Прогнозирование нестационарных временных рядов при несимметричных функциях потерь*

В. Ю. Черных[1]***, М. М. Стенина***[1,2]

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

[1]Московский физико-технический институт, Россия, г. Долгопрудный, Институтский пер., 9

[2]Высшая школа экономики, Россия, г. Москва, ул. Мясницкая, 20

Рассматривается задача прогнозирования временны́х 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 временны́х 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*

DOI: 10.21469/22233792.1.14.01

1 Введение

Рассматривается задача прогнозирования нестационарных временны́х рядов в случае несимметричных функций потерь. Предлагается двухэтапный алгоритм прогнозирования ARIMA + Hist, на первом этапе которого отслеживаются свойства временно́го ряда, обусловливающие его нестационарность, такие как тренд и сезонность. На втором этапе предлагается находить поправку, обеспечивающую оптимальность прогноза в случае несимметричной функции потерь.

Свойства прогнозов временны́х 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 временны́х рядов в условиях несимметричности функции потерь.

Один из используемых методов прогнозирования нестационарных временны́х рядов, авторегрессионное интегрированное скользящее среднее ARIMA [2], позволяет с хорошим качеством прогнозировать временны́е 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

*Работа выполнена при финансовой поддержке РФФИ, проект № 14-07-31046.

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 временны́х 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 временны́х 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 временно́го 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 временно́го 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 и используется только для прогнозирования стационарных временны́х рядов. Такая задача минимизации рассматривалась в работах [9, 10], где математическое ожидание потерь было представлено как свертка функции потерь с функцией плотности распределения значений временно́го ряда. На втором этапе алгоритма ARIMA + Hist в качестве временно́го ряда выступают регрессионные остатки, однако их плотность распределения неизвестна. 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 временны́х 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 $\mathbf{x} = \{(x_i)_{i=1}^T \,|\, x_i \in \mathbb{R}\}$. Также задается горизонт прогнозирования h. Ставится задача прогнозирования этого временно́го ряда, т. е. нахождения регрессионной модели

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

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

Временной ряд $\mathbf{x}$ называется *стационарным*, если для любых v многомерное распределение $x_t, \ldots, x_{t+v}$ не зависит от t, т. е. его свойства не зависят от времени. Из определения немедленно следует, что все значения ряда $x_1, \ldots, x_T$ генерируются из одного распределения $\rho(u)$, которое не меняется во времени. Пусть задана функция потерь $\mathscr{L}(\hat{x}, x)$ и требуется получить прогноз $\hat{x}_{T+1}$ следующего значения x_{T+1} временно́го ряда, минимизируя ожидаемые потери. Предполагается, что следующее значение временно́го ряда генерируется из того же распределения, что и все предыдущие. При этом задача прогнозирования запишется как

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

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

$$L(c) = \mathsf{E}\, \mathscr{L}(c, x_{T+1}) = \int\limits_{-\infty}^{+\infty} \mathscr{L}(c, u)\, \rho(u)\, du.$$

В таком случае задача прогнозирования формулируется как

$$\hat{x}_{T+1} = \arg\min_{c \in \mathbb{R}} \int\limits_{-\infty}^{+\infty} \mathscr{L}(c, u)\, \rho(u)\, du \equiv \arg\min_{c \in \mathbb{R}} L(c). \qquad (1)$$

2.2 Прогнозирование нестационарных временных рядов

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

$$\hat{x}_{T+1} = \hat{x}^{\mathrm{ns}}_{T+1} + \hat{x}^{\mathrm{s}}_{T+1}.$$

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

$$\mathbf{r} = \{(r_i)_{i=1}^T \,|\, r_i = x_i - \hat{x}_i^{\mathrm{ns}}\}$$

были стационарным временны́м 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 $\mathscr{L}_{\mathrm{sq}}(\hat{x}, x) = (\hat{x} - x)^2$, по которой строится функционал потерь:

$$\mathcal{Q}(f^{\mathrm{ns}}, \mathbf{x}) = \frac{1}{T}\sum_{i=1}^{T} \mathscr{L}_{\mathrm{sq}}(f^{\mathrm{ns}}(\mathbf{w}, \mathbf{x}_i, 1), x_{i+1}); \quad \mathbf{x}_i = \{x_1 \cdots x_i\}. \tag{2}$$

Тогда решение задачи минимизации (2) дает вектор параметров искомой регрессионной модели:

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

При этом прогноз вычисляется следующим образом:

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

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

3 Прогнозирование нестационарной компоненты. ARIMA. Методология Бокса–Дженкинса

В данном разделе описывается модель авторегрессионного интегрированного скользящего среднего ARIMA и методология Бокса–Дженкинса прогнозирования временны́х 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 временно́го ряда предлагается использовать для получения прогноза нестационарной компоненты $\hat{x}_{T+1}^{\mathrm{ns}}$. Анализ того, насколько хорошо выбранная модель аппроксимирует временной ряд, по методологии Бокса–Дженкинса, включает проверку регрессионных остатков на несмещенность, стационарность и неавтокоррелированность. Модель считается подходящей для аппроксимации временно́го 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 временно́го ряда получается стационарный ряд регрессионных остатков, который можно использовать для построения прогноза $\hat{x}_{T+1}^{\mathrm{s}}$ стационарной компоненты временно́го ряда.

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

$$x_t = \alpha + \varepsilon_t + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \sum_{i=1}^{p} \theta_i x_{t-i}; \quad \alpha = \mu\left(1 - \sum_{i=1}^{p} \theta_i\right).$$

где $\theta_1, \ldots, \theta_p$, $\psi_1, \ldots, \psi_q$ — константы; ε_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:

$$\theta(L)x_t = \alpha + \psi(L)\varepsilon_t; \quad \theta(L) = 1 - \sum_{i=1}^{p} \theta_i L^i; \quad \psi(L) = 1 + \sum_{i=1}^{q} \psi_i L^i. \tag{3}$$

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

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

описывается моделью (3), при этом модель ARIMA(p, d, q) записывается как

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

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

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

3.1 Методология Бокса–Дженкинса анализа временных рядов

Методология Бокса–Дженкинса используется для оценки параметров модели ARIMA. Согласно этой методологии, порядок дифференцирования временно́го 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* τ *для временно́го ряда* $\mathbf{x}$ *вычисляется по формуле:*

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

Определение 2. *Частичная автокорреляционная функция* PACF$_\tau$ *с лагом автокорреляции* τ *для стационарного временно́го ряда* $\mathbf{x}$ *вычисляется по формуле:*

$$\mathrm{PACF}_\tau = \begin{cases} \mathsf{E}\left[x_{t+1}x_t\right], & \tau = 1; \\ \mathsf{E}\left[(x_{t+\tau} - x_{t+\tau}^{\tau-1})(x_t - x_t^{\tau-1})\right], & \tau \geqslant 2; \end{cases}$$

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

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

где $\beta_1, \ldots, \beta_{\tau-1}$ *— 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 ARIMA назначается с помощью анализа автокорреляционной и частичной автокорреляционной функций. При наличии сезонной компоненты у временно́го ряда на графиках этих функций будут наблюдаться характерные максимумы в лагах, соответствующих периоду S сезонной компоненты.

Необходимые коэффициенты многочленов $\theta(L)$ и $\psi(L)$ оптимизируются при использовании квадратичной функции потерь $\mathscr{L}_{\mathrm{sq}}(\hat{x}, x)$ и функционала потерь (2), основанного на ней.

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

(1) несмещенность $\mathsf{E}\, r_t = 0$;
(2) стационарность $\forall t \hookrightarrow r_t \sim \gamma(u)$;
(3) неавтокоррелированность $\mathsf{E}\, [r_t r_{t+k}] = 0, k \neq 0$,

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

4 Прогнозирование стационарной компоненты. Hist. Влияние функции потерь

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

В случае квадратичной $\mathscr{L}_{\mathrm{sq}}(\hat{x}, x) = (\hat{x} - x)^2$ или абсолютной $\mathscr{L}_{\mathrm{abs}}(\hat{x}, x) = |\hat{x} - x|$ функций потерь добавочный прогноз x^{s}_{T+1} можно найти аналитически, не зная при этом конкретного вида распределения $\gamma(u)$: для квадратичной функции потерь $x^{\mathrm{s}}_{T+1} = \mathsf{E}\, r_t$; для абсолютной $x^{\mathrm{s}}_{T+1} = \mathrm{med}\, \gamma(u)$, 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 $u_{\min} = \min\limits_t r_t$; $u_{\max} = \max\limits_t r_t$. 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

$$u_i = u_{\min} + ia; \ \ a = \frac{u_{\max} - u_{\min}}{n}.$$

На этих отрезках положим значения 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 $\hat{\gamma}(u)$:

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

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

$$L_{\text{hist}}(c) = \int\limits_{u_{\min}}^{u_{\max}} \mathscr{L}(c,u)\hat{\gamma}(u)\, du = \sum_{i=1}^{n} y_i \int\limits_{u_{i-1}}^{u_i} \mathscr{L}(c,u)\, du. \tag{4}$$

4.2 Приближение интеграла

Интеграл от функции потерь, присутствующий в (4), приближается по методу прямоугольников со средней точкой:

$$\int\limits_{u_{i-1}}^{u_i} \mathscr{L}(c,u)\, du \approx a\, \mathscr{L}\left(c, \frac{u_{i-1}+u_i}{2}\right).$$

После этого приближенная функция математического ожидания потерь примет окончательный вид:

$$L_{\text{conv}}(c) = a \sum_{i=1}^{n} y_i\, \mathscr{L}\left(c, \frac{u_{i-1}+u_i}{2}\right). \tag{5}$$

Точность приближений растет с ростом числа отрезков n; в первом случае (4) это связано с уточнением приближения гистограммой исходной плотности распределения, во втором (5) — 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 $\mathscr{L}(\hat{x}, x)$ минимум ищется среди значений функции в ограниченном наборе точек

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

состоящем из середин отрезков разбиений:

$$\hat{x}^{\text{s}}_{T+1} = \operatorname*{arg\,min}_{c \in G} L_{\text{conv}}(c). \tag{6}$$

4.3 Алгоритм Hist

Вход: стационарный ряд регрессионных остатков $\mathbf{r}$, функция потерь $\mathscr{L}(\hat{x}, x)$, число столбцов гистограммы n;

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

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

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 временны́х рядов.

Алгоритм Hist минимизирует математическое ожидание потерь прогнозирования при любом распределении регрессионных остатков $\mathbf{r}$ и произвольной функции потерь $\mathscr{L}(\hat{x}, x)$. Если регрессионные остатки имеют нулевое среднее, то смещение $\hat{x}^s_{T+1}$ будет обусловлено лишь несимметричностью функции потерь. Однако использование алгоритма ARIMA + Hist при двухэтапном прогнозе может уменьшить средние потери и в том случае, если подобранная для прогнозирования нестационарной компоненты модель дает смещенные прогнозы. Это может происходить при смене характера тренда. Например, увеличивается темп роста величины или величина сначала убывала, а потом начинает возрастать. Если изменение однократное и смещение прогнозов ARIMA, обученной на первом характере тренда, постоянное, то при поиске решения задачи минимизации ожидаемых потерь (6) это смещение будет скомпенсировано, что приведет к повышению качества даже в случае симметричной функции потерь. Примером такого ряда может служить ряд [13], во второй половине которого тренд более сильный. В разд. 7 будет показано, что для этого временно́го ряда использование алгоритма Hist действительно дает ощутимый выигрыш в качестве по сравнению с ARIMA даже для квадратичной 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, т. е. Hist строит регрессионную модель нулевого порядка.

5 Двухэтапное прогнозирование. Алгоритм ARIMA + Hist

Как было показано в (2), ARIMA настраивается таким образом, чтобы минимизировать регрессионные остатки для квадратичной функции потерь $\mathscr{L}_{\mathrm{sq}}(\hat{x}, x)$. Если же функция потерь $\mathscr{L}(\hat{x}, x)$, 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 $\mathbf{r}$, т. е. минимизируем остатки в смысле новой функцией потерь $\mathscr{L}(\hat{x}, x)$. Отметим, что первый этап действительно необходим и нельзя давать на вход Hist просто продифференцированный ряд, так как этот алгоритм не учитывает некоторые особенности, которые не уходят при дифференцировании, например сезонность.

Итоговой прогноз суммируется из двух $\hat{x}_{T+1} = \hat{x}^{\text{ns}}_{T+1} + \hat{x}^{\text{s}}_{T+1}$. По сути же Hist добавляет одинаковый на всем 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 $\mathscr{L}(\hat{x}, x)$;
Выход: прогноз $\hat{x}_{T+1}$;

1: подобрать подходящую для временнóго ряда модель ARIMA по методологии Бокса–Дженкинса;
2: вычислить прогноз нестационарной компоненты $\hat{x}^{\text{ns}}_{T+1}$ 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 $\hat{x}^{\text{s}}_{T+1}$ с помощью алгоритма Hist;
6: $\hat{x}_{T+1} = \hat{x}^{\text{ns}}_{T+1} + \hat{x}^{\text{s}}_{T+1}$;

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 временны́е ряды [13–17], изображенные на рис. 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.

