

# Прогнозирование нестационарных временных рядов при несимметричных функциях потерь\*

*В. Ю. Черных<sup>1</sup>, М. М. Стенина<sup>1,2</sup>*

vladimir.chernykh@phystech.edu, mmedvednikova@gmail.com

<sup>1</sup>Московский физико-технический институт; <sup>2</sup>Высшая школа экономики

Рассматривается задача прогнозирования временных рядов при несимметричных функциях потерь. Предлагается двухэтапный алгоритм прогнозирования  $ARIMA + Hist$ . На первом этапе используется авторегрессионное интегрированное скользящее среднее  $ARIMA$  с сезонной компонентой в случае необходимости. Параметры модели подбираются согласно методологии Бокса-Дженкинса. На втором этапе проводится анализ регрессионных остатков и находится оптимальная добавка к прогнозу, полученному на первом шаге, минимизирующая математическое ожидание потерь. Для оценки ожидаемых потерь используется свертка функции потерь с гистограммой регрессионных остатков. Работа предлагаемого двухэтапного алгоритма иллюстрируется на временных рядах с различными элементами нестационарности (тренд, сезонность) и для различных симметричных и несимметричных функций потерь. Демонстрируется, что качество прогнозов двухэтапного алгоритма превосходит качество прогнозов модели  $ARIMA$  в случае несимметричных функций потерь.

**Ключевые слова:** *прогнозирование, временные ряды, нестационарность, ARIMA, свертка с функцией потерь, несимметричная функция потерь*

## Forecasting nonstationary time series under asymmetric loss\*

*V. Y. Chernykh<sup>1</sup>, M. M. Stenina<sup>1,2</sup>*

<sup>1</sup>Moscow Institute of Physics and Technology; <sup>2</sup>Higher School of Economics

The problem of forecasting time series under asymmetric loss functions is considered in this paper. We present a new two-step forecasting algorithm  $ARIMA + Hist$ . At the first step autoregression integrated moving average algorithm  $ARIMA$  with seasonal components is used. Parameters of the model is selected according to Box-Jenkins methodology. At the second step the analysis of regression residuals is taken place and optimal addition to the forecast of the first step which minimize the expected value of losses is found. Expected loss is estimated by convolution of loss function with histogram of regression residuals. We demonstrate the work of their algorithm on time series with different types of nonstationarity (i.e. trend or seasonality) and for different symmetric and asymmetric loss functions. The results obtained during this experiment show that the quality of the forecast of two-step  $ARIMA + Hist$  exceed the quality of usual  $ARIMA$  in case of asymmetric loss functions.

**Keywords:** *forecasting, time series, nonstationary, ARIMA, convolution with loss function, asymmetric loss*

## 1 Введение

Рассматривается задача прогнозирования нестационарных временных рядов в случае несимметричных функций потерь. Предлагается двухэтапный алгоритм прогнозирования

---

\*Работа выполнена при финансовой поддержке РФФИ, проект № 13-07-13139.

$ARIMA + Hist$ , на первом этапе которого отслеживаются свойства временного ряда, обуславливающие его нестационарность, такие как тренд и сезонность. На втором этапе предлагается находить поправку, обеспечивающую оптимальность прогноза в случае несимметричной функции потерь.

Свойства прогнозов временных рядов при использовании несимметричных функций потерь были исследованы в работе [1], авторы которой отмечают смещенность оптимальных прогнозов при несимметричных потерях и делают вывод о необходимости разработки специальных методов прогнозирования временных рядов в условиях несимметричности функции потерь.

Один из используемых методов прогнозирования нестационарных временных рядов, авторегрессионное интегрированное скользящее среднее  $ARIMA$  [2], позволяет с хорошим качеством прогнозировать временные ряды с трендом, а также при небольшой модификации и ряды с сезонной компонентой. Однако настройка параметров этого алгоритма осуществляется путем минимизации квадратичной функции потерь, но функция потерь, по которой производится оценка качества прогноза, может существенно отличаться от квадратичной. Это приводит к тому, что оптимальный прогноз для модели  $ARIMA$  является несмещенным, а регрессионные остатки должны удовлетворять условиям описанным далее. Ввиду вышесказанного, модель  $ARIMA$  не подходит для решения задачи прогнозирования в случае несимметричной функции потерь, что отмечается в [1, 3].

Авторами работ [4, 5] были предложены модификации модели  $ARIMA$ , позволяющие учесть несимметричность функции потерь при настройке параметров алгоритма. Однако обе предложенные модификации сложны в реализации, не позволяют использовать пакеты для прогнозирования временных рядов, в которых есть стандартные реализации  $ARIMA$ , и требуют для каждой функции потерь создания и обучения индивидуальной модели, что неприемлемо в промышленных задачах. Еще одним методом, предложенным для работы с несимметричными функциями потерь, является квантильная регрессия [6]. Она позволяет находить оптимальный смещенный прогноз для несимметричных функций потерь кусочно-линейного вида, но не дает возможности работать с функциями потерь других видов, а также применима только для стационарных временных рядов.

Предлагаемый алгоритм  $ARIMA + Hist$  использует результат из [7] о том, что при несимметричной функции потерь оптимальный прогноз смещен, причем его смещение зависит только от функции потерь и дисперсии временного ряда. Также используется идея из статьи [8], где автор для построения прогноза использовал авторегрессионную модель с минимизацией квадратичной функции потерь для получения несмещенного прогноза, и анализ регрессионных остатков для оценки оптимального смещения прогноза.

Алгоритм  $ARIMA + Hist$  строит прогноз в два этапа. На первом этапе используется модель  $ARIMA$  с сезонной компонентой в случае необходимости, параметры которой подбираются при помощи анализа временного ряда по методологии Бокса-Дженкинса [2]. На этом этапе получается несмещенный прогноз. На втором этапе производится анализ регрессионных остатков модели  $ARIMA$  с целью оценки оптимального смещения прогноза для минимизации математического ожидания потерь. Оптимальное смещение находится при помощи алгоритма  $Hist$ . Финальный прогноз получается путем прибавления к несмещенному прогнозу, полученному с помощью  $ARIMA$ , найденной при помощи алгоритма  $Hist$  добавки.

Алгоритм  $Hist$  является обобщением алгоритма квантильной регрессии [6]. Он находит приближенное решение задачи минимизации математического ожидания потерь. Такая задача минимизации рассматривалась в работах [9, 10], где математическое ожидание

потерь было представлено как свертка функции потерь с функцией плотности распределения значений временного ряда. На втором этапе алгоритма  $ARIMA + Hist$  в качестве временного ряда выступают регрессионные остатки, однако их плотность распределения не известна. В качестве оценки плотности используется гистограмма значений регрессионных остатков, как предложено в [11]. В алгоритме  $Hist$  используется ряд упрощений задачи минимизации свертки функции потерь с оценкой плотности распределения регрессионных остатков, которые приводят к задаче приближенного нахождения минимума путем перебора конечного числа значений, из которых выбирается то, которое обеспечивает наименьшее значение свертки.

Основное преимущество  $ARIMA + Hist$  состоит в том, что не накладывается ограничений на класс функций потерь, которые можно использовать в задаче прогнозирования.

Алгоритм тестируется на наборе временных рядов, обладающих различными элементами нестационарности. Качество полученных прогнозов сравнивается с качеством прогнозов модели  $ARIMA$  при использовании различных функций потерь. Демонстрируется, что чем более несимметричная будет функция потерь, тем более существенный выигрыш в качестве можно будет получить при помощи двухэтапного алгоритма  $ARIMA + Hist$  по сравнению с  $ARIMA$ .

