.

Прогноз плотности инфляции публикуется и по сей день на протяжении последних 10 лет каждый квартал как Банком Англии, так и Национальным институтом экономика и социальных исследований в их ежегодных прогнозах. Прогноз плотности информирует пользователя о рисках, вовлекаемых в использование прогноза.

Самым распространенным способом построения прогноза плотности является его представление в виде нормального распределения с параметрами (μ, σ) . Очень часто используется метод экспертных оценок. Пусть имеется N прогнозов плотности $p_i(t)$, сделанных экспертами. Тогда результирующий прогноз определяется как взвешенная линейная комбинация экспертных прогнозов.

$$p(t) = \sum_{i=1}^N w_i p_i(t),$$

где w_i - набор неотрицательных весов, сумма которых равна единице.

При построении комбинации экспертных вопросов веса могут быть известны. А как определять веса в композиции прогнозы плотности, если они неизвестны? Допустим,

у нас есть N моделей для определения прогноза плотности и строится линейная композиция этих моделей. В работе [11] предлагается два варианта определения весов:

1) Самый простой и наиболее прямой способ – это сделать веса всех прогнозов плотности в композиции равными, то есть $w_i = 1/N$, $i = 1, \dots, N$.

2) Другим альтернативным методом является Байесовская модель усреднения. Маржинальная (или граничная) вероятность каждой модели $p(y_i)$ используется для подсчета весовых коэффициентов:

$$w_i = \frac{p(y_i)p(M_i)}{\sum_{i=1}^N p(y_i)p(M_i)}, \quad i = 1, \dots, N.$$

где $p(M_i) = 1/N$.

Если прогнозы плотности представлены нормальным распределением, но можно подсчитать весовые коэффициенты и построить результирующий прогноз плотности. Но ситуация усложняется, если прогнозы плотности не являются нормально распределенными или вообще прогноз плотности нельзя представить аналитически никаким распределением. В таком случае построение композиции описанными способами является трудноразрешимой задачей. В данной работе приводится эффективный способ построения прогноза плотности основанный на стандартных методах решения задач прогнозирования.

2.2.2.2 Способы представления прогноза плотности

Существует три способа представления прогноза плотности: аналитический (при помощи алгебраических выражений), графический и с помощью таблиц. Из этих способов нельзя выделить самый лучший, так как некоторые способы могут скрывать одни особенности прогноза и наоборот акцентировать внимание на другие.

2.2.2.3 Оценивание прогноза плотности

Оценивание прогноза плотности является довольно трудной задачей. При оценивании точечного прогноза всегда можно выбрать соответствующую функцию потерь, при оценивании прогноза плотности выбор функции потерь является довольно трудной задачей.

В настоящее время данная задача является открытой. Известно несколько подходов к решению задачи оценивания прогноза плотности, при которых функции потерь не используются. Эти методы оценивания основываются на вероятностном интеграле преобразования (Probability integral transform, PIT). Этот метод был предложен в 1998 году Диаболдом, хотя понятие integral transform использовалось ещё в 1952 года Розенблаттом. Пусть $f_t(y_t)$ — некоторая модель, а $p_t(y_t)$ — последовательность прогнозов плотности. Вероятностный интеграл преобразования (PIT) вычисляется следующим образом:

$$z_t = \int_{-\infty}^{y_t} p_t(u) du = P_t(y_t).$$

Данный метод основывается на связи между данными генерирующимися моделью с последовательностью прогнозов плотности $p_t(y_t)$. Суть PIT устанавливает следующая теорема:

Теорема. Пусть $\{y_t\}_{t=1}^m$ генерируется из $\{f_t(y_t|\Omega_t)\}_{t=1}^m$, где $\Omega_t = \{y_{t-1}, y_{t-2}, \dots\}$. Если последовательность прогнозов плотности $\{p_t(y_t)\}_{t=1}^m$ совпадает с $\{f_t(y_t|\Omega_t)\}_{t=1}^m$, то последовательность вероятностных интегралов преобразования $\{y_t\}_{t=1}^m$ по отношению к $\{p_t(y_t)\}_{t=1}^m$ является реализацией m независимых случайных величин, равномерно распределенных на отрезке $[0, 1]$.

Таким образом, для оценивания прогноза плотности данная теория предлагает оценить, является ли последовательность вероятностных интегралов преобразования независимыми случайными величинами, равномерно распределенными на отрезке $[0,1]$. Данный метод также легко обобщается на n -мерный случай. В случае оценивания прогноза плотности, построенного на h шагов вперед, наиболее просто разделить z серии на группы, для каждой проверить независимость, одинаковое и равномерное распределение. Например, для корректных прогнозов на 2 шага вперед, подгруппы $\{z_1, z_3, z_5\}$ и $\{z_2, z_4, z_6\}$ должны быть независимыми, равномерно и одинакового распределенными на отрезке $[0,1]$, хотя полная серия не обязательно должна быть таковой.

Интерпретация вероятностных интегралов преобразования (PIT) приводится на рис.2.4. На графиках синим цветом показывается «правильное» распределение $N(0,1)$, красным – прогнозируемое распределение. В первом случае прогнозируемое распределение совпадает с «правильным», в остальных случаях различно: $N(0,0.5)$, $N(0,2)$, $N(1,1)$. На графиках (рис 2.4) показана результирующая PIT-гистограмма для каждого из рассмотренных случаев.

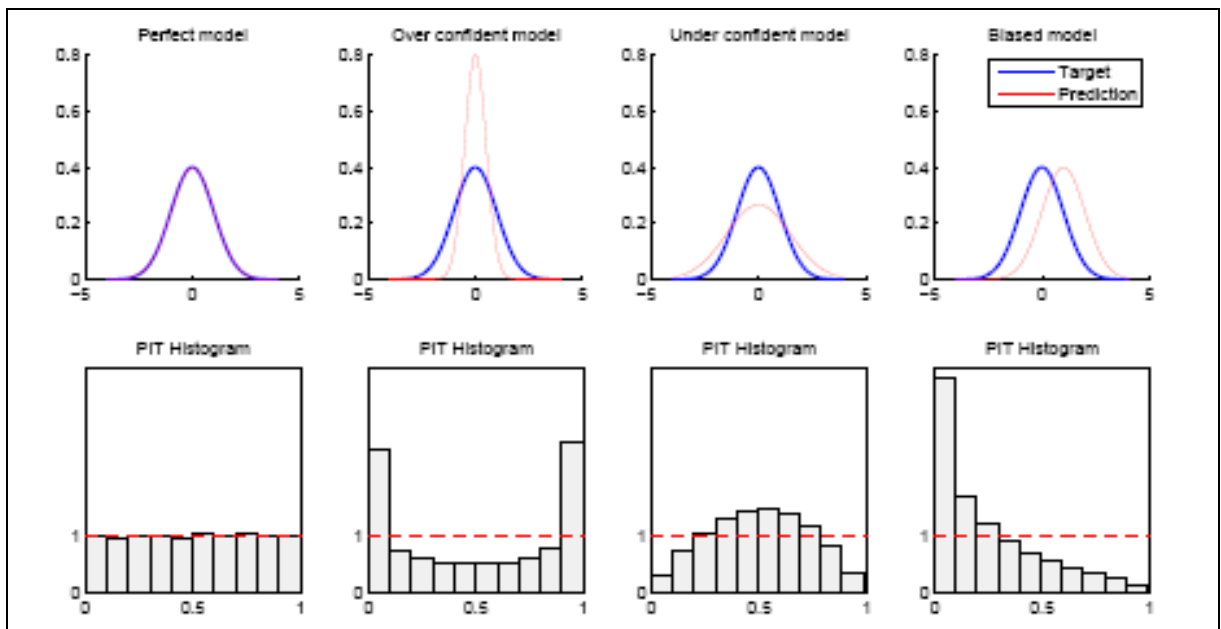


Рис. 2.4 Интерпретация вероятностных интегралов преобразования (PIT).

Для проверки правильности прогноза плотности существует целый ряд статистических тестов. Далее приведем наиболее распространенные из них.

- 1) Тест Колмогорова-Смирнова.
- 2) Тест Андерсона и Дарлинга. Является более сильным, чем тест Колмогорова-Смирнова.
- 3) Тест Льюнга-Бокса для проверки независимости $\{z_t\}$.
- 4) Тест Хонга для проверки независимости и равномерного распределения $\{z_t\}$.

Многие ученые предлагают преобразовать значения PIT (Probability integral transforms) в нормально распределенные с помощью преобразований обратной нормальной интегральной функции плотности. Использование этих преобразований объясняется тем, что тесты нормальности являются более мощными, чем тесты равномерности распределения. Приведем наиболее распространенные тесты проверки нормальности распределения:

- 1) Параметрический тест Берковица. Имеет смысл лишь для установления ненормальности путем первых двух моментов.
- 2) Тест Jarque-Bera (JB). Тест для проверки нормальности $\{z_t^*\}$. Для использования этого теста необходимо определить коэффициенты асимметрии и эксцесса. Не используется для нахождения расхождения в первых двух моментах. Одним из недостатков является ненадежность при оценивании для очень больших T .
- 3) Тест Doornik-Hansen (DH). Основывается на случайной выборке, к тому же не может определить расхождение в первых двух моментах.
- 4) Тест Diebold-Mariano (DB). Является надежным методом для проверки нормального распределения с математическим ожиданием 0 и дисперсией 1 $N(0,1)$.

В данной работе предлагается другой, вычислительно эффективный способ построения прогноза плотности и формула для вычисления потерь.

2.3 Композиции алгоритмов при симметричной функции потерь.

Очень часто приходится сталкиваться с нестационарностью, огромным количеством временных рядов (например, в задаче прогнозирования потребительского спроса для некоторых предприятий их порядок может быть близок к 10^5) или с отсутствием наилучшей модели временного ряда. В таком случае ни один из стандартных алгоритмов прогнозирования не может обладать достаточно хорошим качеством прогнозирования и тогда имеет смысл строить композиции алгоритмов прогнозирования. Преимущество композиций алгоритмов в том, что погрешности алгоритмов могут компенсировать друг друга. Другими словами, увеличения точности прогнозов достигается за счет уменьшения дисперсий ошибок.

В этом разделе рассмотрим способы построения композиций алгоритмов при симметричной функции потерь, которые использовались в данной работе. Предполагаем, что каждый базовый алгоритм B_i и корректирующая операция F дают в качестве результата прогноз потребительского спроса за прогнозируемый период в будущем. Таким образом, корректирующая операция A строится в виде суперпозиции: $A = F(B_1, \dots, B_p)$.

Будем строить корректирующую операцию по формуле

$$A_t = \sum_{i=1}^p w_t^i \cdot B_t^i,$$

где B_t^i – прогнозы базовых алгоритмов в момент времени t , w_t^i – весовые коэффициенты. На веса можно накладывать следующие ограничения:

$$w_t^i \geq 0 \quad \text{и} \quad \sum_{i=1}^p w_t^i = 1.$$

Первое ограничение типа неравенства $w_t^i \geq 0$ отвечает за монотонность корректирующей операции, то есть при увеличении (уменьшении) ответа любого из базовых алгоритмов B_1, \dots, B_p ответ корректирующей операции A неубывает (невозрастает). Второе ограничение типа равенства $\sum_{i=1}^p w_t^i = 1$ определяет условие нормировки, его необходимость определяется по результатам экспериментов. Выполнение

обоих условий означает выпуклость корректирующей операции, то есть ее ответ лежит в интервале между максимальным и минимальным прогнозами базовых алгоритмов.

2.3.1 Простое усреднение

Корректор представляет собой среднее арифметическое всех базовых алгоритмов

$$A = \frac{1}{p} \cdot \sum_{i=1}^p B_i. \text{ Далее вычисляется ошибка прогноза: } \varepsilon_t = y_t - A_t.$$

2.3.2 Выбор наилучшего алгоритма среди базовых

Построение корректора производится на основе результатов базовых алгоритмов. Из всех базовых алгоритмов выбирается тот, результаты которого давали наименьшую ошибку на большем количестве алгоритмов на информации обучения, то есть $A_t = B_t^i$, для которого средняя ошибка прогнозов $\varepsilon_t = y_t - B_t^i$ на предыстории была бы наименьшей.

2.3.3 Метод наименьших квадратов

Корректор построен таким образом, что зависит от всех алгоритмов из набора базовых алгоритмов. Каждому алгоритму присваивается вес w_t^i , который корректируется в процессе обучения таким образом, чтобы результат, выдаваемый корректором, был наилучший. В каждый момент времени t корректор строится по формуле $A_t = \sum_{i=1}^p w_t^i \cdot B_t^i$, где B_t – матрица из $(t-1)$ -ой строки.

Рассматривается 2 случая: с наложением ограничения неотрицательности весов и без него. В первом случае решалась задача математического программирования, имеющая вид:

$$\begin{cases} A_t = \sum_{i=1}^p w_t^i \cdot B_t^i \\ \sum_{i=1}^p w_t^i = 1 \\ w_t^i \geq 0 \end{cases},$$

во втором – решаемая задача математического программирования имела вид

$$\begin{cases} A_t = \sum_{i=1}^p w_t^i \cdot B_t^i \\ \sum_{i=1}^p w_t^i = 1 \end{cases}.$$

Для нахождения весов модели используется метод наименьших квадратов (МНК).

$$\sum_{s=1}^{t-1} \left(\sum_{i=1}^p w_t^i \cdot B_s^i - y_s \right)^2 \rightarrow \min_{w_t^1, \dots, w_t^p}$$

После перенастройки корректора по интервалу времени $[1, t-1]$ вычисляется прогноз алгоритмической композиции в момент времени t , ошибка прогноза $\varepsilon_t = y_t - A_t$ и ошибки усредняются по всем t . Полученная оценка называется оценкой скользящего контроля.

Для настройки МНК с одним ограничением равенства вида $\sum_{i=1}^p w_t^i = 1$ и одним ограничением неравенства вида $w_t^i \geq 0$ можно использовать стандартные методы математического программирования.

В системе MATLAB есть встроенные функции для решения задач вида

$$\frac{1}{2} w^T \cdot H \cdot w + f^T \cdot w \rightarrow \min_w$$

с ограничениями $A \cdot w \leq b$, $A_{eq} \cdot w = b_{eq}$, $lb \leq w \leq ub$, где H , A , A_{eq} являются матрицами, а f , b , b_{eq} , lb , ub и w – векторами.