6.2 Функции потерь

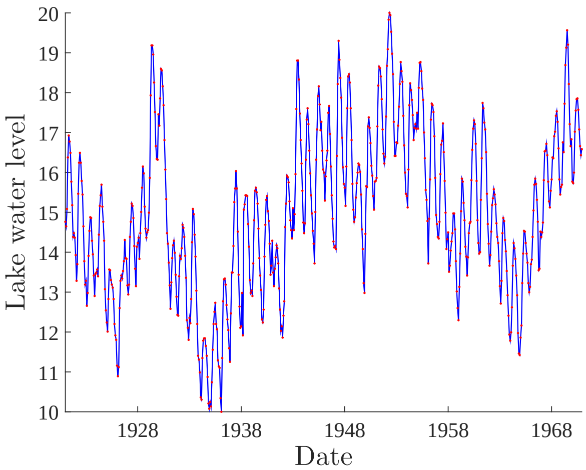
Эксперименты проводились для трех различных функций потерь:

$$\mathscr{L}_{\text{sq}}(\hat{x}, x) = (\hat{x} - x)^2\,; \tag{7}$$

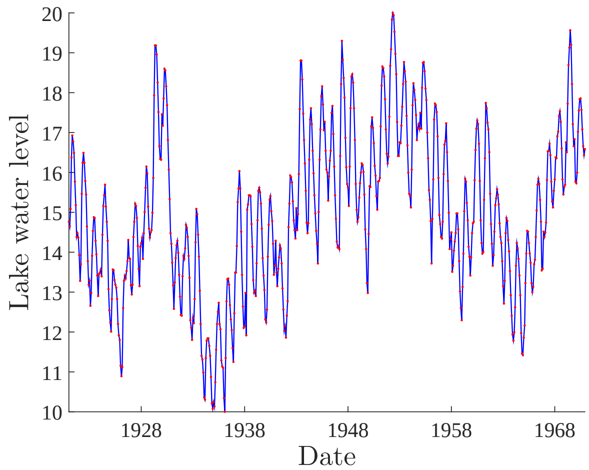
$$\mathscr{L}_{\text{abs}}(\hat{x}, x) = |\hat{x} - x|\,; \tag{8}$$

$$\mathscr{L}_{\text{asym}}(\hat{x}, x) = \begin{cases} \dfrac{1}{2}|\hat{x} - x|, & x \leqslant \hat{x}\,; \\ 2|\hat{x} - x|, & x > \hat{x}\,. \end{cases} \tag{9}$$

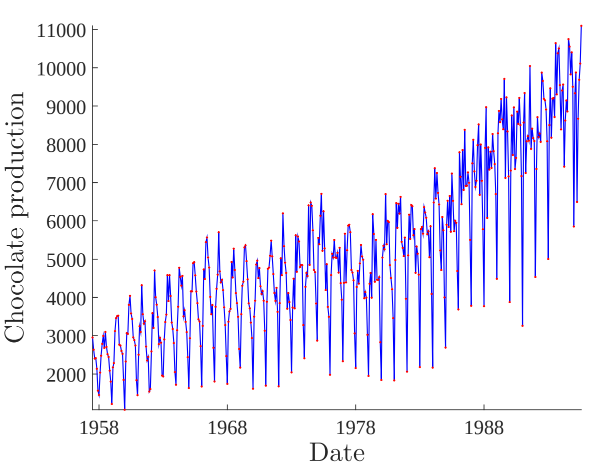
Графики квадратичной, абсолютной и ассимметричной функций потерь изображены на рис. 2. Все три функции выпуклые, достигают минимума при совпадении прогноза и действительного значения временнóго ряда. Первые две функции симметричные, последняя — несимметричная кусочно-линейная функция.