## 2 Задача прогнозирования временных рядов

Данные представляют собой временной ряд  $\mathbf{x} = \{(x_i)_{i=1}^T \mid x_i \in \mathbb{R}\}$ . Также задается горизонт прогнозирования  $h$ . Ставится задача прогнозирования этого временного ряда при наиболее общем его виде, т.е. нахождении регрессионной модели

$$f : (\mathbf{w}, \mathbf{x}, h) \mapsto \hat{\mathbf{x}}, \quad (1)$$

где  $\mathbf{w}$  — вектор параметров,  $\hat{\mathbf{x}}$  — вектор прогнозов длины  $h$ . В данной работе прогнозирование производится с горизонтом  $h = 1$ , поэтому вектор прогнозов  $\hat{\mathbf{x}}$  является скаляром и обозначается далее как  $\hat{x}_{T+1}$ .

### 2.1 Прогнозирование стационарных временных рядов.

Назовем временной ряд  $\mathbf{x}$  *стационарным*, если все его значения  $x_1, \dots, x_T$  генерируются из одного распределения  $\rho(u)$ , которое не меняется с ходом времени. Из этого определения сразу следует, что рассматриваемый случайный процесс является стационарным в узком смысле, а наш временной ряд является простой выборкой длины  $T$ . Пусть задана функция потерь  $\mathcal{L}(\hat{x}, x)$  и требуется получить прогноз  $\hat{x}_{T+1}$  следующего значения  $x_{T+1}$  временного ряда, минимизируя ожидаемые потери. Предполагается, что следующее значение временного ряда генерируется из того же распределения, что и все предыдущие. При этом задача прогнозирования запишется как

$$\hat{x}_{T+1} = \arg \min_{c \in \mathbb{R}} \mathbf{E} \mathcal{L}(c, x_{T+1}). \quad (2)$$

Если предположить, что плотность распределения  $\rho(u)$ , из которого генерируются значения временного ряда, известна, математическое ожидание потерь запишется как

$$L(c) = \mathbf{E} \mathcal{L}(c, x_{T+1}) = \int_{-\infty}^{+\infty} \mathcal{L}(c, u) \rho(u) du. \quad (3)$$

В таком случае задача прогнозирования формулируется как

$$\hat{x}_{T+1} = \arg \min_{c \in \mathbb{R}} \int_{-\infty}^{+\infty} \mathcal{L}(c, u) \rho(u) du \equiv \arg \min_{c \in \mathbb{R}} L(c). \quad (4)$$

## 2.2 Прогнозирование нестационарных временных рядов.

В случае, когда временной ряд не является стационарным, необходимо оценить и исключить из временного ряда нестационарные особенности прежде чем минимизировать ожидаемые потери в задаче (4). Таким образом, прогноз  $\hat{x}_{T+1}$  нестационарного временного ряда будет складываться из двух частей: прогноз нестационарной компоненты  $\hat{x}_{T+1}^{\text{ns}}$  и прогноз стационарной компоненты  $\hat{x}_{T+1}^{\text{s}}$

$$\hat{x}_{T+1} = \hat{x}_{T+1}^{\text{ns}} + \hat{x}_{T+1}^{\text{s}}. \quad (5)$$

Алгоритм прогнозирования нестационарной компоненты временного ряда должен быть таким, чтобы регрессионные остатки при прогнозе доступной для обучения истории  $\mathbf{x}$

$$\mathbf{r} = \{(r_i)_{i=1}^T \mid r_i = x_i - \hat{x}_i^{\text{ns}}\} \quad (6)$$

были стационарным временным рядом, значения которого сгенерированы из одного распределения с плотностью  $\gamma(u)$ .

В качестве алгоритма прогнозирования нестационарной части ряда предлагается использовать *ARIMA*. Для оптимизации параметров этот алгоритм использует квадратичную функцию потерь  $\mathcal{L}_{\text{sq}}(\hat{x}, x) = (\hat{x} - x)^2$ , по которой строится функционал потерь:

$$\mathcal{Q}(f^{\text{ns}}, \mathbf{x}) = \frac{1}{T} \sum_{i=1}^T \mathcal{L}_{\text{sq}}(f^{\text{ns}}(\mathbf{w}, \mathbf{x}_i, 1), x_{i+1}), \quad (7)$$

$$\mathbf{x}_i = \{x_1 \dots x_i\}.$$

Тогда решение задачи минимизации (7) даёт вектор параметров искомой регрессионной модели

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \mathcal{Q}(f^{\text{ns}}, \mathbf{x}). \quad (8)$$

При этом прогноз вычисляется следующим образом:

$$\hat{x}_{T+1}^{\text{ns}} = f^{\text{ns}}(\mathbf{w}^*, \mathbf{x}, 1). \quad (9)$$

После получения прогноза нестационарной компоненты временного ряда  $\hat{x}_{T+1}^{\text{ns}}$  прогноз стационарной компоненты  $\hat{x}_{T+1}^{\text{s}}$  может быть получен при помощи оценки плотности распределения  $\gamma(u)$  регрессионных остатков  $\mathbf{r}$  и решения для этой плотности задачи минимизации ожидаемых потерь (4).

## 3 Прогнозирование нестационарной компоненты. *ARIMA*. Методология Бокса-Дженкинса

В данном разделе описывается модель авторегрессионного интегрированного скользящего среднего *ARIMA* и методология Бокса-Дженкинса прогнозирования временных

рядов. Принято записывать модель в виде  $ARIMA(p, d, q)$ , где  $p, d, q \in \mathbb{Z}_+$  — структурные параметры, характеризующие порядок для соответствующих частей модели — авторегрессионной, интегрированной и скользящего среднего.  $ARIMA$  с подходящими параметрами для каждого временного ряда предлагается использовать для получения прогноза нестационарной компоненты  $\hat{x}_{T+1}^{ns}$ . Анализ того, насколько хорошо выбранная модель аппроксимирует временной ряд, по методологии Бокса-Дженкинса, включает проверку регрессионных остатков на несмещенность, гомоскедастичность (постоянство дисперсии), нормальность и неавтокоррелированность. Модель считается подходящей для аппроксимации временного ряда, если все эти свойства выполняются для ряда регрессионных остатков, как это описано в [12]. Таким образом, при выборе подходящей модели  $ARIMA$  для прогнозирования нестационарной компоненты временного ряда получается стационарный ряд регрессионных остатков, который можно использовать для построения прогноза  $\hat{x}_{T+1}^s$  стационарной компоненты временного ряда.

### 3.1 Модель ARMA для временного ряда.

Стационарный временной ряд со средним значением  $\mu$  описывается моделью  $ARMA(p, q)$ , если выполняется

$$x_t = \alpha + \varepsilon_t + \sum_{i=1}^q \psi_i \varepsilon_{t-i} + \sum_{i=1}^p \theta_i x_{t-i}, \quad (10)$$

$$\alpha = \mu \left( 1 - \sum_{i=1}^p \theta_i \right).$$

где  $\theta_1, \dots, \theta_p, \psi_1, \dots, \psi_q$  — константы,  $\varepsilon_t$  — гауссов белый шум с нулевым средним и постоянной дисперсией. Вводя оператор сдвига  $L$ , действующий по правилу  $Lx_i = x_{i-1}$ , можно записать модель  $ARMA(p, q)$  в следующем виде:

$$\theta(L)x_t = \alpha + \psi(L)\varepsilon_t, \quad (11)$$

$$\theta(L) = 1 - \sum_{i=1}^p \theta_i L^i, \quad \psi(L) = 1 + \sum_{i=1}^q \psi_i L^i.$$

### 3.2 Модель ARIMA для временного ряда.

Временной ряд описывается моделью  $ARIMA(p, d, q)$ , если ряд его разностей

$$\nabla^d x_t = (1 - L)^d x_t. \quad (12)$$

описывается моделью (11). При этом модель  $ARIMA(p, d, q)$  записывается как

$$\theta(L)\nabla^d x_t = \alpha + \psi(L)\varepsilon_t. \quad (13)$$

### 3.3 Модель ARIMA с мультипликативной сезонностью.

Временной ряд, обладающий мультипликативной сезонностью с периодом  $S$ , описывается моделью  $SARIMA(p, d, q) \times (P, D, Q)_S$ , если

$$\theta_p(L)\Theta_P(L^S)\nabla^d \nabla_S^D x_t = \alpha + \psi_q(L)\Psi_Q(L^S)\varepsilon_t. \quad (14)$$

### 3.4 Методология Бокса-Дженкинса анализа временных рядов.

Используется для оценки параметров модели *ARIMA*. Согласно этой методологии, порядок дифференцирования временного ряда  $d$  выбирается так, чтобы ряд разностей порядка  $d$  был стационарным. Параметры  $p$  и  $q$  выбирают при помощи анализа автокорреляционной и частичной автокорреляционной функций.