Для того чтобы воспользоваться этой функцией, данные в матричной форме были приведены следующим образом:

$$\Phi = \|B \cdot w - y\|^2 \rightarrow \min_w,$$

$$\Phi = w^T \cdot B \cdot B^T \cdot w - 2 \cdot y \cdot B^T \cdot w + y^2 \rightarrow \min_w,$$

$$\frac{1}{2} \cdot w^T \cdot (B \cdot B^T) \cdot w + (-y \cdot B^T) \cdot w \rightarrow \min_w.$$

Таким образом, используя $H = B \cdot B^T$, $f = -y \cdot B^T$, $A_{eq} = [a_i = 1, i = \overline{1, p}]$, $b_{eq} = 1$, $lb = 0$, вычисляем весовые коэффициенты корректирующей операции с помощью встроенной в MATLAB функции.

2.3.4 Метод наименьших квадратов с регуляризацией

Корректор строится таким же образом как в предыдущем пункте $A_t = \sum_{i=1}^p w_t^i \cdot B_t^i$,

ограничения на веса остаются прежними $\sum_{i=1}^p w_t^i = 1$, $w_t^i \geq 0$. Однако при перенастройке

корректора в каждый момент времени возникает проблема неустойчивости весовых коэффициентов. Потребуем, чтобы веса не изменялись сильно при переходе от момента времени t к моменту времени $t+1$. Для формализации этого естественного требования

вводится штрафное слагаемое $\lambda \cdot \sum_{i=1}^p (w_t^i - w_{t-1}^i)^2$ в функционал качества:

$$\sum_{s=1}^{t-1} \left(\sum_{i=1}^p w_s^i \cdot B_s^i - y_s \right)^2 + \lambda \cdot \sum_{i=1}^p (w_t^i - w_{t-1}^i)^2 \rightarrow \min_{w_t^1, \dots, w_t^p}.$$

Вид этого слагаемого соответствует обычной практике регуляризации задач, некорректно поставленных по Тихонову. Параметр λ позволяет регулировать стабильность весов по времени. Этот параметр находится следующим образом: вычисляется средняя ошибка прогнозов по скользящему контролю для каждого фиксированного λ и выбирается такое значения $\lambda = \lambda^*$, при котором ошибка минимальна.

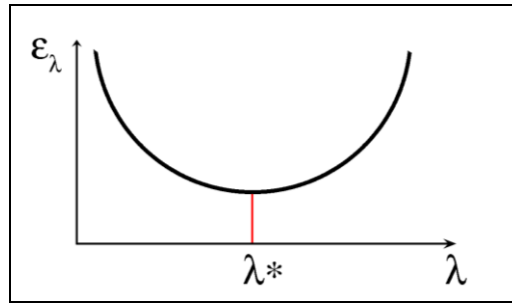


Рис. 2.5. Определение оптимального значения параметра регуляризации λ^*

Для инициализации процесса настройки весов в первый момент времени $t = 1$ всем весам приписывается значение $w_i^1 = \frac{1}{p}$, затем решается задача нахождения весов методом МНК. При переходе от момента времени t , $t = 2, \dots$ к моменту времени $t+1$ решается задача нахождения весов методом МНК с регуляризацией.

Эту задачу также можно свести к встроенной в MATLAB функции $\frac{1}{2} w^T \cdot H \cdot w + f^T \cdot w \rightarrow \min_w$. При аналогичном приведении данных получаем $H = B \cdot B^T + \lambda \cdot I_p$, $f = -(y \cdot B^T + \lambda \cdot w_{t-1}^T)$, ограничения на A_{eq} , b_{eq} и lb остаются прежними. Далее задача решается встроенной в систему MATLAB функцией.

Приведение данных:

$$\Phi = \|B \cdot w - y\|^2 + \lambda \cdot (w - w_{t-1})^T \cdot (w - w_{t-1}) \rightarrow \min_w,$$

$$\Phi = w^T \cdot B \cdot B^T \cdot w - 2 \cdot y \cdot B^T \cdot w + y^2 + \lambda \cdot w^T \cdot w - 2 \cdot \lambda \cdot w^T \cdot w_{t-1} + \lambda \cdot w_{t-1}^T \cdot w_{t-1} \rightarrow \min_w,$$

$$\frac{1}{2} \cdot w^T \cdot (B \cdot B^T + \lambda \cdot I_p) \cdot w + (-y \cdot B^T - \lambda \cdot w_{t-1}^T) \cdot w \rightarrow \min_w.$$

2.3.5 Метод локальной адаптации весов с регуляризацией

Корректор построен таким же образом, что рассмотренные выше: $A_t = \sum_{i=1}^p w_i^i \cdot B_i^i$.

Кардинальное отличие состоит в том, что учитывается история продаж товара не во все моменты времени t , $t = 1, 2, \dots, t-1$, а только за предыдущий момент $t-1$ времени t . Веса корректора также вычисляются методом наименьших квадратов с регуляризацией, то есть в модель введено штрафное слагаемое вида $\lambda \cdot \sum_{i=1}^p (w_i^i - w_{t-1}^i)^2$, которое позволяет регулировать стабильность весов по времени. Таким образом, веса находятся по следующей схеме:

$$\left(\sum_{i=1}^p w_i^i \cdot B_{t-1}^i - y_{t-1} \right)^2 + \lambda \cdot \sum_{i=1}^p (w_i^i - w_{t-1}^i)^2 \rightarrow \min,$$

в которой ограничения на веса остаются прежними: $\sum_{i=1}^p w_i^i = 1$, $w_i^i \geq 0$. Заметим, что этот случай соответствует рассмотренному в предыдущем разделе при $K = 0$.

Инициализация корректора повторяет шаги построения корректора методом наименьших квадратов с регуляризацией. Веса в начальный момент времени $t=1$ приравняются $w_t^i = \frac{1}{p}$, и в следующий момент времени $t, t=2, \dots$ находятся методом наименьших квадратов с регуляризацией.

Параметр λ находится следующим образом: вычисляется ошибка прогноза $\varepsilon_t(\lambda) = y_t - A_t$ для каждого фиксированного λ и выбирается такое значения $\lambda = \lambda^*$, при котором ошибка минимальна (рис. 2.5).

Предполагается, что в момент времени $t-1$ корректор настроен соответствующим образом, тогда для нахождения весов в момент времени t необходимо провести ту же процедуру.

В данной работе производится проверка эффективности корректора такого вида. Очевидно, что время работы такого алгоритма значительно меньше времени работы других корректоров, описанных в работе.

Задачу нахождения весовых коэффициентов методом локальной адаптации с регуляризацией также можно свести к встроенной в MATLAB функции $\frac{1}{2} w^T \cdot H \cdot w + f^T \cdot w \rightarrow \min_w$, при $H = B \cdot B^T + \lambda \cdot I_p$, (где B – вектор, а не матрица), $f = -(y \cdot B^T + \lambda \cdot w_{t-1}^T)$, ограничения на A_{eq} , b_{eq} и lb остаются прежними.

2.4 Композиции алгоритмов при несимметричной функции потерь.

В этом разделе рассмотрим способы построения композиций алгоритмов при несимметричной функции потерь. Также, как и в прошлом разделе, будем строить корректирующую операцию по формуле

$$A_t = \sum_{i=1}^p w_t^i \cdot B_t^i,$$

где B_t^i – прогнозы базовых алгоритмов в момент времени t , w_t^i – весовые коэффициенты. На веса можно накладывать следующие ограничения:

$$w_t^i \geq 0 \quad \text{и} \quad \sum_{i=1}^p w_t^i = 1.$$

По сути, методы построения композиции алгоритмов при несимметричной функции потерь аналогичны описанным методам для симметричной функции потерь, однако главным отличием является то, что для нахождения весов уже нельзя использовать метод наименьших квадратов, а нужно использовать метод наименьших модулей. Рассмотрим основные методы построения композиций.

2.4.1 Метод наименьших модулей.

Корректор построен таким образом, что зависит от всех алгоритмов из набора базовых алгоритмов. Каждому алгоритму присваивается вес w_t^i . В каждый момент

времени t корректор строится по формуле $A_t = \sum_{i=1}^p w_t^i \cdot B_t^i$. Веса настраиваются при заданном t по критерию минимума несимметричной кусочно-линейной функции потерь $F(A_t, y_t)$.

Рассмотрим 3 случая: без ограничений на сумму весов, с одним ограничением на сумму весов: $\sum_{i=1}^p w_t^i = 1$ и с двумя ограничениями на сумму весов: $\sum_{i=1}^p w_t^i = 1$ при $i \in I_{sym}$ и $\sum_{i=1}^p w_t^i = 1$ при $i \in I_{nsym}$, где I_{sym} и I_{nsym} - множества номеров базовых алгоритмов, настроенных, соответственно, при симметричной и несимметричной функциях потерь. Соответственно, получаем три задачи математического программирования, с ограничениями на сумму весов либо:

$$\left\{ \begin{array}{l} A_t = \sum_{i=1}^p w_t^i \cdot B_t^i; \\ w_t^i \geq 0; \\ \sum_{i=1}^p w_t^i = 1; \end{array} \right. ,$$

либо:

$$\left\{ \begin{array}{l} A_t = \sum_{i=1}^p w_t^i \cdot B_t^i; \\ w_t^i \geq 0; \\ \sum_{i=1}^p w_t^i = 1, \quad i \in I_{sym}; \\ \sum_{i=1}^p w_t^i = 1, \quad i \in I_{nsym}. \end{array} \right.$$

Либо без ограничений на сумму весов:

$$\left\{ \begin{array}{l} A_t = \sum_{i=1}^p w_t^i \cdot B_t^i; \\ w_t^i \geq 0 \end{array} \right. .$$

Для нахождения весов модели используется метод наименьших модулей (МНМ).

$$\sum_{s=1}^{t-1} F(A_s, y_s) \longrightarrow \min_{w_t^1, \dots, w_t^p} ,$$

где $A_s = \sum_{i=1}^p w_t^i \cdot B_s^i$.

Настройка метода наименьших модулей с одним ограничением – неравенством вида $w_t^i \geq 0$ сводится к задаче линейного программирования, которую можно решить с помощью симплекс-метода. Введем дополнительные переменные $\eta_s^+ = a \left(\sum_{i=1}^p w_t^i \cdot B_s^i - y_s \right)_+$,

$\eta_s^- = b \left(y_s - \sum_{i=1}^p w_t^i \cdot B_s^i \right)_-$. Тогда исходная задача примет следующий вид:

$$\left\{ \begin{array}{l} \sum_{s=1}^{t-1} (\eta_s^+ + \eta_s^-) \rightarrow \min_{\eta_s^+, \eta_s^-, w_t^i}; \\ b \sum_{i=1}^p w_t^i B_s^i \geq b y_s - \eta_s^-, \quad s = 1, \dots, t-1; \\ a \sum_{i=1}^p w_t^i B_s^i \leq a y_s + \eta_s^+, \quad s = 1, \dots, t-1; \\ \eta_s^+, \eta_s^-, w_t^i \geq 0, \quad i = 1, \dots, p. \end{array} \right.$$

Такая задача с помощью простейших преобразований сводится к стандартной задаче линейного программирования:

$$\left\{ \begin{array}{l} c^T x \rightarrow \min; \\ Ax \leq b; \\ A_{eq} x = b_{eq}; \\ lb \leq x \leq ub. \end{array} \right.$$

для которой существуют в системе MATLAB встроенные функции.

Настройка метода наименьших модулей с ограничениями вида $w_t^i \geq 0$ и $\sum_{i=1}^p w_t^i = 1$ или $w_t^i \geq 0$ и $\sum_{i=1}^p w_t^i = 1$ при $i \in I_{sym}$, $\sum_{i=1}^p w_t^i = 1$ при $i \in I_{nsym}$ сводится к задаче линейного программирования аналогичным образом.

2.4.2 Метод локальной адаптации весов с регуляризацией.

Данный метод полностью идентичен симметричной функции потерь, кроме формулы для вычисления весов. В случае несимметричной функции потерь веса вычисляются по следующей формуле:

$$F(A_{t-1}, y_{t-1}) + \lambda \sum_{i=1}^p (w_t^i - w_{t-1}^i)^2 \longrightarrow \min_{w_t^1, \dots, w_t^p},$$

где $A_{t-1} = \sum_{i=1}^p w_t^i \cdot B_{t-1}^i$.

Настройка метода локальной адаптации весов с регуляризацией с одним ограничением неравенства вида $w_t^i \geq 0$ сводится к задаче квадратичного программирования, для решения которой в системе MATLAB существуют встроенные функции. Введем дополнительные переменные $\eta^+ = a \left(\sum_{i=1}^p w_t^i \cdot B_{t-1}^i - y_{t-1} \right)_+$,

$\eta^- = b \left(y_{t-1} - \sum_{i=1}^p w_t^i \cdot B_{t-1}^i \right)_-$. Тогда исходная задача примет следующий вид:

$$\left\{ \begin{array}{l} \eta^+ + \eta^- + \lambda(w_t^i - w_{t-1}^i)^2 \rightarrow \min_{\eta^+, \eta^-, w_t^i}; \\ b \sum_{i=1}^p w_t^i B_{t-1}^i \geq by_{t-1} - \eta^-; \\ a \sum_{i=1}^p w_t^i B_{t-1}^i \leq ay_{t-1} + \eta^+; \\ \eta^+, \eta^-, w_t^i \geq 0, \quad i = 1, \dots, p. \end{array} \right.$$

Такая задача с помощью простейших преобразований сводится к стандартной задаче квадратичного программирования:

$$\left\{ \begin{array}{l} \frac{1}{2} x^T Hx + f^T x \rightarrow \min; \\ Ax \leq b; \\ A_{eq} x = b_{eq}; \\ lb \leq x \leq ub. \end{array} \right. ,$$

для которой существуют в системе MATLAB встроенные функции.

Настройка метода наименьших модулей с ограничениями вида $w_t^i \geq 0$ и $\sum_{i=1}^p w_t^i = 1$ или $w_t^i \geq 0$ и $\sum_{i=1}^p w_t^i = 1$ при $i \in I_{sym}$, $\sum_{i=1}^p w_t^i = 1$ при $i \in I_{nsym}$ сводится к задаче квадратичного программирования аналогичным образом.

Очевидно, что время работы такого метода значительно меньше времени работы метода наименьших модулей. В полученной задаче квадратичного программирования число переменных $(p+3)$ и число ограничений $(p+5)$ зависят только от числа базовых алгоритмов p .

2.5 Различные варианты решения задачи.

Рассмотрим различные варианты решения задачи прогнозирования при несимметричной функции потерь:

1) Точечный прогноз \hat{y}_t настраивается по симметричной функции потерь, переводится в прогноз плотности p_t , потери считаются с помощью свертки прогноза плотности с несимметричной функцией потерь.

2) Точечный прогноз \hat{y}_t настраивается по несимметричной функции потерь, потери считаются по несимметричной функции потерь.

3) Точечные прогнозы B_t^i настраиваются по симметричной функции потерь и подаются на вход композиции, корректирующая композиция, настроенная по симметричной функции потерь преобразует точечные прогнозы B_t^i в точечный прогноз \hat{y}_t , точечный прогноз \hat{y}_t переводится в прогноз плотности p_t , потери считаются с помощью свертки прогноза плотности с несимметричной функцией потерь (рис.2.6).

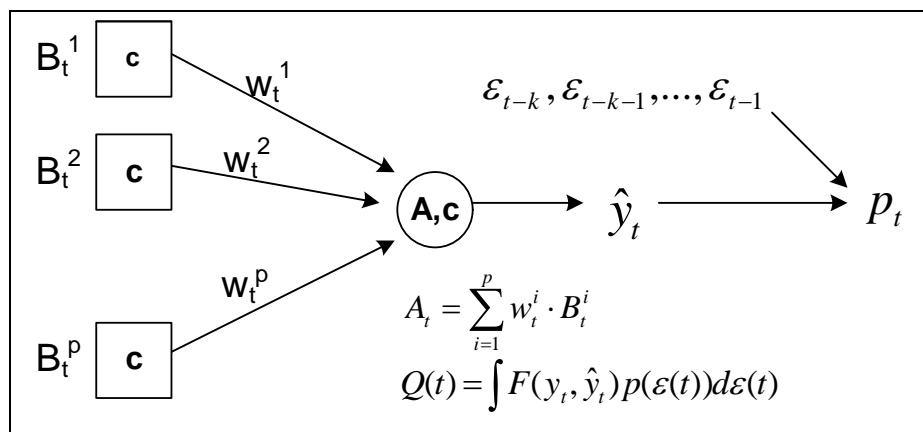


Рис. 2.6. Первый вариант построения адаптивных композиций алгоритмов прогнозирования при несимметричной функции потерь.

4) Точечные прогнозы B_t^i настраиваются по несимметричной функции потерь и подаются на вход композиции, корректирующая операция, настроенная по несимметричной функции потерь, преобразует точечные прогнозы B_t^i в точечный прогноз \hat{y}_t , потери считаются по несимметричной функции потерь (рис.2.7).

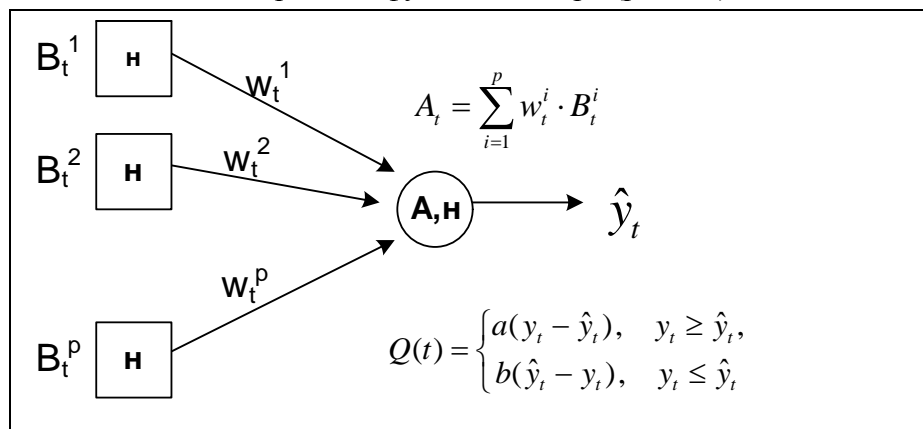


Рис. 2.7. Второй вариант построения адаптивных композиций алгоритмов прогнозирования при несимметричной функции потерь.

5) Прогнозы плотности p_t^i подаются на вход композиции, корректирующая композиция преобразует прогнозы плотностей p_t^i в прогноз плотности p_t и считаются потери по свертке прогноза плотности с несимметричной функцией потерь (рис.2.8).

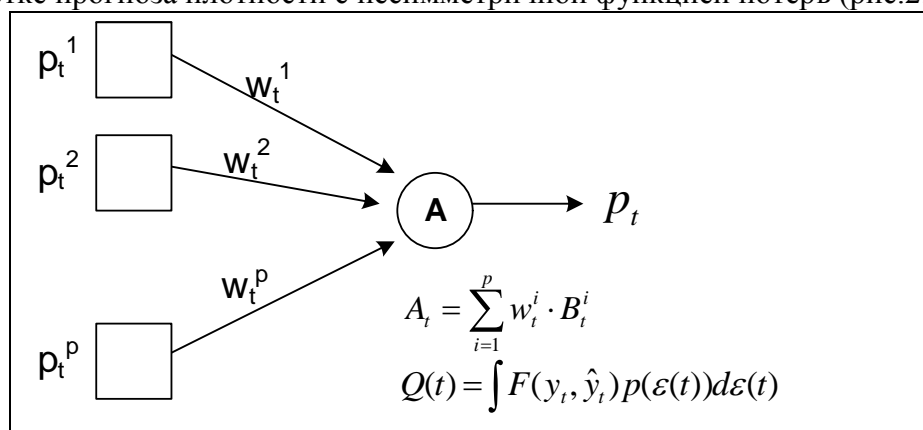


Рис. 2.8. Третий вариант построения адаптивных композиций алгоритмов прогнозирования при несимметричной функции потерь.

В данной работе рассматриваются первые три типа композиций, проводятся вычислительные эксперименты. Задача построения композиций алгоритмов, в случае подачи на вход корректирующей операции прогнозов плотности представляет задачу будущих исследований. Задача построения композиций алгоритмов для случая, когда на вход корректирующей операции подаются точечные прогнозы, настроенные по несимметричной функции потерь была рассмотрена в более ранних работах.

3 Об одном эффективном методе построения прогноза плотности.

В данной работе предлагается новый способ построения прогноза плотности, основанный на оценивании плотности распределения ошибок по скользящему контролю. Также был проведен эксперимент, целью которого было подтвердить возможность построения прогноза плотности с помощью стандартных методов точечного прогнозирования.

3.1 Построение прогноза плотности.

Предлагаемый метод состоит в следующем: за основу берется любой стандартный метод точечного прогнозирования при квадратичной функции потерь. С помощью стандартных методов прогнозирования в каждый момент времени получаем точечный прогноз \hat{y}_t на следующий промежуток времени. Далее строится временной ряд ошибок точечных прогнозов в предыдущие моменты времени в режиме скользящего среднего: $\varepsilon_{t-k}, \varepsilon_{t-k-1}, \dots, \varepsilon_{t-1}$, затем по нему строится вариационный ряд ошибок $\varepsilon_1^{(k)}, \varepsilon_2^{(k)}, \dots, \varepsilon_k^{(k)}$ и эмпирическая плотность распределения ошибок. Для получения прогноза плотности в заданный момент времени эмпирическая плотность распределения ошибок сдвигается на величину точечного прогноза \hat{y}_t в данный момент времени: $\varepsilon_1^{(k)} + \hat{y}_t, \varepsilon_2^{(k)} + \hat{y}_t, \dots, \varepsilon_k^{(k)} + \hat{y}_t$. Аналитически функция распределения прогноза в заданный момент времени запишется следующим образом:

$$\begin{cases} P(y_t \leq \varepsilon_1^{(k)} + \hat{y}_t) = 0, \\ P(\varepsilon_i^{(k)} + \hat{y}_t \leq y_t \leq \varepsilon_{i+1}^{(k)} + \hat{y}_t) = \frac{i}{k}, \\ P(\varepsilon_k^{(k)} + \hat{y}_t \leq y_t) = 1 \end{cases} \quad (3.1)$$

Вычисление величины потерь в данный момент времени выполняется с помощью свертки эмпирической плотности с несимметричной функцией потерь.

Лемма 2. Пусть $p(\varepsilon_t)$ — прогноз плотности в момент времени t , $\varepsilon_t = y_t - \hat{y}_t$. Тогда потери в момент времени t вычисляются по формуле:

$$Q(t) = \int_{-\infty}^{+\infty} F(\varepsilon_t) p(\varepsilon_t) d\varepsilon_t = \frac{1}{k} \sum_{i=1}^k F(\varepsilon_i^{(k)})$$

Доказательство:

Воспользуемся формулой интегрирования по частям и тем что, производная от функции распределения равна плотности распределения, получим:

$$Q(t) = \int_{-\infty}^{+\infty} F(\varepsilon_t) p(\varepsilon_t) d\varepsilon_t = \int_{-\infty}^{+\infty} F(\varepsilon_t) dP(\varepsilon_t) = F(\varepsilon_t) P(\varepsilon_t) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} P(\varepsilon_t) F'(\varepsilon_t) d(\varepsilon_t) \Rightarrow$$

$$Q(t) = F_{\max} \cdot 1 - F_{\min} \cdot 0 - \int_{-\infty}^{+\infty} P(\varepsilon_t) F'(\varepsilon_t) d(\varepsilon_t) = F_{\max} - \int_{-\infty}^{+\infty} P(\varepsilon_t) F'(\varepsilon_t) d(\varepsilon_t).$$

Используя формулу (3.1) для подсчета функции распределения, получим:

$$Q(t) = F_{\max} - \left(\int_{-\infty}^{\varepsilon_1^{(k)}} 0 \cdot F'(\varepsilon_t) d(\varepsilon_t) + \int_{\varepsilon_1^{(k)}}^{\varepsilon_2^{(k)}} \frac{1}{k} F'(\varepsilon_t) d(\varepsilon_t) + \int_{\varepsilon_2^{(k)}}^{\varepsilon_3^{(k)}} \frac{2}{k} F'(\varepsilon_t) d(\varepsilon_t) + \dots + \int_{\varepsilon_{k-1}^{(k)}}^{\varepsilon_k^{(k)}} \frac{k-1}{k} F'(\varepsilon_t) d(\varepsilon_t) \right) \Rightarrow$$

$$Q(t) = F_{\max} - \left(\frac{1}{k} (F(\varepsilon_2^{(k)}) - F(\varepsilon_1^{(k)})) + \frac{2}{k} (F(\varepsilon_3^{(k)}) - F(\varepsilon_2^{(k)})) + \dots + \frac{k-1}{k} (F_{\max} - F(\varepsilon_k^{(k)})) \right) \Rightarrow$$

$$Q(t) = \frac{1}{k} (F(\varepsilon_1^{(k)}) + F(\varepsilon_2^{(k)}) + \dots + F(\varepsilon_k^{(k)})), \text{ что и требовалось доказать.}$$

3.2 Эксперимент.

Проводился эксперимент, целью которого было показать, что для прогнозирования плотности можно воспользоваться предложенным выше способом. В эксперименте рассматривалось два метода для построения композиции алгоритмов прогнозирования:

Метод 1. Базовые алгоритмы выдают точечные прогнозы, к этим прогнозам применяется корректирующая операция, преобразующая набор точечных прогнозов в точечный прогноз. Потери считаются при помощи стандартной квадратичной функции потерь.

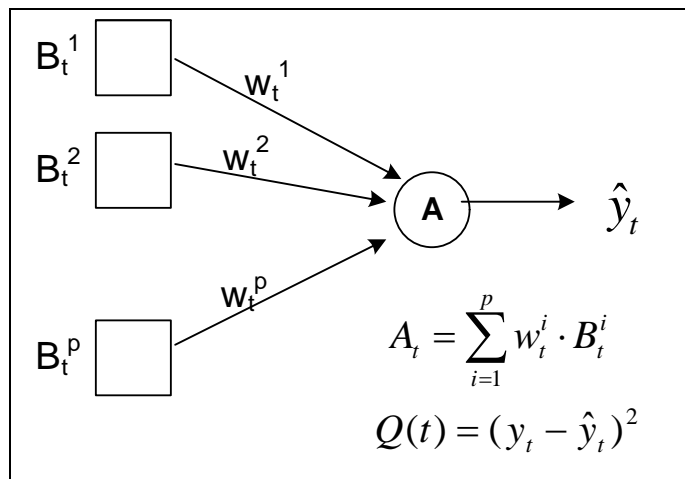


Рис. 3.2 Схема метода 1 данного эксперимента.

Метод 2. Точечные прогнозы на выходе корректирующей операции преобразуются в прогноз плотности. Потери считаются при помощи свертки прогноза плотности с несимметричной функцией потерь.

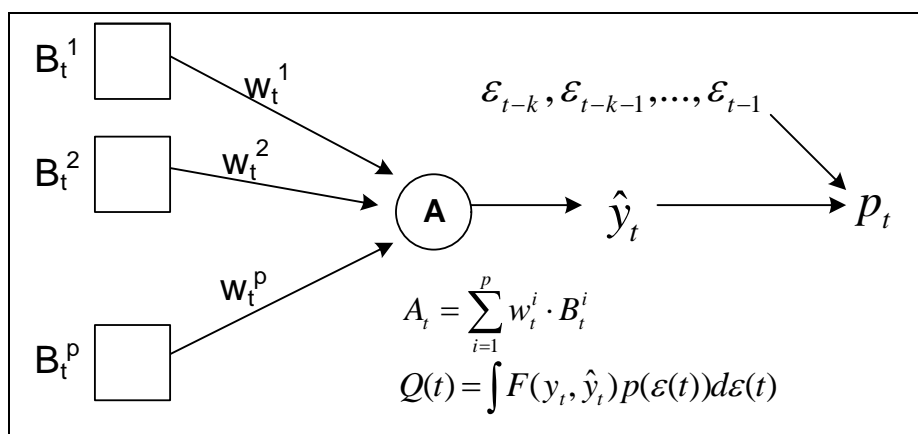


Рис. 3.3 Схема метода 2 данного эксперимента.