**Определение 1.** Автокорреляционная функция  $ACF_\tau$  с лагом автокорреляции  $\tau$  для временного ряда  $\mathbf{x}$  вычисляется по формуле

$$ACF_\tau = \frac{\sum_{i=1}^{T-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x})}{\sum_{i=1}^T (x_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{T} \sum_{i=1}^T x_i.$$

**Определение 2.** Частичная автокорреляционная функция  $PACF_\tau$  с лагом автокорреляции  $\tau$  для стационарного временного ряда  $\mathbf{x}$  вычисляется по формуле

$$PACF_\tau = \begin{cases} \mathbf{E}[x_{t+1}x_t], & \tau = 1; \\ \mathbf{E}[(x_{t+\tau} - x_{t+\tau}^{\tau-1})(x_t - x_t^{\tau-1})], & \tau \geq 2, \end{cases}$$

$$x_t^{\tau-1} = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \dots + \beta_{\tau-1} x_{t+\tau-1},$$

$$x_{t+\tau}^{\tau-1} = \beta_1 x_{t+\tau-1} + \beta_2 x_{t+\tau-2} + \dots + \beta_{\tau-1} x_{t+1},$$

где  $\beta_1, \dots, \beta_{\tau-1}$  — коэффициенты линейной регрессии.

Выбор параметров  $p$  и  $q$  осуществляется из следующих соображений:

- 1) в модели *ARIMA*( $p, d, 0$ ) автокорреляционная функция экспоненциально затухает или имеет синусоидальный вид, а частичная автокорреляционная функция значимо отличается от нуля при лагах, не больших  $p$ ;
- 2) в модели *ARIMA*( $0, d, q$ ) частичная автокорреляционная функция экспоненциально затухает или имеет синусоидальный вид, а автокорреляционная функция значимо отличается от нуля при лагах, не больших  $q$ .

Множество структурных параметров сезонной компоненты *SARIMA* назначается с помощью анализа автокорреляционной и частичной автокорреляционной функций. При наличии сезонной компоненты у временного ряда на графиках этих функций будут наблюдаться характерные максимумы в лагах, соответствующих периоду  $S$  сезонной компоненты.

Необходимые коэффициенты многочленов  $\theta(L), \psi(L)$  оптимизируются при использовании квадратичной функции потерь  $\mathcal{L}_{sq}(\hat{x}, x)$  и функционала потерь (7), основанного на ней.

После оптимизации параметров модели проводится анализ остатков. Регрессионные остатки проверяются на

- 1) несмещенность  $\mathbf{E} r_t = 0$ ,
- 2) стационарность  $\forall t \leftrightarrow r_t \sim \gamma(u)$ ,
- 3) неавтокоррелированность  $\mathbf{E}[r_t r_{t+k}] = 0, k \neq 0$ ,
- 4) нормальность  $r_t \sim N(\mu, \sigma^2)$ ,
- 5) гомоскедастичность  $\mathbf{D} r_t = \sigma_a^2$ .

Если регрессионные остатки обученной модели обладают всеми этими свойствами, то модель признается подходящей для аппроксимации анализируемого временного ряда.

## 4 Прогнозирование стационарной компоненты. Hist. Влияние функции потерь

После обучения модели *ARIMA* с выбранными параметрами даваемый ею прогноз  $\hat{x}_{T+1}^{ns}$  учитывает характерные особенности временного ряда  $\mathbf{x}$ , но не функции потерь  $\mathcal{L}(\hat{x}, x)$ . Пусть ряд регрессионных остатков  $\mathbf{r}$  описывается неизвестной плотностью распределения  $\gamma(u)$ . Предлагается построить добавочный прогноз  $\hat{x}_{T+1}^s$  для стационарного ряда из регрессионных остатков, который минимизирует математическое ожидание потерь (4). Такая добавка к несмещенному прогнозу  $\hat{x}_{T+1}^{ns}$  позволит учесть особенности несимметричной функции потерь.

В случае квадратичной  $\mathcal{L}_{sq}(\hat{x}, x) = (\hat{x} - x)^2$  или абсолютной  $\mathcal{L}_{abs}(\hat{x}, x) = |\hat{x} - x|$  функций потерь добавочный прогноз  $x_{T+1}^s$  можно найти аналитически, не зная при этом конкретного вида распределения  $\gamma(u)$ : для квадратичной функции потерь  $x_{T+1}^s = \mathbf{E} r_t$ , для абсолютной  $x_{T+1}^s = \text{med} \gamma(u)$ , что можно получить продифференцировав  $L(c)$ . В случае же более общего вида функции потерь задача не поддается аналитическому решению без знания конкретного распределения  $\gamma(u)$ , а в практических задачах оно, как правило, неизвестно.

Алгоритм *Hist* предлагает следующий путь для решения этой проблемы. Он состоит из двух приближений функции  $L(c)$ .

### 4.1 Оценка плотности $\gamma(u)$ гистограммой.

Плотность вероятности  $\gamma(u)$  приближается гистограммой значений ряда, т.е. кусочно-постоянной функцией  $\hat{\gamma}(u)$ . Обозначим  $u_{\min} = \min_t r_t$ ,  $u_{\max} = \max_t r_t$ . Они существуют, так как рассматриваются только конечные множества  $\mathbf{r}$ . Произведем разбиение области  $[u_{\min}; u_{\max}]$  на  $n$  отрезков  $[u_i; u_{i+1}]$  равной длины, где

$$u_i = u_{\min} + i \cdot a, \quad a = \frac{u_{\max} - u_{\min}}{n}. \quad (15)$$

На этих отрезках положим значение функции  $\hat{\gamma}(u)$  постоянными и равными  $y_i$  на отрезке  $[u_{i-1}; u_i]$ , где  $y_i$  пропорционально количеству точек ряда  $\mathbf{r}$ , значения которых  $r_t \in [u_{i-1}; u_i]$ .

Точное значение  $y_i$  определяется из условия нормировки функции  $\hat{\gamma}(u)$

$$\int_{u_{\min}}^{u_{\max}} \hat{\gamma}(u) du = \sum_{i=1}^n y_i (u_{i-1} - u_i) = a \sum_{i=1}^n y_i = 1. \quad (16)$$

Тогда  $\hat{\gamma}(u)$  есть оценка плотности распределения. При использовании этого приближения функция математического ожидания потерь  $L(c)$  оценивается как

$$L_{\text{hist}}(c) = \int_{u_{\min}}^{u_{\max}} \mathcal{L}(c, u) \hat{\gamma}(u) du = \sum_{i=1}^n y_i \int_{u_{i-1}}^{u_i} \mathcal{L}(c, u) du. \quad (17)$$

## 4.2 Приближение интеграла.

Интеграл от функции потерь, присутствующий в (17), приближается по методу прямоугольников со средней точкой.

$$\int_{u_{i-1}}^{u_i} \mathcal{L}(c, u) du \approx a \mathcal{L}\left(c, \frac{u_{i-1} + u_i}{2}\right). \quad (18)$$

После этого приближённая функция математического ожидания потерь примет окончательный вид

$$L_{\text{conv}}(c) = a \sum_{i=1}^n y_i \mathcal{L}\left(c, \frac{u_{i-1} + u_i}{2}\right). \quad (19)$$

Точность приближений растёт с ростом числа отрезков  $n$ ; в первом случае (17) это связано с уточнением приближения гистограммой исходной плотности распределения, во втором (19) — с уточнением оценки интеграла.