Целью эксперименты было показать возможность использования описанного выше метода. Для этого необходимо в каждом из подходов отранжировать базовые алгоритмы и композиции алгоритмов. Если в обоих методах алгоритмы отранжируются одинаково, то для построения прогноза плотности можно воспользоваться стандартными методами точечного прогнозирования.

В качестве базовых алгоритмов использовались алгоритмы экспоненциального сглаживания с учетом факторов сезонности и тренда:

Базовый алгоритм 1. Экспоненциальное сглаживание без учета сезонности и тренда.

Базовый алгоритм 2. Экспоненциальное сглаживание с учетом экспоненциального роста.

Базовый алгоритм 3. Экспоненциальное сглаживание с учетом аддитивного роста.

Базовый алгоритм 4. Экспоненциальное сглаживание с учетом мультипликативного сезонного эффекта.

Базовый алгоритм 5. Экспоненциальное сглаживание с учетом аддитивного сезонного эффекта.

Композиции базовых алгоритмов, дающих точечные прогнозы, строились следующими способами:

- 1) Выбор наилучшего алгоритма среди базовых (Выбор лучшего);
- 2) Простое усреднение среди базовых алгоритмов (Среднее);
- 3) Метод наименьших квадратов с ограничением неотрицательности весов (МНК М);
- 4) Метод наименьших квадратов без ограничения неотрицательности весов (МНК неМ);
- 5) Метод наименьших квадратов с регуляризацией с ограничением неотрицательности весов (МНК-рег М);
- 6) Метод наименьших квадратов с регуляризацией без ограничения неотрицательности весов (МНК-рег неМ);
- 7) Локальная адаптация весов с регуляризацией с ограничением неотрицательности весов (ЛАВР М);
- 8) Локальная адаптация весов с регуляризацией без ограничения неотрицательности весов (ЛАВР неМ).

3.3 Результаты вычислительных экспериментов.

Вычислительный эксперимент был проведен на реальных временных рядах объемов продаж в супермаркетах. Всего использовалось около 40 временных рядов, в каждом из которых строилось примерно по 600 прогнозов в каждый момент времени. В таблице 1 (Приложение) показаны результаты сравнения методов для базовых алгоритмов. В каждой ячейке показано какое место занимает соответствующий алгоритм для соответствующего временного ряда по величине ошибки.

В таблице 2 (Приложение) показаны результаты сравнения методов для композиций алгоритмов.

Из таблицы 2 (Приложение) видно, что для обоих методов базовые алгоритмы ранжируются, за редким исключением, одинаково. Композиции алгоритмов для обоих методов в большинстве случаев ранжируются также одинаково. Отличия вызваны тем, что ошибки для композиций алгоритмов в этих случаях различаются совсем незначительно. Таким образом, для прогнозирования плотности не обязательно строить какие-то специальные методы (например, основанные на решении трудных оптимизационных задач с неквадратичным функционалом потерь), а можно воспользоваться стандартными методами точечного прогнозирования.

Таблица 3.1 Сумма рангов базовых алгоритмов для сравниваемых методов для всех рядов.

Метод	Базовый 1	Базовый 2	Базовый 3	Базовый 4	Базовый 5
Метод 1	64	161	121	101	108
Метод 2	69	149	127	98	112

Таблица 3.2 Сумма рангов композиций алгоритмов для сравниваемых методов для всех рядов.

Метод	Среднее	Выбор лучшего	МНК М	МНК неМ	МНК-рег М	МНК-рег неМ	ЛАВР М	ЛАВР неМ
Метод 1	211	202	88	115	162	198	170	185
Метод 2	209	199	80	113	157	202	176	194

Кроме этого, результаты эксперимента показали, что композиции алгоритмов дают меньшие потери, чем базовые алгоритмы. А наилучшим среди композиций является метод наименьших квадратов с ограничением неотрицательности весов (Таблица 3.2).

4 Сравнение метода квантильной регрессии и метода, основанного на оценивании плотности распределения по скользящему контролю

В данном разделе приводится сравнение двух методов построения прогнозов при несимметричной функции потерь:

- 1) Стандартного метода – метода квантильной регрессии;
- 2) Метода, основанного на оценивании плотностей распределения ошибок по скользящему контролю.

Функция потерь имеет вид, приведенный на Рис.1.2 и задается следующей формулой:

$$F(y_t, \hat{y}_t) = \begin{cases} a(y_t - \hat{y}_t), & y_t \geq \hat{y}_t, \\ b(\hat{y}_t - y_t), & y_t \leq \hat{y}_t. \end{cases}$$

Цель данного эксперимента – сравнить методы прогнозирования плотности и квантильной регрессии по скорости и качеству прогнозирования. Показать, что предлагаемый метод построения прогнозов плотности является гораздо более вычислительно эффективным, чем метод квантильной регрессии, а также возможность его использования наряду с методом квантильной регрессии для построения прогнозов при несимметричной функции потерь..

Для чистоты эксперимента для каждого метода использовалась одна и та же авторегрессионная модель. Рассматривались следующие модели:

- 1) Авторегрессионная модель без учета весовых коэффициентов: $Q(t) = \sum_{i=1}^t F(y_i, \hat{y}_i)$;
- 2) Авторегрессионная модель с учетом весовых коэффициентов: $Q(t) = \sum_{i=1}^t w^{t-i} F(y_i, \hat{y}_i)$.

Приведем способы вычисления прогнозов для каждого метода для каждой авторегрессионной модели.

4.1 Вычисление прогнозов для авторегрессионной модели без учета весовых коэффициентов.

4.1.1 Метод квантильной регрессии.

Определение прогноза в момент времени t является решением следующей задачи:

$$\left(\sum_{i: y_i \geq \hat{y}_{t+1}} \theta |y_i - \hat{y}_{t+1}| + \sum_{i: y_i \leq \hat{y}_{t+1}} (1-\theta) |y_i - \hat{y}_{t+1}| \right) \rightarrow \min_{\hat{y}_{t+1}}, \text{ где}$$

y_i - реальное значение в момент времени t ,

$$\theta = a/(a+b).$$

Такая задача с помощью следующих замен:

$$u_i^+ = \begin{cases} y_i - \hat{y}_{t+1}, & y_i \geq \hat{y}_{t+1} \\ 0, & y_i \leq \hat{y}_{t+1} \end{cases} \text{ и } u_i^+ = \begin{cases} 0, & y_i \geq \hat{y}_{t+1} \\ \hat{y}_{t+1} - y_i, & y_i \leq \hat{y}_{t+1} \end{cases}$$

легко сводится к следующей задаче линейного программирования:

$$\begin{cases} \theta 1' u^+ + (1-\theta) 1' u^- \rightarrow \min_{u^+, u^-, \hat{y}_{t+1}} \\ u^+ 1' - u^- 1' + \hat{y}_{t+1} = y, \\ u^+, u^- \geq 0 \end{cases},$$

для которой существуют встроенные в MATLAB методы решения.

4.1.2 Метод, основанный на оценивании плотности распределения по скользящему контролю.

Определение точечного прогноза в момент времени $t+1$ является решением следующей задачи:

$$\sum_{i=1}^t (y_i - \hat{y}_{t+1})^2 \rightarrow \min_{\hat{y}_{t+1}}.$$

Очевидно, что точечный прогноз равен среднему арифметическому всех предыдущих значений ряда:

$$\hat{y}_{t+1} = \frac{1}{t} \sum_{i=1}^t y_i.$$

Далее, точечный прогноз \hat{y}_{t+1} с помощью операций, описанных в п.3.1 преобразовывается в прогноз плотности.

4.2 Вычисление прогнозов для авторегрессионной модели с учетом весовых коэффициентов.

4.2.1 Метод квантильной регрессии.

Определение прогноза в момент времени t является решением следующей задачи:

$$\left(\sum_{i: y_i \geq \hat{y}_{t+1}} w^{t-i} \theta |y_i - \hat{y}_{t+1}| + \sum_{i: y_i \leq \hat{y}_{t+1}} w^{t-i} (1-\theta) |y_i - \hat{y}_{t+1}| \right) \rightarrow \min_{\hat{y}_{t+1}}, \text{ где}$$

y_i - реальное значение в момент времени t ,

$\theta = a/(a+b)$.

Такая задача с помощью следующих замен:

$$u_i^+ = \begin{cases} y_i - \hat{y}_{t+1}, & y_i \geq \hat{y}_{t+1} \\ 0, & y_i \leq \hat{y}_{t+1} \end{cases} \quad \text{и} \quad u_i^- = \begin{cases} 0, & y_i \geq \hat{y}_{t+1} \\ \hat{y}_{t+1} - y_i, & y_i \leq \hat{y}_{t+1} \end{cases}$$

легко сводится к следующей задаче линейного программирования:

$$\begin{cases} \theta w^{t-i} u^+ + (1-\theta) w^{t-i} u^- \rightarrow \min_{u^+, u^-, \hat{y}_{t+1}} \\ u^+ 1' - u^- 1' + \hat{y}_{t+1} = y, \\ u^+, u^- \geq 0 \end{cases},$$

для которой существуют встроенные в MATLAB методы решения.

4.2.2 Метод, основанный на оценивании плотности распределения по скользящему контролю.

Определение точечного прогноза в момент времени $t+1$ является решением следующей задачи:

$$\sum_{i=1}^t w^{t-i} (y_i - \hat{y}_{t+1})^2 \rightarrow \min_{\hat{y}_{t+1}}.$$

По лемме 1, точечный прогноз вычисляется с помощью формулы экспоненциального сглаживания:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t, \quad \alpha = 1-w, \quad w \in (0,1)$$

Далее, точечный прогноз \hat{y}_{t+1} с помощью операций, описанных в п.3.1 преобразовывается в прогноз плотности.

4.3 Результаты вычислительных экспериментов.

Вычислительный эксперимент был проведен на реальных временных рядах объемов продаж в супермаркетах. Всего использовалось около 40 временных рядов по продажам различных продовольственных товаров. В каждом ряду было примерно 800 значений. Для каждого ряда строилось от 400 до 600 прогнозов. Описанные методы сравнивались по скорости работы и по средней величине потерь.

Для метода квантильной регрессии потери в каждый момент времени считались по следующей формуле:

$$F(y_t, \hat{y}_t) = \begin{cases} a(y_t - \hat{y}_t), & y_t \geq \hat{y}_t, \\ b(\hat{y}_t - y_t), & y_t \leq \hat{y}_t. \end{cases}$$

В случае вычисления прогнозов по методу, основанному на оценивании плотности распределения по скользящему контролю, потери, согласно лемме 2, вычислялись следующим образом:

$$F(y_t, \hat{y}_t) = \frac{1}{k} \sum_{i=1}^k F(\varepsilon_i^{(k)}).$$

В таблице 3 (Приложение) приводится сравнение величины потерь рассматриваемых методов для авторегрессионной функции без учета весовых коэффициентов для каждого временного ряда. Рассматривается различный наклон несимметричной функции потерь: 1 к 1, 1 к 5, 5 к 1. Результаты показали, что случая симметричной функции потерь предлагаемый метод прогнозирования плотности в 20% случаев дает лучшие результаты. В остальных случаях, потери превышают потери в методе квантильной регрессии не более чем на 30%. В случае увеличения наклона несимметричной функции потерь (например, 5 к 1 или 1 к 5), потери превышают потери в методе квантильной регрессии примерно в 2 раза. Это вызвано влиянием «хвостов» полученных распределений. Для улучшения качества необходимо сглаживать эти «хвосты». Что касается скорости работы, метод квантильной регрессии является вычислительно неэффективным. Эксперимент показал, что обработка одного временного ряда в случае прогнозирования плотности прогноза занимала до 30 секунд, в то время, как для работы метода квантильной регрессии затрачивалось 15-20 минут (в несколько десятков раз больше, чем для предложенного метода).

На рисунке 1 (Приложение) приводятся графики изменения средней величины потерь со временем. Графики приведены для различных временных рядов и для различных коэффициентов наклона несимметричной функции потерь. Красным цветом обозначены потери для метода прогнозирования плотности, синим цветом – для метода квантильной регрессии.

В таблице 4 (Приложение) приводится сравнение величины потерь рассматриваемых методов для авторегрессионной функции с учетом весовых коэффициентов для каждого временного ряда. Рассматриваются различные весовые коэффициенты по арифметической прогрессии: от 0.5 до 0.95 с шагом 0.05. Также рассматривается различный наклон функции потерь: 1 к 1, 5 к 1, 1 к 5.

На рисунке 2 (Приложение) приведена зависимость величины средних потерь от весового коэффициента. Как видно из графиков, практически для всех случаев имеется ярко выраженный минимум функции потерь. Эксперимент еще раз подтверждает вычислительную неэффективность метода квантильной регрессии. Обработка временных рядов для предложенного в работе нового метода занимала на несколько порядков меньше времени.

4.4 Преимущества и недостатки метода прогнозирования плотности и метода квантильной регрессии

Проведенная работа позволяет сформулировать основные преимущества и недостатки рассматриваемых методов, а также основные направления дальнейших работ по улучшению качества и скорости прогнозирования.

Метод прогнозирования плотности.

Преимущества:

1) Вычислительно эффективен, даже на больших временных рядах (порядка 1000 значений) и большой предысторией;

2) Используется для произвольной функции потерь, а не только для несимметричной L_1 -нормы;

3) Если потери от недопрогноза не сильно отличаются от потерь от перепрогноза, данный метод дает результаты, сравнимые по величине потерь с методом квантильной регрессии;

4) В качестве прогноза выдается не точечное значение, а некоторая функция распределения, которая может быть преобразована в функцию плотности прогноза и несет в себе гораздо больше информации чем просто точечный прогноз.

Недостатки:

1) Большая величина потерь, если потери от недопрогноза и от перепрогноза отличаются в несколько раз.

Дальнейшие работы: увеличение качества прогнозирования за счет сглаживания «хвостов» получаемых функций распределения.

Метод квантильной регрессии.