Работа алгоритма заключается в поиске  $c^*$ , на котором достигается минимум  $L_{\text{conv}}(c)$  и взятии значения  $c^*$  в качестве прогноза. В силу возможной сложности функции потерь  $\mathcal{L}(\hat{x}, x)$  минимум ищется среди значений функции в ограниченном наборе точек

$$G = \left\{ \frac{u_0 + u_1}{2}, \dots, \frac{u_{n-1} + u_n}{2} \right\}, \quad (20)$$

состоящем из середин отрезков разбиений:

$$\hat{x}_{T+1}^s = \arg \min_{c \in G} L_{\text{conv}}(c). \quad (21)$$

## 4.3 Алгоритм *Hist*.

**Вход:** стационарный ряд регрессионных остатков  $\mathbf{r}$ , функция потерь  $\mathcal{L}(\hat{x}, x)$ ;

**Выход:** прогноз  $\hat{x}_{T+1}^s$ , минимизирующий математическое ожидание потерь;

- 1: задать число столбцов гистограммы  $n$ ;
- 2: вычислить ширину столбцов гистограммы  $a = \frac{\max \mathbf{r} - \min \mathbf{r}}{n}$  и координаты концов отрезков постоянства  $u_0, u_1, \dots, u_n$  для функции  $\hat{\gamma}(u)$ ;
- 3: построить гистограмму, найти функцию  $\hat{\gamma}(u)$ , отнормировав гистограмму, получить значения функции на отрезках постоянства  $y_1, \dots, y_n$ ;
- 4: найти значения свертки  $\sum_{i=1}^n y_i \mathcal{L}\left(c, \frac{u_i + u_{i-1}}{2}\right)$  для всех  $c \in \left\{ \frac{u_0 + u_1}{2}, \dots, \frac{u_{n-1} + u_n}{2} \right\}$ ;
- 5: выбрать  $c^*$ , дающее минимальное значение свертки;
- 6:  $\hat{x}_{T+1}^s = c^*$ .

Основной параметр алгоритма, который можно варьировать, — число столбцов гистограммы  $n$ . При малых  $n$  оценка плотности распределения  $\hat{\gamma}(u)$  получается огрубленной, при больших  $n$  — более детальной. В следующем разделе будут приведены результаты исследования свойств алгоритма и на регрессионных остатках различных временных рядов.

Алгоритм *Hist* минимизирует математическое ожидание потерь прогнозирования при любом распределении регрессионных остатков  $\mathbf{r}$  и произвольной функции потерь  $\mathcal{L}(\hat{x}, x)$ .



Если регрессионные остатки имеют нулевое среднее, то смещение  $\hat{x}_{T+1}^s$  будет обусловлено лишь несимметричностью функции потерь. Однако использование алгоритма *Hist* при двухэтапном прогнозе может уменьшить средние потери и в том случае, если подобранная для прогнозирования нестационарной компоненты модель дает смещенные прогнозы. При поиске решения задачи минимизации ожидаемых потерь (21) это смещение будет скомпенсировано, что приведет к повышению качества даже в случае симметричной функции потерь.

Отметим также, что прогноз *Hist* является константой в том смысле, что не зависит от горизонта прогнозирования, и для любого будущего момента времени  $t$  ответ будет одним и тем же. То есть, иными словами, *Hist* строит регрессионную модель нулевого порядка.

## 5 Двухэтапное прогнозирование. Алгоритм *ARIMA + Hist*

В предыдущем разделе было показано, что при правильном подборе параметров модели *ARIMA* ряд из регрессионных остатков представляет собой белый шум  $\varepsilon_t$ , см. [12]. Но это и означает, что ряд из регрессионных остатков стационарен и сгенерирован из распределения  $\mathcal{N}(0; \sigma_a^2)$ .

Как было сказано выше в (7), *ARIMA* настраивается таким образом, чтобы минимизировать регрессионные остатки для квадратичной функции потерь  $\mathcal{L}_{sq}(\hat{x}, x)$ . Если же функция потерь  $\mathcal{L}(\hat{x}, x)$ , по которой производится оценка качества прогноза, не является квадратичной, то и регрессионные остатки в общем случае минимальными не будут.

Далее запускается алгоритм *Hist* с действительной функцией ошибок на ряде  $\varepsilon_t$  — то есть минимизируем остатки в смысле новой функцией потерь  $\mathcal{L}(\hat{x}, x)$ .

Итоговой прогноз суммируется из двух  $\hat{x}_{T+1} = \hat{x}_{T+1}^{ns} + \hat{x}_{T+1}^s$ . По сути же *Hist* добавляет одинаковый на всём горизонте сдвиг вверх или вниз от исходного прогноза *ARIMA* в зависимости от конкретного вида функции потерь.

### 5.1 Алгоритм *ARIMA + Hist*.

---

**Вход:** временной ряд  $\mathbf{x}$ , функция потерь  $\mathcal{L}(\hat{x}, x)$ ;

**Выход:** прогноз  $\hat{x}_{T+1}$ ;

---

- 1: подобрать подходящую для временного ряда модель *ARIMA* по методологии Бокса-Дженкинса;
  - 2: вычислить прогноз нестационарной компоненты  $\hat{x}_{T+1}^{ns}$  с помощью выбранной модели *ARIMA*;
  - 3: вычислить регрессионные остатки  $\mathbf{r}$  для выбранной модели *ARIMA*;
  - 4: задать число столбцов  $n$  в гистограмме для алгоритма *Hist*;
  - 5: вычислить прогноз стационарной компоненты  $\hat{x}_{T+1}^s$  с помощью алгоритма *Hist*;
  - 6:  $\hat{x}_{T+1} = \hat{x}_{T+1}^{ns} + \hat{x}_{T+1}^s$ ;
- 

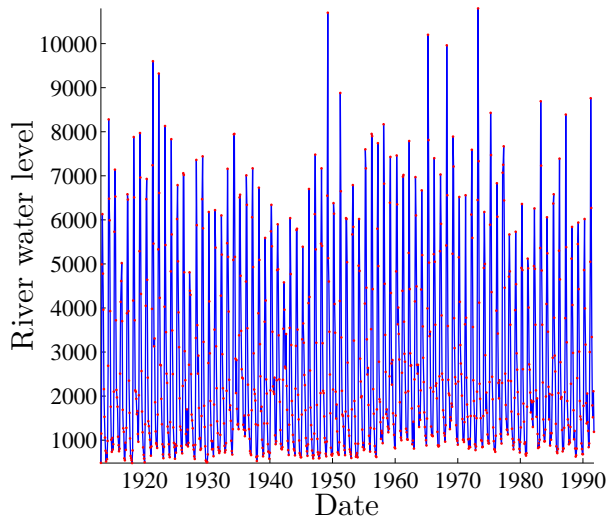
Предложенный двухэтапный алгоритм прогнозирования временных рядов позволяет прогнозировать различные нестационарные временные ряды в условиях несимметричности потерь. Он не накладывает на вид функции потерь ограничений, она может быть несимметричной или симметричной, дифференцируемой либо нет.