Преимущества:

1) Высокое качество прогнозирования.

Недостатки:

1) Используется лишь для несимметричной L_1 -нормы. Нельзя использовать для произвольной функции потерь;

2) Очень низкая скорость прогнозирования.

Дальнейшие работы: увеличение скорости прогнозирования за счет более глубокого изучения метода линейного прогнозирования и замены стандартного симплекс-метода на метод, который позволял бы решать такие задачи прогнозирования гораздо быстрее.

5 Заключение

В данной работе рассматривалась задача построения адаптивных композиций алгоритмов прогнозирования при несимметричной функции потерь.

Данное исследование связано с решением актуальной прикладной задачи прогнозирования объемов продаж в сети супермаркетов. Специфика задачи заключается в том, что число рядов может исчислять сотнями тысяч, поэтому вычислять прогнозы требуется быстро. Кроме этого, в задаче прогнозирования потребительского спроса несимметричная функция потерь более адекватно оценивает величину потерь, чем стандартная квадратичная функция потерь.

В работе исследованы методы построения адаптивных композиций над адаптивными базовыми алгоритмами прогнозирования.

Предлагается метод построения прогноза плотности, основанный на оценивании плотности распределения ошибок по скользящему контролю. Также был проведен эксперимент, подтверждающий возможность построения прогноза плотности с помощью стандартных методов точечного прогнозирования.

В работе приводится экспериментальное сравнение (на реальных данных объемов продаж) двух методов прогнозирования при несимметричной функции потерь: квантильной регрессии и метода прогнозирования плотности. Результаты экспериментов показали, что метод прогнозирования плотности является вычислительно эффективным и позволяет строить прогнозы по времени на несколько порядков быстрее, чем метод квантильной регрессии. К тому же, метод прогнозирования плотности может применяться не только для несимметричной L_1 -нормы, но и для любой произвольной функции потерь. Главным недостатком метода прогнозирования плотности является низкое качество прогнозирования, в случае если величины потерь при недопрогнозе и перепрогнозе отличаются в несколько раз.

Основными направлениями дальнейших работ является улучшение качества построения прогноза плотности и увеличение скорости прогнозирования в методе квантильной регрессии. Улучшение качества построения прогноза плотности может быть улучшено при помощи сглаживания «хвостов» получаемых функций распределения. Для увеличения скорости построения прогнозов в методе квантильной регрессии необходимо предложить более вычислительно эффективные методы решения задачи линейного программирования, чем стандартный симплекс-метод.

В работе также представлен общий обзор задачи прогнозирования потребительского спроса. Рассмотрены различные методы решения задач прогнозирования: нейронные сети, векторная авторегрессия, адаптивные модели, модели ARIMA, GARCH и другие методы.

Приложения

Таблица 1. Ранги базовых алгоритмов для сравниваемых методов для каждого ряда.

№ ряда	Метод	Базовый 1	Базовый 2	Базовый 3	Базовый 4	Базовый 5
2934	Метод 1	1	2	3	4	5
	Метод 2	1	2	3	4	5
1488	Метод 1	1	5	4	3	2
	Метод 2	1	5	4	3	2
1604	Метод 1	2	5	4	3	1
	Метод 2	3	5	4	2	1
2988	Метод 1	1	4	2	5	3
	Метод 2	1	4	3	5	2
3314	Метод 1	3	2	4	1	5
	Метод 2	3	2	4	1	5
3367	Метод 1	1	5	4	3	2
	Метод 2	2	5	4	3	1
3488	Метод 1	1	5	3	2	4
	Метод 2	1	5	3	2	4
3760	Метод 1	1	3	2	5	4
	Метод 2	1	2	3	5	4
4122	Метод 1	2	5	4	3	1
	Метод 2	2	5	4	3	1
4176	Метод 1	3	5	4	1	2
	Метод 2	3	5	4	1	2
4730	Метод 1	3	5	2	1	4
	Метод 2	3	5	2	1	4
4929	Метод 1	1	3	4	5	2
	Метод 2	1	2	4	5	3
4936	Метод 1	3	5	4	1	2
	Метод 2	2	5	4	1	3
5184	Метод 1	1	5	3	2	4
	Метод 2	1	4	3	2	5
5620	Метод 1	1	3	2	5	4
	Метод 2	1	3	2	5	4
1742	Метод 1	2	5	1	3	4
	Метод 2	2	5	1	3	4
2425	Метод 1	1	5	4	3	2
	Метод 2	1	5	4	3	2
2934	Метод 1	1	2	3	4	5
	Метод 2	1	2	3	4	5
6304	Метод 1	1	3	2	4	5
	Метод 2	1	3	2	5	4
6305	Метод 1	1	5	4	3	2
	Метод 2	1	5	4	3	2
6310	Метод 1	2	5	4	1	3
	Метод 2	2	5	4	1	3
6314	Метод 1	1	5	4	2	3
	Метод 2	1	5	4	2	3
449	Метод 1	3	5	4	2	1
	Метод 2	3	5	4	2	1
6353	Метод 1	3	5	4	2	1

	Метод 2	3	5	4	1	2
6631	Метод 1	1	4	2	3	5
	Метод 2	1	4	2	3	5
7954	Метод 1	2	4	1	5	3
	Метод 2	3	1	2	5	4
8492	Метод 1	2	5	4	3	1
	Метод 2	2	3	5	4	1
8937	Метод 1	1	3	2	4	5
	Метод 2	1	2	3	4	5
9128	Метод 1	3	5	4	1	2
	Метод 2	3	5	4	1	2
10190	Метод 1	2	5	4	3	1
	Метод 2	3	5	4	2	1
10898	Метод 1	1	5	2	3	4
	Метод 2	1	3	2	4	5
10899	Метод 1	1	5	3	2	4
	Метод 2	1	5	3	2	4
12543	Метод 1	2	3	4	1	5
	Метод 2	2	3	4	1	5
12597	Метод 1	2	5	4	1	3
	Метод 2	2	4	5	1	3
12602	Метод 1	3	5	4	1	2
	Метод 2	3	5	4	2	1
12966	Метод 1	2	5	4	3	1
	Метод 2	3	5	4	1	2
13188	Метод 1	2	5	4	3	1
	Метод 2	3	5	4	1	2

Таблица 2. Ранги композиций алгоритмов для сравниваемых методов для каждого ряда.

№ ряда	Метод	Среднее	Выбор лучшего	МНК М	МНК неМ	МНК-рег М	МНК-рег неМ	ЛАВР М	ЛАВР неМ
2934	Метод 1	8	2	3	1	4	5	6	7
	Метод 2	8	2	3	1	4	5	6	7
1488	Метод 1	7	4	1	3	5	2	6	8
	Метод 2	7	5	1	3	4	2	6	8
1604	Метод 1	8	7	2	1	6	5	4	3
	Метод 2	8	5	1	2	7	6	4	3
2988	Метод 1	1	3	6	8	7	4	5	2
	Метод 2	1	5	3	6	7	4	8	2
3314	Метод 1	2	8	3	4	1	5	6	7
	Метод 2	1	8	2	3	4	6	5	7
3367	Метод 1	7	8	2	3	1	6	4	5
	Метод 2	5	8	2	3	1	4	6	7
3488	Метод 1	8	3	1	2	4	7	5	6
	Метод 2	8	3	1	2	4	7	5	6
3760	Метод 1	8	1	2	3	4	7	6	5
	Метод 2	8	1	2	3	4	6	7	5
4122	Метод 1	6	3	1	2	5	4	7	8
	Метод 2	5	3	1	2	4	6	7	8
4176	Метод 1	4	8	2	1	6	7	5	3
	Метод 2	6	8	2	1	5	7	4	3
4730	Метод 1	4	7	2	3	1	5	6	8
	Метод 2	6	8	2	5	1	4	3	7
4929	Метод 1	3	7	2	6	5	8	1	4
	Метод 2	5	7	4	6	3	8	1	2
4936	Метод 1	7	5	2	3	1	6	4	8
	Метод 2	7	5	1	2	3	6	4	8
5184	Метод 1	4	1	2	3	5	8	6	7
	Метод 2	2	1	3	4	5	8	6	7
5620	Метод 1	7	1	2	3	4	5	6	8
	Метод 2	7	1	2	3	4	5	6	8
1742	Метод 1	7	8	5	6	2	4	1	3
	Метод 2	4	8	1	2	7	5	3	6
2425	Метод 1	8	6	1	2	7	5	3	4
	Метод 2	8	3	1	2	7	4	5	6
2934	Метод 1	8	2	3	1	4	5	6	7
	Метод 2	8	2	3	1	4	5	6	7
6304	Метод 1	7	6	1	3	2	8	5	4
	Метод 2	7	1	2	3	4	5	8	6
6305	Метод 1	7	3	2	5	6	8	4	1
	Метод 2	7	4	2	6	5	8	3	1
6310	Метод 1	8	4	2	1	7	6	5	3
	Метод 2	7	3	1	2	6	5	4	8
6314	Метод 1	8	4	2	1	3	6	7	5
	Метод 2	8	3	2	1	4	6	7	5
449	Метод 1	7	8	4	6	2	5	3	1
	Метод 2	6	8	5	7	1	4	2	3
6353	Метод 1	7	8	1	2	5	6	3	4
	Метод 2	7	8	1	2	3	6	4	5
6631	Метод 1	4	8	2	1	6	3	7	5

№ ряда	Метод	Среднее	Выбор лучшего	МНК М	МНК неМ	МНК-рег М	МНК-рег неМ	ЛАВР М	ЛАВР неМ
	Метод 2	4	5	2	1	6	3	8	7
7954	Метод 1	2	8	6	7	4,5	4,5	1	3
	Метод 2	2,3	8	6	7	4,5	4,5	1	2,3
8492	Метод 1	5	8	3	4	2	1	6	7
	Метод 2	4	8	3	5	2	1	6	7
8937	Метод 1	2	8	1	3	5	6	7	4
	Метод 2	1	8	2	3	5	6	7	4
9128	Метод 1	6	3	2	1	5	7	4	8
	Метод 2	7	4	2	1	5	8	3	6
10190	Метод 1	8	6	1	2	7	3	5	4
	Метод 2	8	7	1	2	3	4	6	5
10898	Метод 1	4	7	1	2	5	6	3	8
	Метод 2	5	7	2	3	4	8	1	6
10899	Метод 1	3	8	2	1	7	6	4	5
	Метод 2	3	8	2	1	7	6	4	5
12543	Метод 1	6	7	2	1	5	4	8	3
	Метод 2	6	8	2	1	4	5	7	3
12597	Метод 1	4	8	5	7	3	6	2	1
	Метод 2	6	8	4	5	2	7	3	1
12602	Метод 1	5	4	2	3	6	7	1	8
	Метод 2	6	3	1	4	5	8	2	7
12966	Метод 1	3	7	6	8	5	4	2	1
	Метод 2	3	8	4	6	5	7	2	1
13188	Метод 1	8	3	1	2	5	4	6	7
	Метод 2	8	7	1	2	4	3	6	5

Таблица 3. Сравнение методов квантильной регрессии и метода построения плотности для авторегрессионной модели без учета весов. Рассмотрены различные варианты наклона функции потерь.

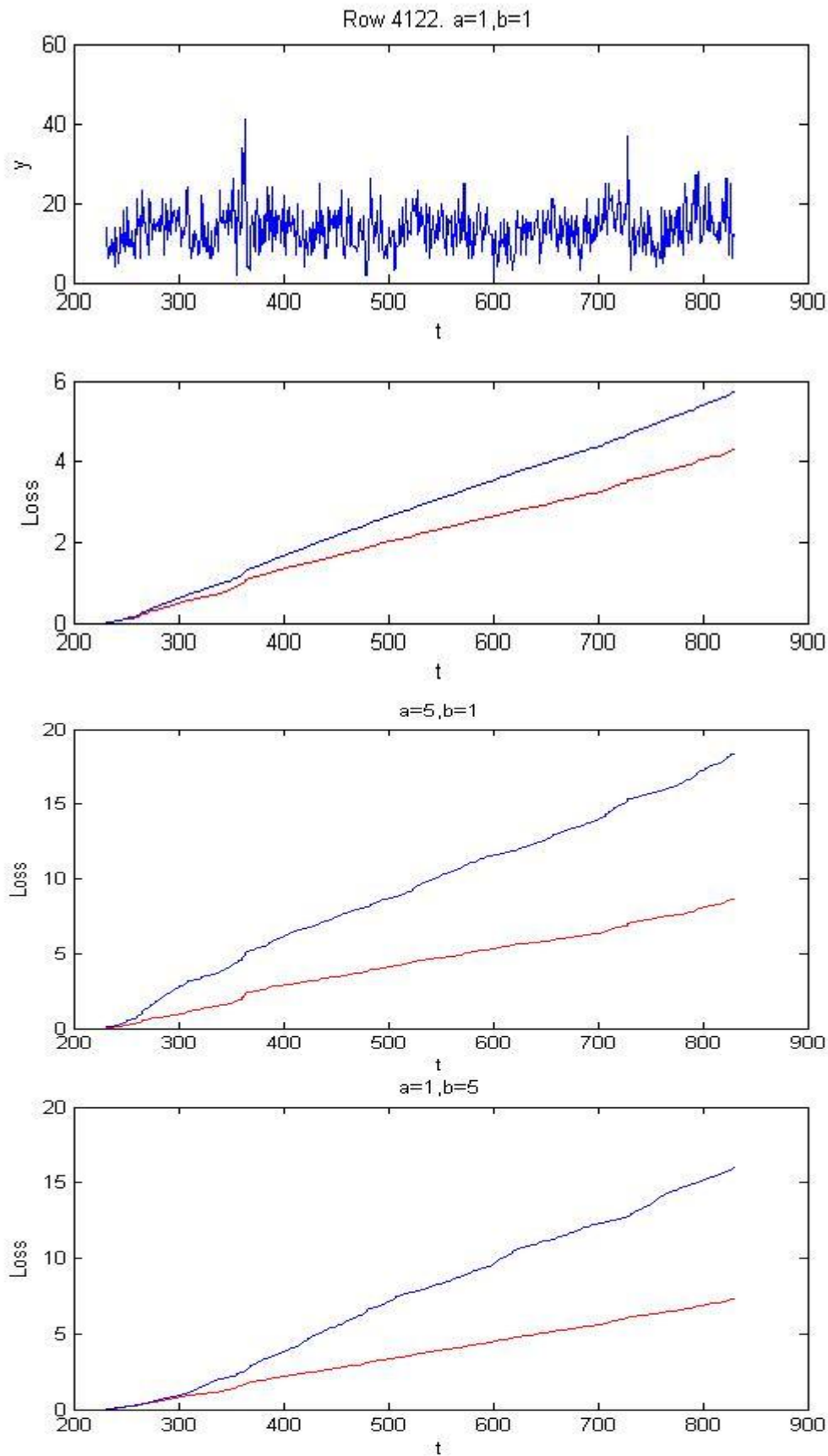
Метод 1. Квантильная регрессия.						
Метод 2. Построения прогнозов путем построения прогнозов плотности.						
	a=1, b=1		a=5, b=1		a=1, b=5	
№ ряда	Потери Метод 1	Потери Метод 2	Потери Метод 1	Потери Метод 2	Потери Метод 1	Потери Метод 2
449	4,44	5,39	10,85	16,72	7,16	18,9
1742	3,7	3,65	8,17	12,14	5,63	9,8
2425	5,61	7,55	11,84	22,68	9,5	22,44
2934	6,96	6,43	15,625	22,217	9,483	16,65
6304	4,76	6,05	10,738	16,705	7,87	19,46
6305	6,59	9,13	14,34	27,34	11,3	27,75
6310	3,73	4,85	7,94	14,28	6,18	14,73
6314	8,1	11,05	17,81	31,53	14,38	36,14
3314	10,04	14,41	25,39	42,58	16,64	42,53
3367	27,3	31,39	61,29	102,13	42,71	86,47
3371	4,47	4,7	10,45	15,93	8,76	12,56
3397	51,3	54,73	136,13	170,32	98,76	153,56
3488	15,32	15,87	31,23	50,54	25,25	44,72
3760	7,76	10,34	16,12	31,42	12,88	31
4122	4,3	5,66	8,638	18,541	7,3	16,127
4176	31,08	21,97	68,54	54,139	46,825	74,81
4730	5,14	7,14	12,75	22,11	8,14	20,74
4929	5,81	8,07	12	27,27	10,58	23,6
4936	8,6	10,86	24,35	33,98	11,52	30,47
5184	5,02	6,81	11,38	20,39	8,1	20,48
5620	11,21	10,35	19,55	36,05	18,34	28,06
6353	3,18	3,9	6,07	12,18	6,3	12,5
6631	5,21	6,16	12,64	19,29	7,82	17,98
7954	8,55	10,34	20,57	33,11	15,41	30,12
8492	15,06	18,74	30,56	59,4	25,46	47,85
8937	6,75	6,37	13,88	21,31	12,21	19,89
9128	5,47	5,21	12,5	17,39	8,41	17,39
10190	3,16	4,34	6,45	12,88	6,02	12,88
10230	4,85	6,71	11,97	20,68	9,54	16,4
10897	6,29	7,84	14,73	23,15	12,16	20,18
10898	10,09	10,82	24,52	28,64	20,76	27,81
10899	9,96	11,37	22,34	34,45	18,65	32,71
11167	6,52	7,64	15,05	23,45	12,5	19,76
12543	4,94	5,93	11,57	17,72	8,14	14,54
12597	4,95	6,14	12,71	19,37	9,06	17,85
12602	4,23	5,18	10,2	16,58	8,76	15,46
12966	4,35	6,32	11,56	19,08	9,15	17,76
13051	19,68	22,79	60,42	72,41	45,53	65,68
13188	6,44	7,23	13,84	20,5	8,74	15,45
13312	6,2	7,6	14,9	22,4	12,32	19,24

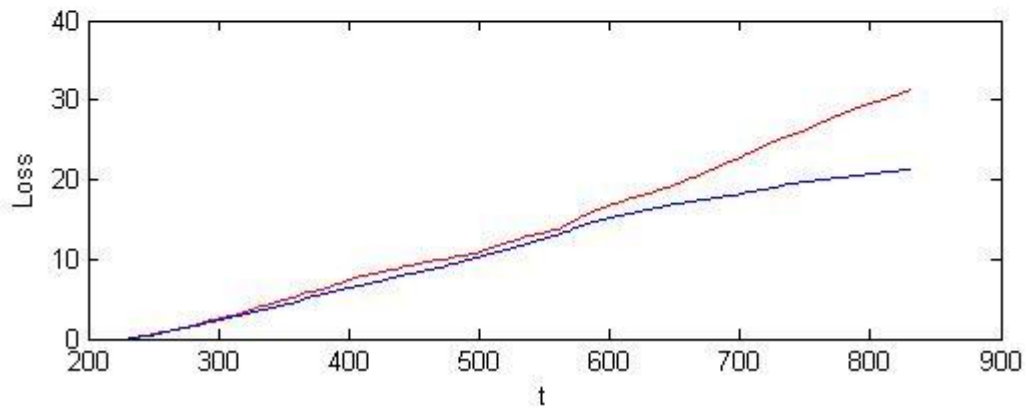
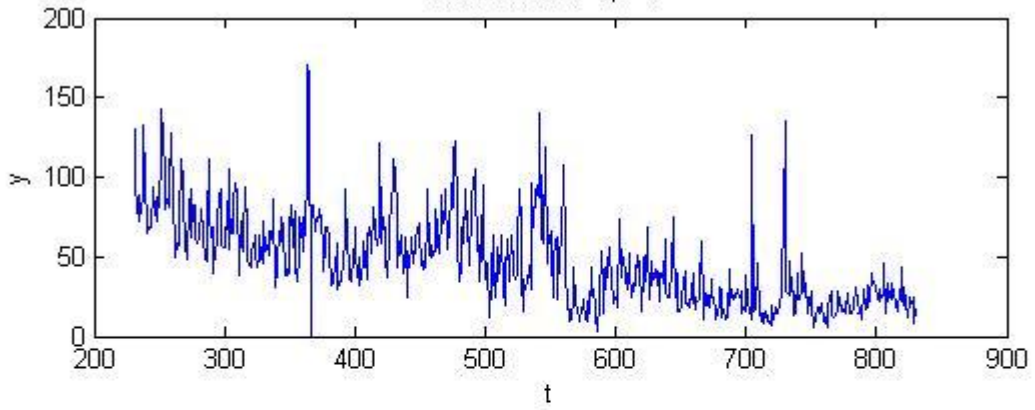
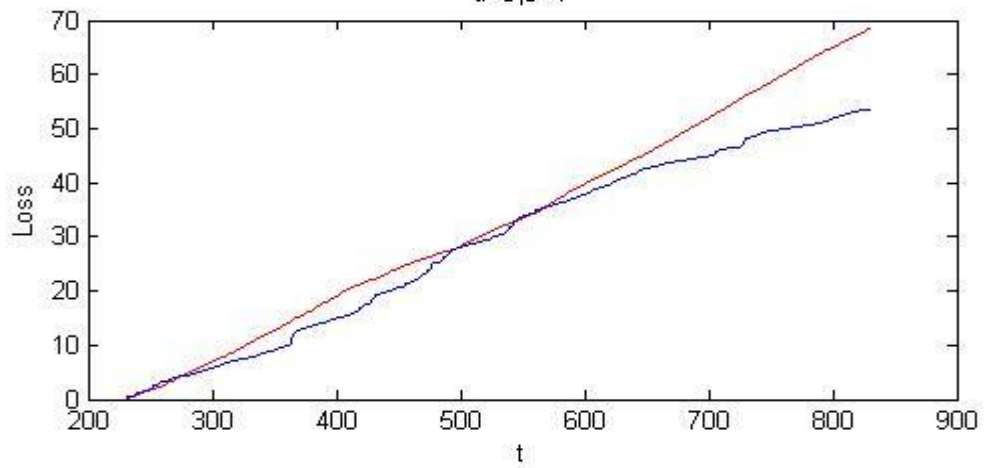
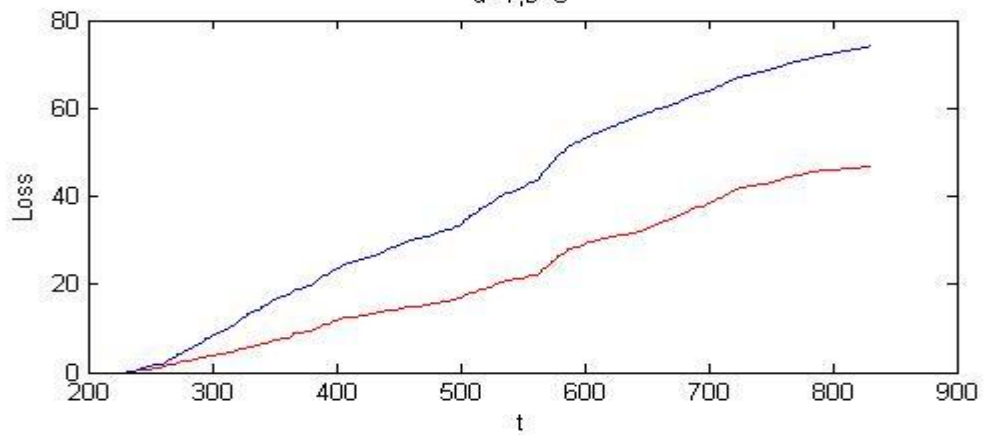
Таблица 4. Сравнение методов квантильной регрессии и метода построения плотности для авторегрессионной модели с учетом весовых коэффициентов. Рассмотрены различные варианты наклона функции потерь.

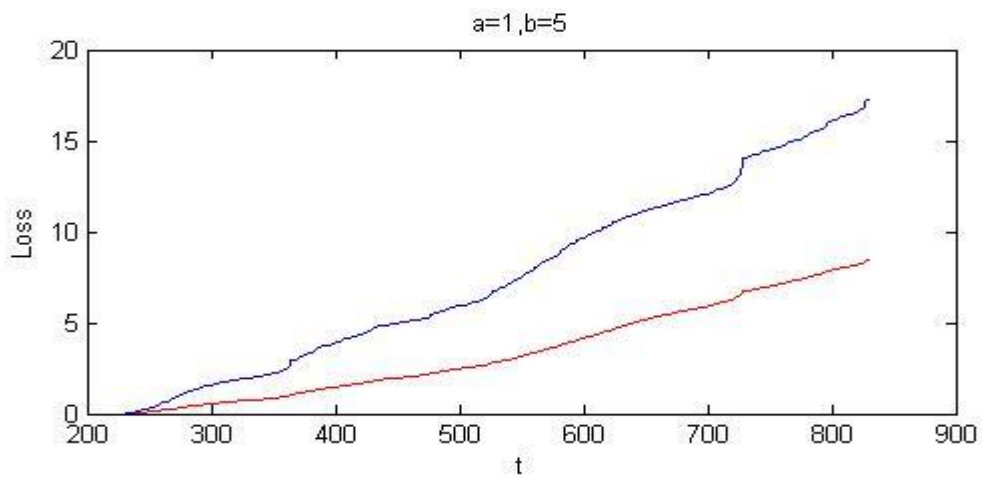
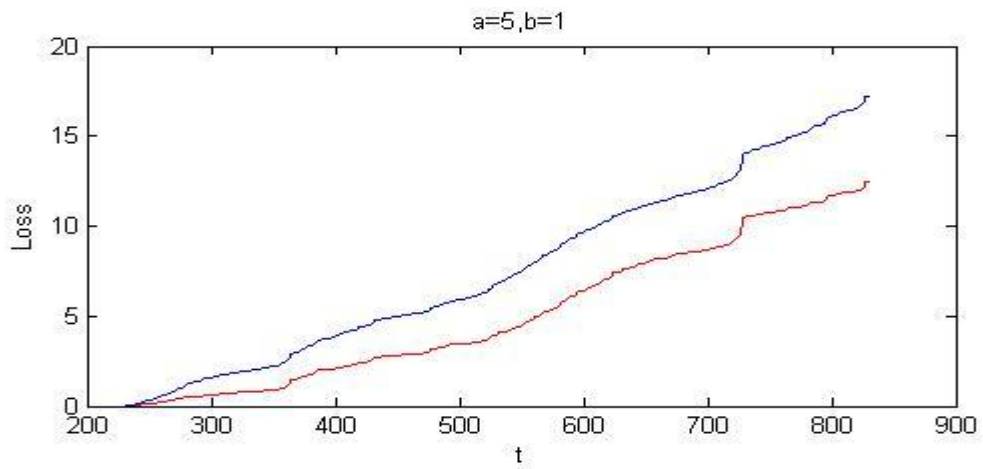
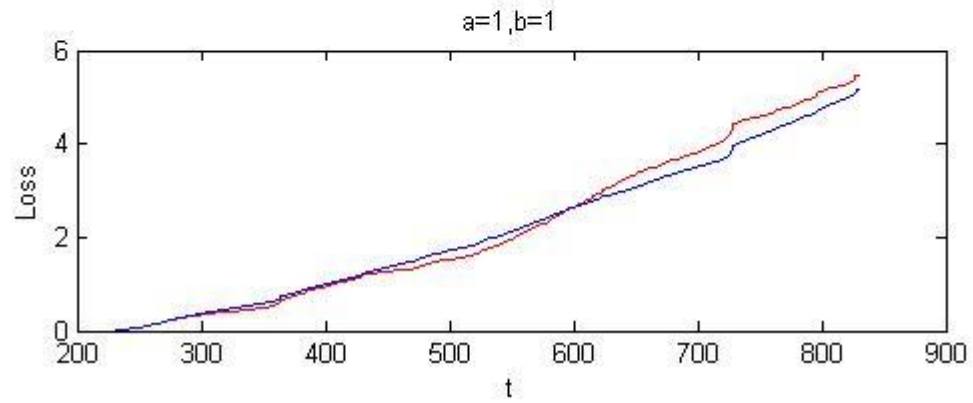
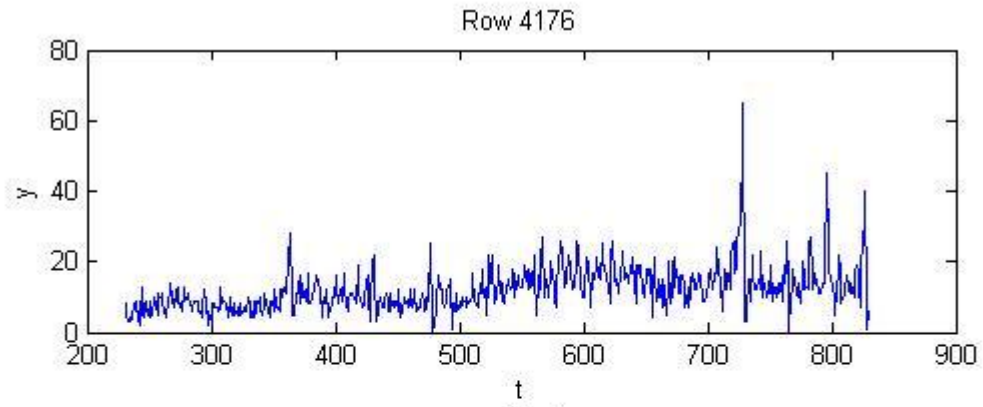
Метод 1. Квантильная регрессия.							
Метод 2. Построения прогнозов путем построения прогнозов плотности.							
		a=1, b=1		a=5, b=1		a=1, b=5	
№ ряда	Весовой коэфф., w	Потери Метод 1	Потери Метод 2	Потери Метод 1	Потери Метод 2	Потери Метод 1	Потери Метод 2
1742	0,5	3,233	3,837	6,751	11,507	5,705	11,517
	0,55	2,872	3,79	6,567	11,366	5,54	11,375
	0,6	2,8	3,745	6,455	11,231	5,155	11,279
	0,65	2,78	3,702	6,321	11,101	4,963	11,11
	0,7	2,765	3,662	6,289	10,98	4,901	10,99
	0,75	2,677	3,624	6,187	10,865	4,737	10,878
	0,8	2,643	3,591	6,045	10,763	4,6	10,786
	0,85	2,582	3,568	5,599	10,681	4,556	10,725
	0,9	2,58	3,559	5,576	10,636	4,532	10,72
	0,95	2,6	3,578	6,437	10,669	4,657	10,789
5620	0,5	5,999	7,181	12,969	21,553	11,644	21,537
	0,55	5,4	7,138	12,82	21,44	11,324	21,391
	0,6	5,39	7,1	12,575	21,348	11,105	21,253
	0,65	5,495	7,079	12,187	21,319	10,52	21,16
	0,7	5,545	7,076	12,112	21,355	10,27	21,103
	0,75	5,435	7,099	11,889	21,494	10,05	21,105
	0,8	5,465	7,146	11,766	21,746	9,495	21,136
	0,85	5,455	7,24	11,004	22,229	9,525	21,211
	0,9	5,765	7,483	12,846	23,396	9,535	21,506
	0,95	6,475	8,366	13,431	27,236	10,365	22,6
2934	0,5	6,879	7,699	15,77	23,088	10,19	23,105
	0,55	5,845	7,55	14,287	22,642	9,74	22,661
	0,6	5,689	7,405	14,109	22,206	9,122	22,226
	0,65	5,66	7,263	13,739	21,779	8,69	21,799
	0,7	5,55	7,123	13,249	21,362	8,256	21,378
	0,75	5,376	6,985	13,008	20,951	8,061	20,960
	0,8	5,246	6,85	12,129	20,553	7,983	20,546
	0,85	4,845	6,72	12,01	20,176	7,701	20,141
	0,9	4,976	6,6	12,145	19,837	7,535	19,739
	0,95	5,096	6,482	12,435	19,596	7,457	19,301
6353	0,5	3,7555	4,135	9,73	12,393	7,31	12,417
	0,55	3,205	4,068	6,549	12,193	6,99	12,218
	0,6	3,165	4,008	6,334	12,013	6,725	12,039
	0,65	3,09	3,952	21,682	11,843	6,415	11,87
	0,7	3,025	3,9	17,854	11,69	6,22	11,72
	0,75	2,9	3,852	26,4	11,544	6,125	11,572
	0,8	2,88	3,8	26,101	11,41	6,035	11,44
	0,85	2,85	3,767	16,288	11,284	5,75	11,32
	0,9	2,84	3,737	7,5	11,194	5,615	11,23
	0,95	2,78	3,735	5,37	11,154	5,68	11,259