## 6 Исследование свойств алгоритма *Hist*

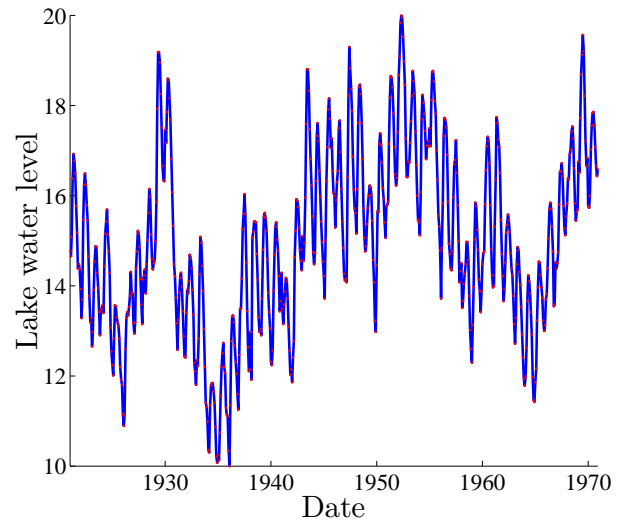
В данном разделе описываются и исследуются свойства алгоритма *Hist*, основанного на свертке гистограммы с функцией потерь.

## 6.1 Используемые временные ряды.

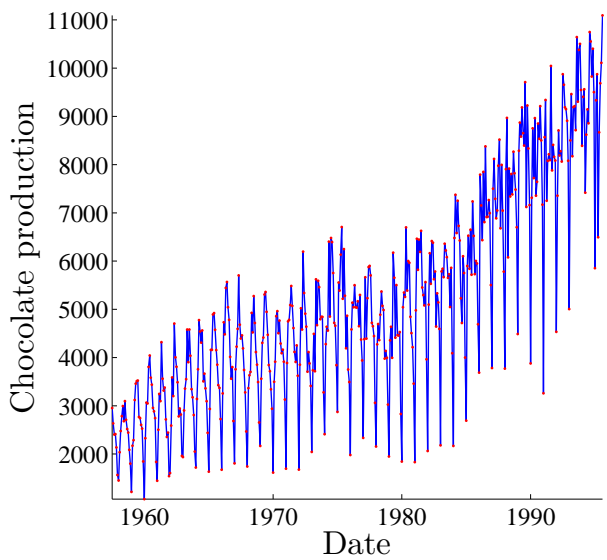
Для исследования свойств алгоритма *Hist* и последующего вычислительного эксперимента были использованы временные ряды [13–17], изображенные на рисунке 1. Рассматриваемые ряды отличаются друг от друга длиной истории, наличием или отсутствием сезонности и тренда, диапазоном значений. После первого этапа алгоритма *ARIMA* + *Hist* получаются стационарные ряды регрессионных остатков. Именно на этих рядах и проводится исследование свойств *Hist*. Первый этап алгоритма и сравнение качества прогнозов описаны в следующем разделе.



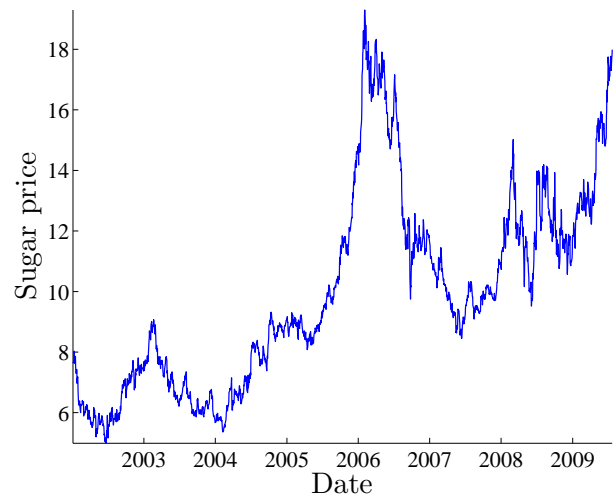
(a) Fraser River at hope



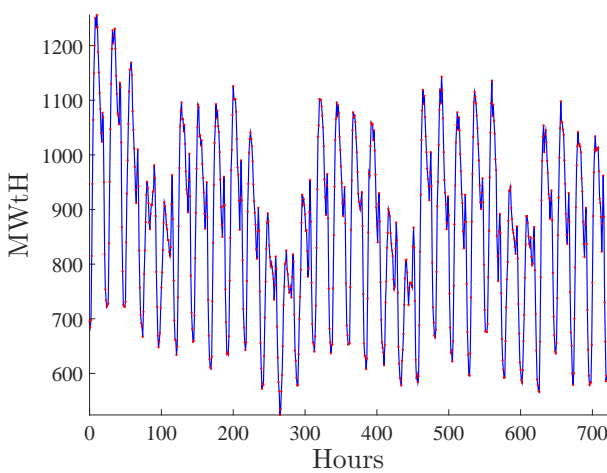
(б) Monthly Lake Erie Levels



(в) Monthly production of chocolate



(г) Sugar price

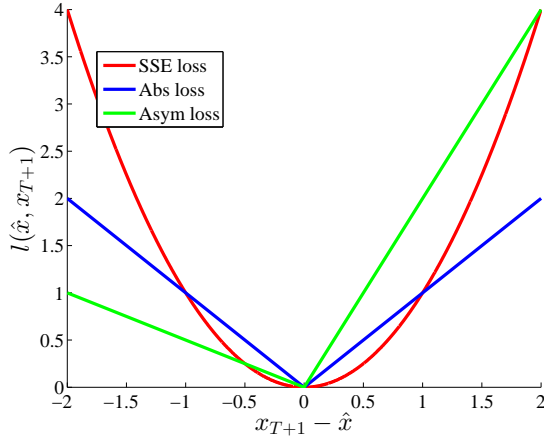


(д) Electricity consumption July 2005

Рис. 1 Временные ряды.

## 6.2 Функции потерь.

Эксперименты проводились для трех различных функций потерь:



$$\mathcal{L}_{\text{sq}}(\hat{x}, x) = (\hat{x} - x)^2; \quad (22)$$

$$\mathcal{L}_{\text{abs}}(\hat{x}, x) = |\hat{x} - x|; \quad (23)$$

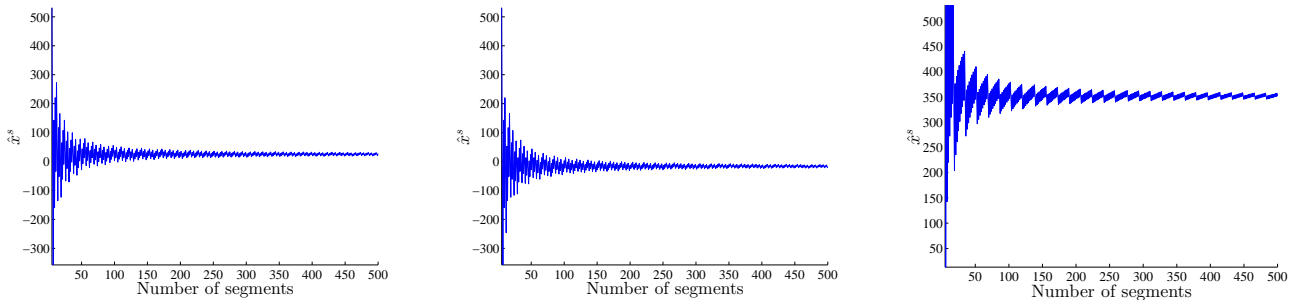
$$\mathcal{L}_{\text{asym}}(\hat{x}, x) = \begin{cases} \frac{1}{2}|\hat{x} - x|, & x \leq \hat{x}, \\ 2|\hat{x} - x|, & x > \hat{x}. \end{cases} \quad (24)$$

Рис. 2 Функции потерь.