Метод 1. Квантильная регрессия.							
Метод 2. Построения прогнозов путем построения прогнозов плотности.							
		a=1, b=1		a=5, b=1		a=1, b=5	
№ ряда	Весовой коэфф., w	Потери Метод 1	Потери Метод 2	Потери Метод 1	Потери Метод 2	Потери Метод 1	Потери Метод 2
6631	0,5	4,7	5,851	13,546	17,872	13,715	17,236
	0,55	4,064	5,781	13,429	17,688	13,546	17
	0,6	4,064	5,717	13,019	17,518	13,425	16,784
	0,65	4,008	5,562	12,865	17,347	12,968	16,567
	0,7	3,97	5,585	12,648	17,167	12,678	16,346
	0,75	3,886	5,509	12,356	16,958	12,365	16,1
	0,8	3,752	5,424	11,978	16,718	12,109	15,828
	0,85	3,723	5,345	11,768	16,49	11,878	15,58
	0,9	3,71	5,286	11,54	16,31	11,657	15,42
	0,95	3,85	5,536	11,34	16,449	11,78	15,75
7954	0,5	6,22	6,9	11,225	20,718	11,83	20,719
	0,55	5,415	6,815	11,237	20,548	11,34	20,432
	0,6	5,32	6,723	11,11	20,2	11,1	20,139
	0,65	5,295	6,633	21,25	19,953	10,725	19,85
	0,7	5,175	6,547	10,32	19,722	10,445	19,563
	0,75	5,055	6,467	10,37	19,518	10,305	19,287
	0,8	4,99	6,393	14,75	19,346	9,79	19,015
	0,85	4,99	6,332	14,66	19,238	9,435	18,754
	0,9	4,985	6,285	10,99	19,242	9,1	18,472
	0,95	4,985	6,283	9,35	19,642	8,91	18,06
8492	0,5	11,209	12,759	22,522	38,044	21,132	38,511
	0,55	10,269	12,599	21,96	37,512	21,392	38,085
	0,6	10,295	12,445	20,66	36,984	19,888	37,69
	0,65	10,11	12,305	20,3	36,476	18,768	37,354
	0,7	10,08	12,179	19,48	35,985	18,6	37,089
	0,75	9,65	12,054	18,63	35,46	18,55	36,869
	0,8	9,6	11,934	18,508	34,88	17,74	36,719
	0,85	9,6	11,828	17,657	34,253	17,46	36,716
	0,9	9,37	11,773	17,555	33,544	17,97	37,086
	0,95	9,47	11,782	17,685	32,363	17,99	38,33
8937	0,5	4,32	4,677	8,773	14,204	7,025	13,86
	0,55	3,825	4,623	8,852	14,052	6,85	13,685
	0,6	3,69	4,578	8,628	13,933	6,502	13,539
	0,65	3,63	4,543	11,696	13,844	6,36	13,416
	0,7	3,625	4,518	8,367	13,796	6,355	13,318
	0,75	3,57	4,509	9,368	13,791	6,125	13,259
	0,8	3,495	4,524	10,098	13,887	6,15	13,258
	0,85	3,495	4,571	7,815	14,102	6,1	13,327
	0,9	3,56	4,683	8	14,547	6,28	13,55
	0,95	3,94	4,94	8,247	15,446	6,645	14,198

Рисунок 1. Графики изменения средней величины потерь со временем для рассматриваемых методов для авторегрессионной модели без учета весовых коэффициентов. Рассмотрены различные варианты наклона функции потерь. Красным цветом показаны потери для метода построения плотности прогноза, синим – для метода квантильной регрессии.



Row 4176. $a=1, b=1$  $a=5, b=1$  $a=1, b=5$ 



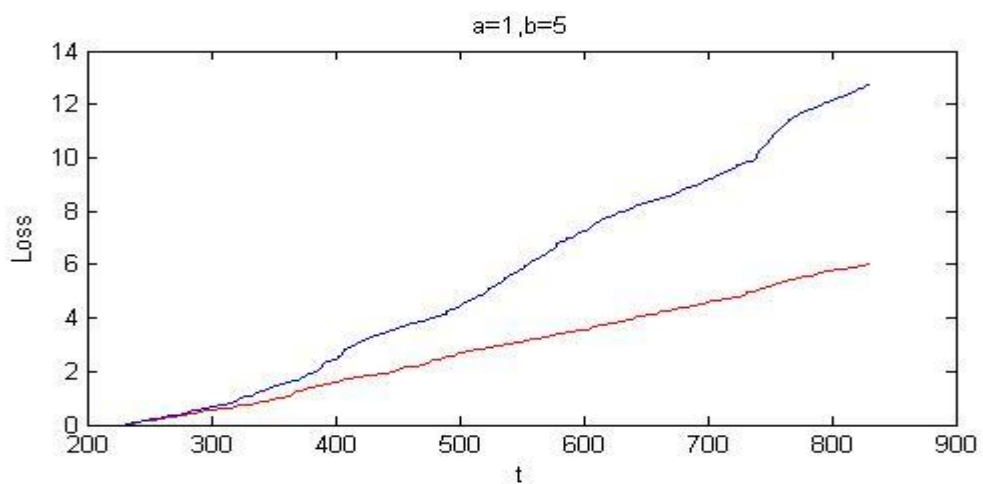
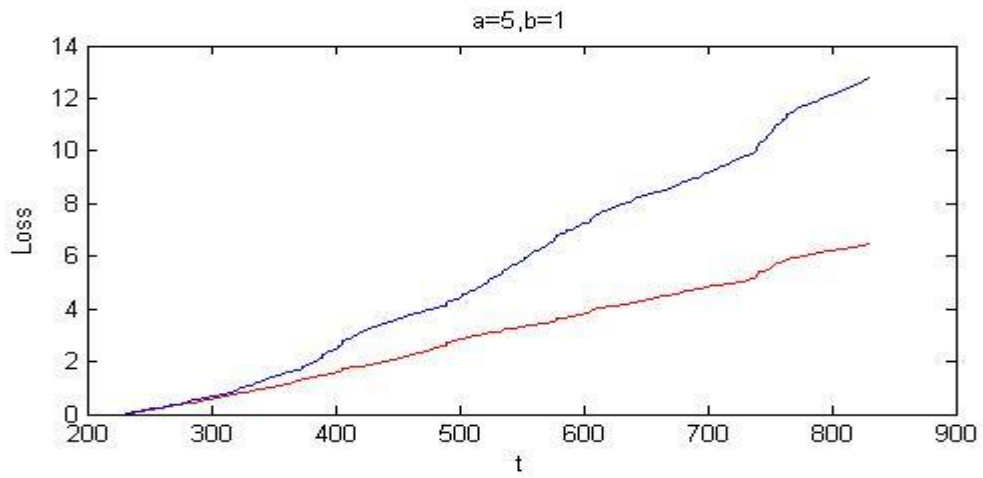
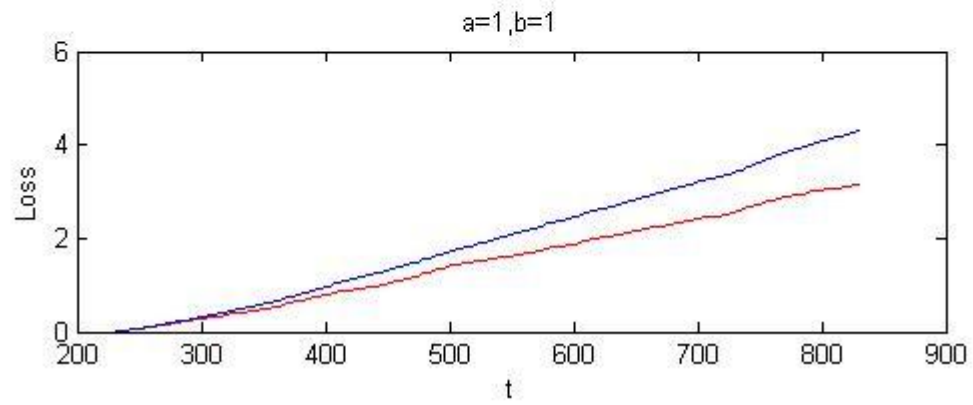
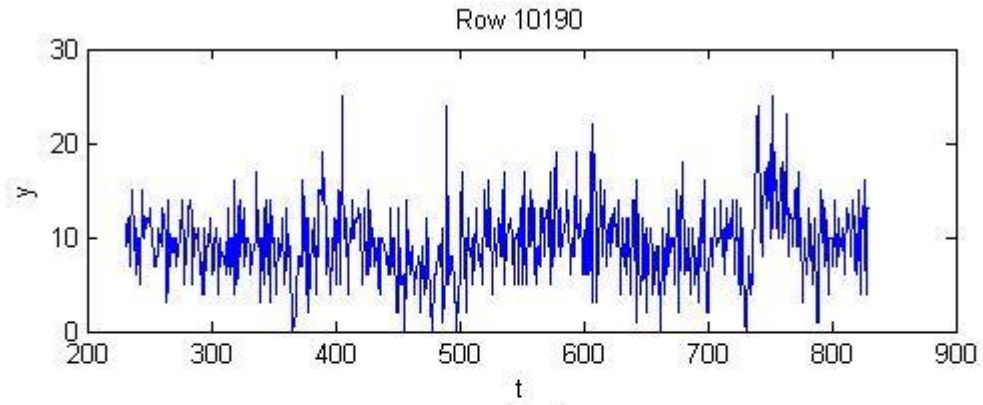


Рисунок 2. Графики зависимости средней величины потерь от весовых коэффициентов для рассматриваемых методов для различных временных рядов. Рассмотрены различные варианты наклона функции потерь. Красным цветом показаны потери для метода построения плотности прогноза, синим – для метода квантильной регрессии.

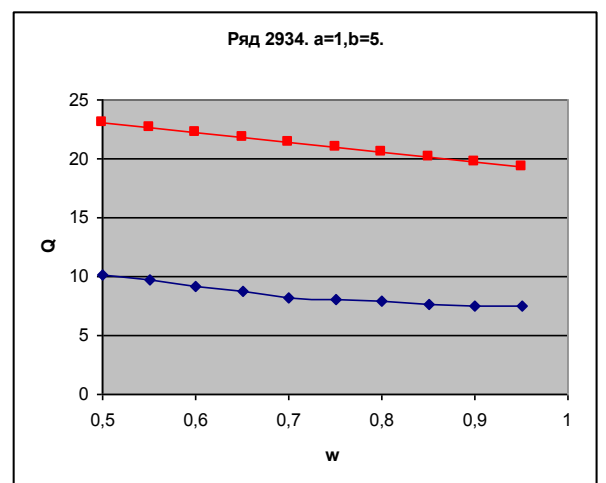
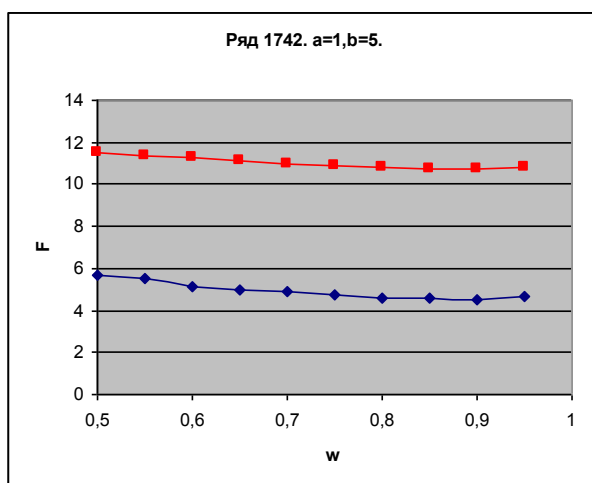
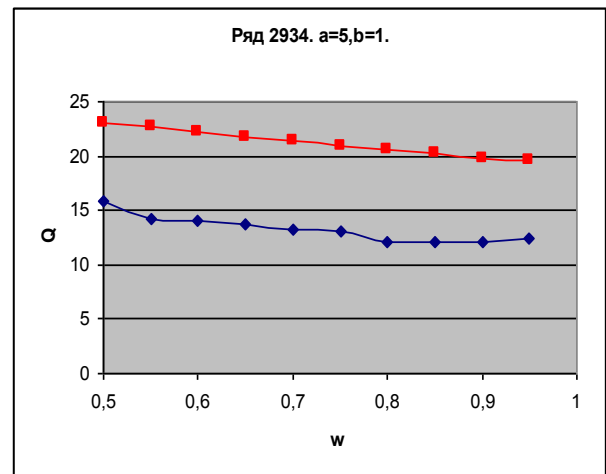
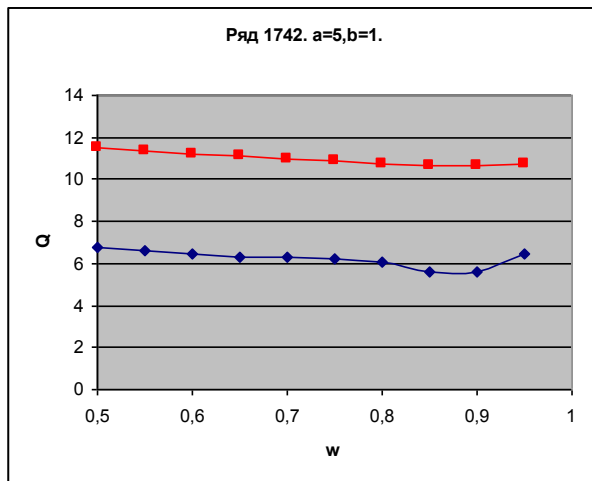
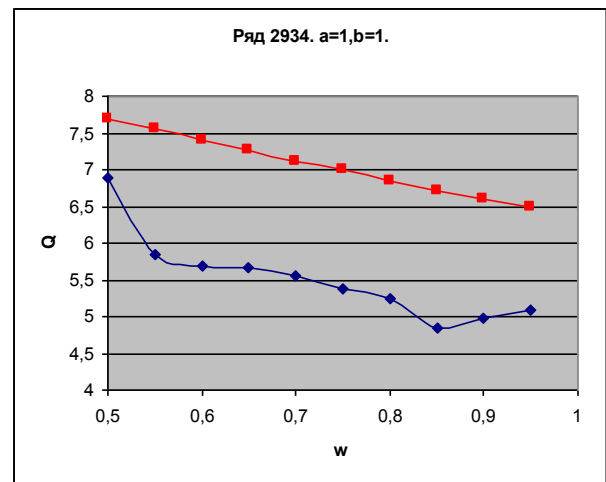
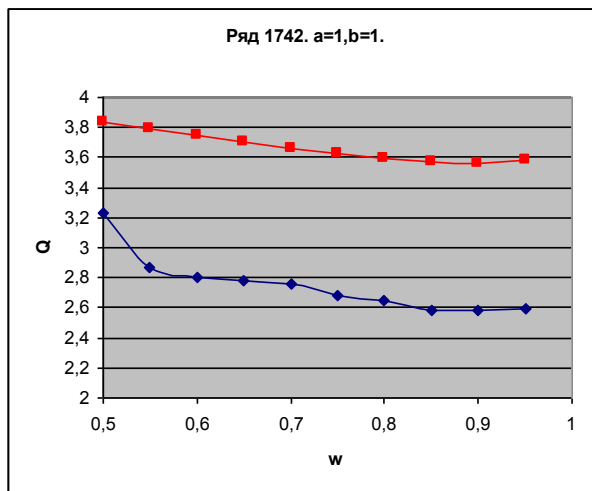
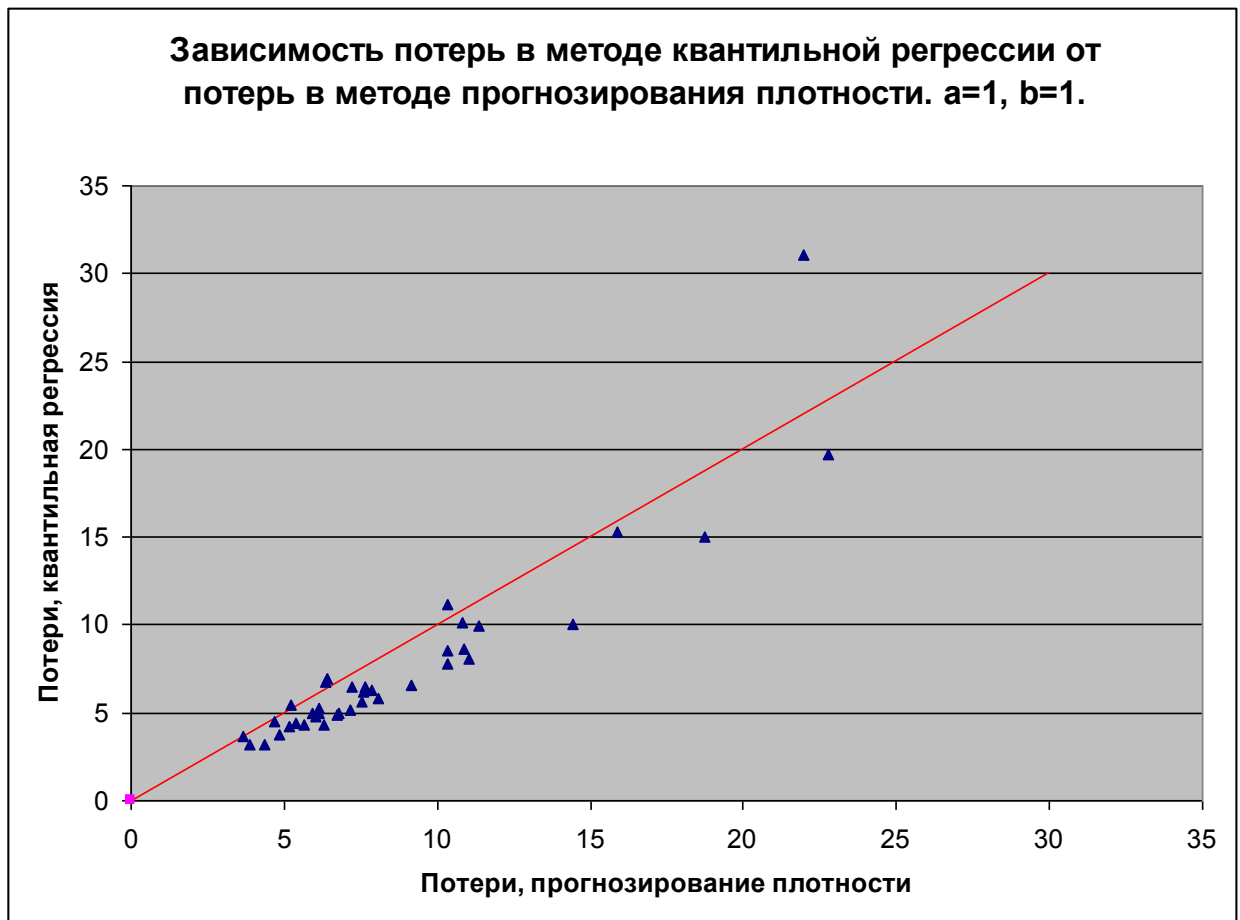
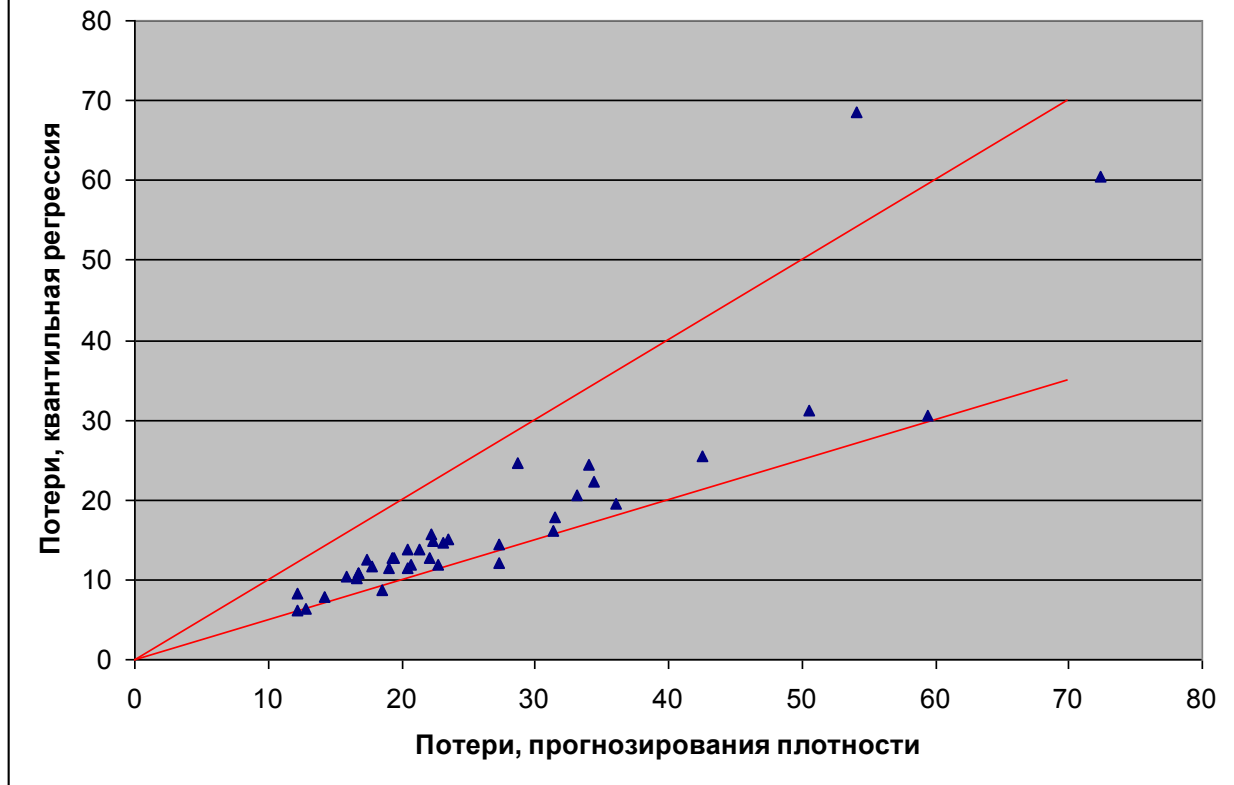


Рисунок 3. Зависимость потерь в методе квантильной регрессии от потерь в методе прогнозирования плотности.



Зависимость потерь в методе квантильной регрессии от потерь в методе прогнозирования плотности. $a=5$, $b=1$.



Список литературы.

- [1] Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики. — 2004. — No. 13. — С. 5–36.
- [2] Журавлёв Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. — 1978. — Т. 33. — С. 5–68.
- [3] Рудаков К. В. Алгебраическая теория универсальных и локальных ограничений для алгоритмов распознавания. — Диссертация на соискание учёной степени д.ф.-м.н., М.: ВЦ РАН. — 1992.
- [4] Постникова Е. (2000) Квантильная регрессия. НГУ.
- [5] Vapnik V. The nature of statistical learning theory. — 2 edition. — Springer-Verlag, New York, 2000.
- [6] Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: Учебное пособие. — М.: Финансы и статистика, 2003.
- [7] Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика // Москва, Дело 2004.
- [8] Peter F. Christoffersen, Francis X. Diebold. Further results on forecasting and model selection under asymmetric loss // Department of Economics University of Pennsylvania.
- [9] Воронцов К. В., Егорова Е. В. Динамически адаптируемые композиции алгоритмов прогнозирования // Вычислительный центр РАН, г.Москва, Россия.
- [10] Tay A. S., Wallis K. F. (2000) Density forecasting: A Survey // Journal of forecasting, №19, p. 235–254.
- [11] Gerard H., Nimark K. Combining multivariate density forecasts using predictive criteria // Reserve Bank of Australia, RDP-2008-02. — 2008.
- [12] Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. — М.: ФИЗМАТЛИТ, 2006.
- [13] Stephen G.Hall, James Mitchell. Density Forecast Combination: Imperial College, London and NIESR, 2004.
- [14] www.offline.computerra.ru/2006/623/248928/.