Графики квадратичной, абсолютной и ассиметричной функций потерь изображены на рис. 2. Все три функции выпуклые, достигают минимума при совпадении прогноза и действительного значения временного ряда. Первые две функции симметричные, последняя — несимметричная кусочно-линейная функция.

## 6.3 Свойства прогноза алгоритма Hist.

Чтобы определить, как зависит поведение алгоритма *Hist* от функции потерь и количества столбцов в гистограмме, для каждого ряда регрессионных остатков были построены графики зависимости прогноза алгоритма *Hist* от количества столбцов в гистограмме для каждой функции потерь (22)–(24). Графики изображены на рис. 3–7. На каждом графике по оси абсцисс отложено количество столбцов гистограммы, по оси ординат — прогноз, полученный алгоритмом *Hist* при использовании заданной функции потерь и гистограммы с заданным числом столбцов.



(а) Квадратичная функция потерь.

(б) Абсолютная функция потерь.

(в) Ассиметричная функция потерь.

Рис. 3 Прогнозы алгоритма *Hist* для регрессионных остатков ряда Fraser River at hope.

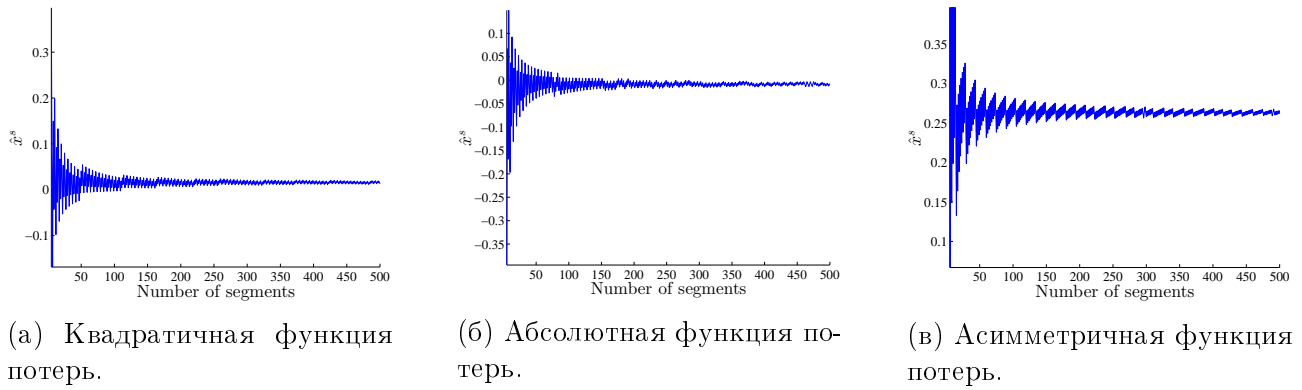


Рис. 4 Прогнозы алгоритма *Hist* для регрессионных остатков ряда Monthly Lake Erie Levels.

На рис. 3–7 видно, что для всех временных рядов и любой функции потерь с увеличением числа столбцов гистограммы полученные прогнозы стабилизируются вокруг предельного значения. Для симметричных функций потерь (22), (23) предельное значение для прогнозов близко к нулю, что означает, что для симметричных функций потерь алгоритм *Hist* не дает существенной поправки к прогнозу нестационарной компоненты, полученному с помощью модели *ARIMA*. В то же время для несимметричной функции потерь (24) предельное значение прогнозов существенно больше нуля. Это значит, что суммарный прогноз будет значительно превышать прогноз нестационарной компоненты, поскольку рассматриваемая функция потерь (24) штрафует недопрогноз гораздо сильнее, чем перепрогноз.

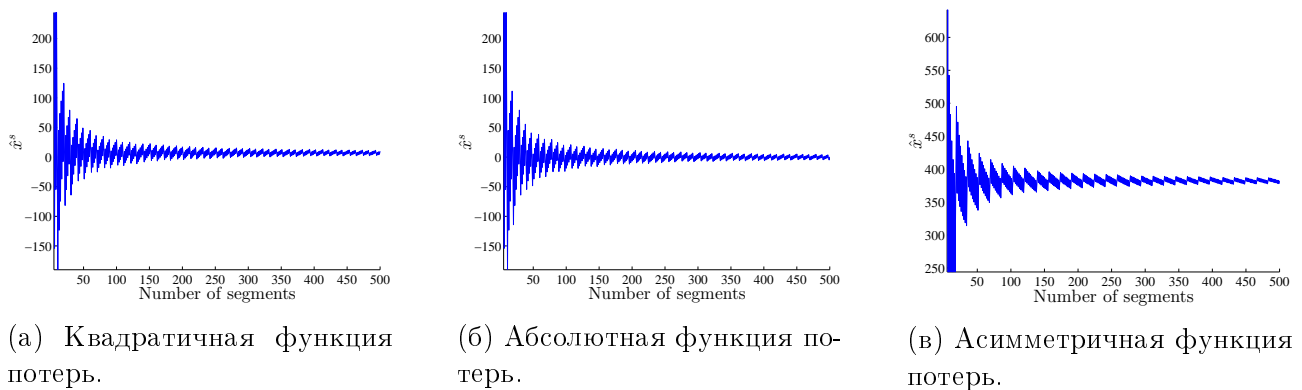


Рис. 5. Прогнозы алгоритма *Hist* для регрессионных остатков ряда Monthly production of chocolate confectionery in Australia.

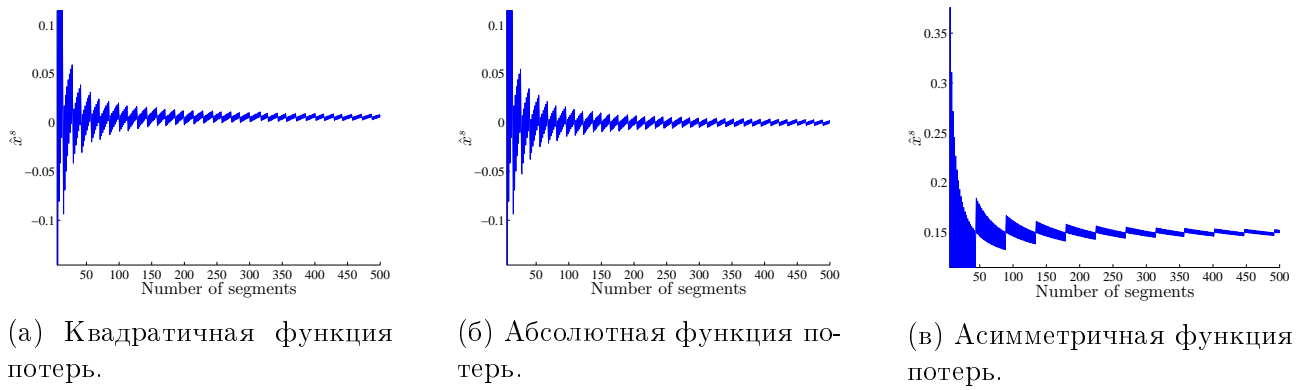


Рис. 6 Прогнозы алгоритма *Hist* для регрессионных остатков ряда Sugar price.

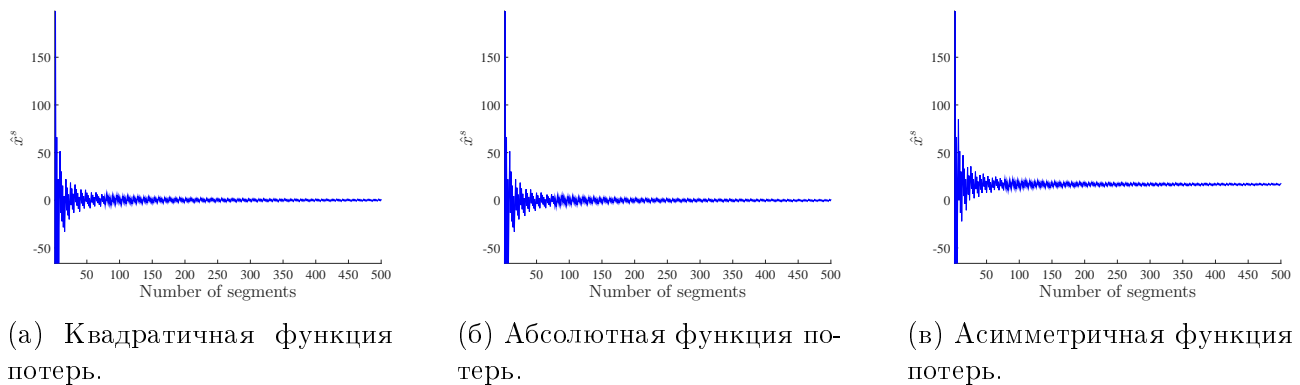


Рис. 7 Прогнозы алгоритма *Hist* для регрессионных остатков ряда Electricity consumption.

Стабилизация прогнозов алгоритма *Hist* с увеличением количества столбцов в гистограмме связана с увеличением точности оценки плотности распределения регрессионных остатков  $\gamma(u)$ , о которой говорилось в разделе с описанием алгоритма *Hist*. Однако для конечных временных рядов добиться сходимости прогнозов к предельному значению с любой наперед заданной точностью невозможно из-за конечного количества доступных данных для оценки плотности распределения  $\gamma(u)$ .

## 7 Вычислительный эксперимент

Целью проведенного вычислительного эксперимента является сравнение средних потерь прогнозирования различных временных рядов для различных функций потерь при использовании модели *ARIMA* и предложенного двухэтапного алгоритма *ARIMA+Hist*. Рассмотрены пять различных временных рядов [13–17], изображенных на рисунке 1, и три функции потерь (22)–(24), одна из которых несимметричная. Получено экспериментальное подтверждение того, что при несимметричных потерях использование двухэтапного прогнозирования позволяет уменьшить средние потери.

**Таблица 1** Выбранные модели *ARIMA* для прогноза нестационарной компоненты

Временной ряд	Модель <i>ARIMA</i>
Fraser River at hope	$SARIMA(1, 0, 0) \times (1, 0, 1)_{12}$
Monthly Lake Erie Levels	$SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$
Monthly production of chocolate confectionery in Australia	$SARIMA(1, 1, 1) \times (1, 0, 1)_{12}$
Sugar price	$ARIMA(1, 0, 0)$
Electricity consumption	$SARIMA(2, 1, 2) \times (1, 0, 1)_{24,168}$

Для каждого временного ряда, изображенного на рис. 1, подбиралась модель *ARIMA* по методологии Бокса-Дженкинса [2]. Выбранные модели для каждого временного ряда показаны в таблице 1.

Для прогнозирования стационарной компоненты используются регрессионные остатки алгоритма *ARIMA*.

Для сравнения качества прогнозов модели *ARIMA* и связки алгоритмов *ARIMA* + *Hist* 20% последних точек каждого временного ряда использовались как контрольные. Для каждой контрольной точки по доступной истории временного ряда (все точки от первой до предшествующей рассматриваемой контрольной) обучалась выбранная для временного ряда модель *ARIMA*, затем для обученной модели вычислялся ряд регрессионных остатков. По ряду регрессионных остатков обучался алгоритм *Hist* с заданной функцией потерь и заданным количеством столбцов в гистограмме. Прогноз для контрольной точки складывался из прогноза *ARIMA* и *Hist*. Эксперимент был проведен для функций потерь (22)–(24) и вариантов алгоритма *Hist* с 20, 50, 300, 500 столбцами в гистограмме. Средние потери для каждой функции потерь приведены для всех вариантов алгоритма в таблицах ниже.

**Таблица 2** Средние потери прогнозирования для временного ряда Fraser River at hope

Алгоритм	Квадратичная функция потерь	Абсолютная функция потерь	Асимметричная функция потерь
<i>ARIMA</i>	52400	498	616
<i>ARIMA</i> + <i>Hist</i> (20)	53500	495	523
<i>ARIMA</i> + <i>Hist</i> (50)	<b>52200</b>	496	516
<i>ARIMA</i> + <i>Hist</i> (300)	52500	493	516
<i>ARIMA</i> + <i>Hist</i> (500)	52400	<b>492</b>	<b>515</b>

**Таблица 3** Средние потери прогнозирования для временного ряда Monthly Lake Erie Levels

Алгоритм	Квадратичная функция потерь	Абсолютная функция потерь	Асимметричная функция потерь
<i>ARIMA</i>	0.172	0.313	0.410
<i>ARIMA</i> + <i>Hist</i> (20)	0.182	0.316	0.315
<i>ARIMA</i> + <i>Hist</i> (50)	0.171	<b>0.313</b>	0.312
<i>ARIMA</i> + <i>Hist</i> (300)	0.171	0.314	0.311
<i>ARIMA</i> + <i>Hist</i> (500)	<b>0.171</b>	0.314	<b>0.311</b>

**Таблица 4.** Средние потери прогнозирования для временного ряда Monthly production of chocolate confectionery in Australia

Алгоритм	Квадратичная функция потерь	Абсолютная функция потерь	Асимметричная функция потерь
<i>ARIMA</i>	71500000	8350	4180
<i>ARIMA + Hist(20)</i>	66000	612	579
<i>ARIMA + Hist(50)</i>	65800	609	575
<i>ARIMA + Hist(300)</i>	65300	610	575
<i>ARIMA + Hist(500)</i>	<b>65100</b>	<b>609</b>	<b>575</b>

**Таблица 5** Средние потери прогнозирования для временного ряда Sugar price

Алгоритм	Квадратичная функция потерь	Абсолютная функция потерь	Асимметричная функция потерь
<i>ARIMA</i>	0.127	0.265	0.340
<i>ARIMA + Hist(20)</i>	0.128	0.267	<b>0.260</b>
<i>ARIMA + Hist(50)</i>	0.127	0.266	0.267
<i>ARIMA + Hist(300)</i>	0.127	0.265	0.266
<i>ARIMA + Hist(500)</i>	<b>0.127</b>	<b>0.265</b>	0.266

**Таблица 6** Средние потери прогнозирования для временного ряда Electricity consumption

Алгоритм	Квадратичная функция потерь	Абсолютная функция потерь	Асимметричная функция потерь
<i>ARIMA</i>	500.13	16.93	19.92
<i>ARIMA + Hist(20)</i>	717.68	19.09	<b>12.74</b>
<i>ARIMA + Hist(50)</i>	588.91	<b>16.47</b>	13.43
<i>ARIMA + Hist(300)</i>	<b>498.35</b>	17.10	13.03
<i>ARIMA + Hist(500)</i>	502.31	16.87	13.08

Как видно из табл. 2–6, при использовании асимметричной функции потерь двухэтапный алгоритм прогнозирования *ARIMA + Hist* позволяет получать среднюю ошибку прогноза существенно ниже, чем прогнозирование с помощью модели *ARIMA*. В большинстве случаев для симметричных функций потерь использование двухэтапного алгоритма прогнозирования не приводит к значительным изменениям по сравнению с прогнозом модели *ARIMA*. Исключение составляет только табл. 4, в которой представлены средние потери прогнозирования для временного ряда Monthly production of chocolate confectionery in Australia. Для этого временного ряда использование двухэтапного алгоритма прогнозирования *ARIMA + Hist* привело к существенному уменьшению потерь для всех функций потерь. Это связано с тем, что, как видно на рис. 1(с), этот временной ряд во второй половине истории имеет более высокий темп роста, чем в первой половине. Обученная преимущественно первой половине истории модель *ARIMA* дает прогнозы в контрольных точках, сильно смещенные в одну сторону относительно реальных значений. При использовании двухэтапного алгоритма прогнозирования *ARIMA + Hist* на втором шаге с помощью алгоритма *Hist* удастся оценить это смещение и сделать более точный прогноз.



## 8 Заключение

Предложен двухэтапный алгоритм прогнозирования нестационарных временных рядов *ARIMA+Hist*, минимизирующий ожидаемые потери. Он не накладывает на вид функции потерь ограничений — она может быть симметричной или несимметричной, дифференцируемой или нет, однако наиболее предпочтительны выпуклые функции. На первом этапе строится прогноз нестационарной компоненты временного ряда путем выбора подходящей модели *ARIMA*. На втором этапе оценивается плотность распределения регрессионных остатков выбранной модели *ARIMA* и строится прогноз стационарной компоненты путем минимизации математического ожидания потерь, которые заданы несимметричной функцией потерь. Финальный прогноз вычисляется как сумма прогнозов нестационарной и стационарной компоненты временного ряда. С помощью вычислительного эксперимента показано, что двухэтапное прогнозирование в случае несимметричной функции потерь позволяет уменьшить средние потери по сравнению с одноэтапным прогнозированием *ARIMA*. Также на практике средние потери можно уменьшить с помощью двухэтапного прогнозирования в случае симметричной функции потерь и смещенных прогнозов нестационарной компоненты временного ряда.

## Литература

- [1] Patton Andrew J, Timmermann Allan. Properties of optimal forecasts under asymmetric loss and nonlinearity // *Journal of Econometrics*. — 2007. — Vol. 140, no. 2. — P. 884–918.
- [2] Box G. E. P., Jenkins G. M., Reinsel G. C. *Time Series Analysis: Forecasting and Control* / Ed. by NJ: Prentice Hall. — 3rd edition. — Englewood Cliffs, 1994.
- [3] Berk Richard. Asymmetric loss functions for forecasting in criminal justice settings // *Journal of Quantitative Criminology*. — 2011. — Vol. 27, no. 1. — P. 107–123.
- [4] Cipra Tomáš. Asymmetric recursive methods for time series // *Applications of Mathematics*. — 1994. — Vol. 39, no. 3. — P. 203–214.
- [5] Koenker Roger, Xiao Zhijie. Quantile autoregression // *Journal of the American Statistical Association*. — 2006. — Vol. 101, no. 475. — P. 980–990.
- [6] Koenker Roger. *Quantile regression*. — Cambridge university press, 2005.
- [7] Christoffersen Peter F, Diebold Francis X. Optimal prediction under asymmetric loss // *Econometric theory*. — 1997. — Vol. 13, no. 06. — P. 808–817.
- [8] Granger Clive WJ. Prediction with a generalized cost of error function // *OR*. — 1969. — June. — Vol. 20, no. 2. — P. 199–207.
- [9] Diebold Francis X., Gunther Todd A., Tay Anthony S. Evaluating density forecasts. — 1997.
- [10] Christoffersen Peter F, Diebold Francis X. Further results on forecasting and model selection under asymmetric loss // *Journal of applied econometrics*. — 1996. — Vol. 11, no. 5. — P. 561–571.
- [11] Nonparametric sequential prediction of time series / Gérard Biau, Kevin Bleakley, László Györfi, György Ottucsák // *Journal of Nonparametric Statistics*. — 2010. — Vol. 22, no. 3. — P. 297–317.
- [12] Кобзарь А. И. *Прикладная математическая статистика*— ФИЗМАТЛИТ, 2006.
- [13] Fraser River at hope. — URL: <https://datamarket.com/data/set/22nm/fraser-river-at-hope-1913-1990#!ds=22nm&display=line>.
- [14] Monthly Lake Erie Levels. — URL: <https://datamarket.com/data/set/22pw/monthly-lake-erie-levels-1921-1970#!ds=22pw&display=line>.

- [15] Monthly production of chocolate confectionery in Australia. — URL: <https://datamarket.com/data/set/22rl/monthly-production-of-chocolate-confectionery-in-australia-tonnes-july-1957!ds=22rl&display=line>.
- [16] Sugar price. — URL: <https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSPForecasting/TimeSeries/Sources/tsSugarPrice.csv>.
- [17] Electricity consumption. — URL: <https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSPForecasting/TimeSeries/Sources/tsEnergyConsumption.csv>.

## References

- [1] Patton Andrew J, Timmermann Allan. Properties of optimal forecasts under asymmetric loss and nonlinearity // *Journal of Econometrics*. — 2007. — Vol. 140, no. 2. — P. 884–918.
- [2] Box G. E. P., Jenkins G. M., Reinsel G. C. *Time Series Analysis: Forecasting and Control* / Ed. by NJ: Prentice Hall. — 3rd edition. — Englewood Cliffs, 1994.
- [3] Berk Richard. Asymmetric loss functions for forecasting in criminal justice settings // *Journal of Quantitative Criminology*. — 2011. — Vol. 27, no. 1. — P. 107–123.
- [4] Cipra Tomáš. Asymmetric recursive methods for time series // *Applications of Mathematics*. — 1994. — Vol. 39, no. 3. — P. 203–214.
- [5] Koenker Roger, Xiao Zhijie. Quantile autoregression // *Journal of the American Statistical Association*. — 2006. — Vol. 101, no. 475. — P. 980–990.
- [6] Koenker Roger. *Quantile regression*. — Cambridge university press, 2005.
- [7] Christoffersen Peter F, Diebold Francis X. Optimal prediction under asymmetric loss // *Econometric theory*. — 1997. — Vol. 13, no. 06. — P. 808–817.
- [8] Granger Clive WJ. Prediction with a generalized cost of error function // *OR*. — 1969. — June. — Vol. 20, no. 2. — P. 199–207.
- [9] Diebold Francis X., Gunther Todd A., Tay Anthony S. Evaluating density forecasts. — 1997.
- [10] Christoffersen Peter F, Diebold Francis X. Further results on forecasting and model selection under asymmetric loss // *Journal of applied econometrics*. — 1996. — Vol. 11, no. 5. — P. 561–571.
- [11] Nonparametric sequential prediction of time series / Gérard Biau, Kevin Bleakley, László Györfi, György Ottucsák // *Journal of Nonparametric Statistics*. — 2010. — Vol. 22, no. 3. — P. 297–317.
- [12] Kobzar A. I. *Prikladnaya matematicheskaya statistika*— FIZMATLIT, 2006.
- [13] Fraser River at hope. — URL: <https://datamarket.com/data/set/22nm/fraser-river-at-hope-1913-1990#!ds=22nm&display=line>.
- [14] Monthly Lake Erie Levels. — URL: <https://datamarket.com/data/set/22pw/monthly-lake-erie-levels-1921-1970#!ds=22pw&display=line>.
- [15] Monthly production of chocolate confectionery in Australia. — URL: <https://datamarket.com/data/set/22rl/monthly-production-of-chocolate-confectionery-in-australia-tonnes-july-1957!ds=22rl&display=line>.
- [16] Sugar price. — URL: <https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSPForecasting/TimeSeries/Sources/tsSugarPrice.csv>.
- [17] Electricity consumption. — URL: <https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSPForecasting/TimeSeries/Sources/tsEnergyConsumption.csv>.