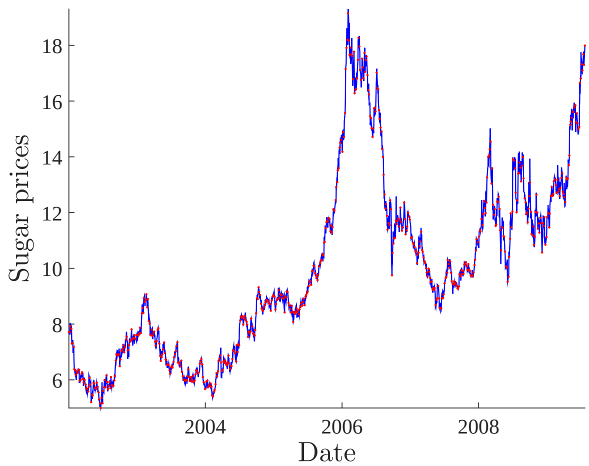
(a) Fraser River at hope



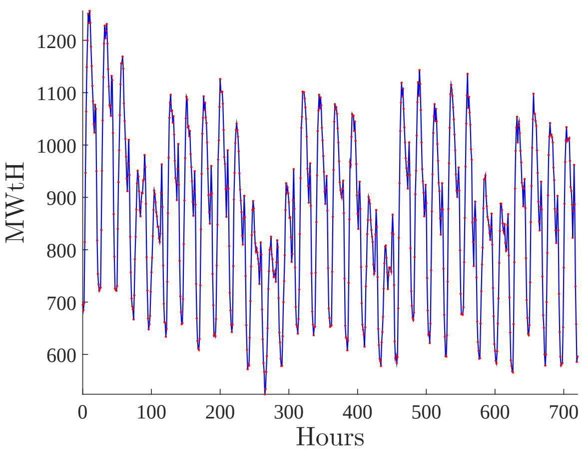
($б$) Monthly Lake Erie Levels



($в$) Monthly production of chocolate confectionery in Australia



($г$) Sugar price



(∂) Electricity consumption July 2005

Рис. 1 Временны́е ряды

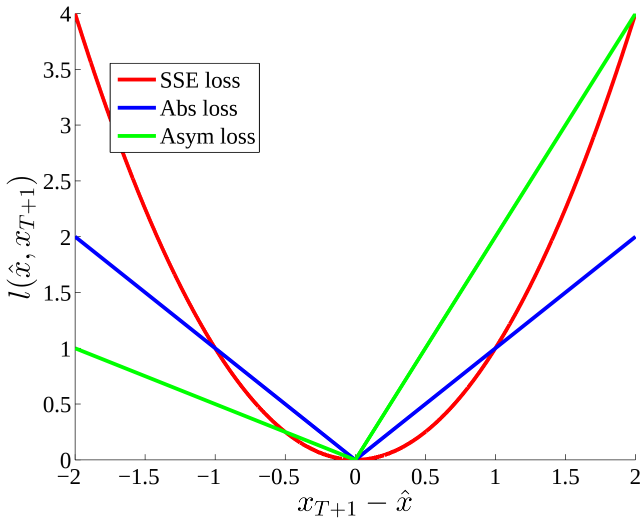


Рис. 2 Функции потерь

6.3 Свойства прогноза алгоритма Hist

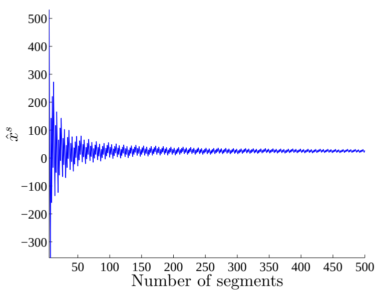
Чтобы определить, как зависит поведение алгоритма Hist от функции потерь и количества столбцов в гистограмме, для каждого ряда регрессионных остатков были построены графики зависимости прогноза алгоритма Hist от количества столбцов в гистограмме для каждой функции потерь (7)–(9). Графики изображены на 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.

На рис. 3–7 видно, что для всех временны́х рядов и любой функции потерь с увеличением числа столбцов гистограммы полученные прогнозы стабилизируются вокруг предельного значения. Для симметричных функций потерь (7) и (8) 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 (9) предельное значение прогнозов существенно больше нуля. Это значит, что суммарный прогноз будет значительно превышать прогноз нестационарной компоненты, поскольку рассматриваемая функция потерь (9) штрафует недопрогноз гораздо сильнее, чем перепрогноз.

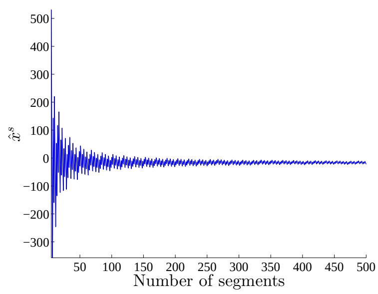
Стабилизация прогнозов алгоритма Hist с увеличением количества столбцов в гистограмме связана с увеличением точности оценки плотности распределения регрессионных остатков $\gamma(u)$, о которой говорилось в разд. 4. Однако для конечных временны́х рядов добиться сходимости прогнозов к предельному значению с любой наперед заданной точностью невозможно из-за конечного количества доступных данных для оценки плотности распределения $\gamma(u)$.

7 Вычислительный эксперимент

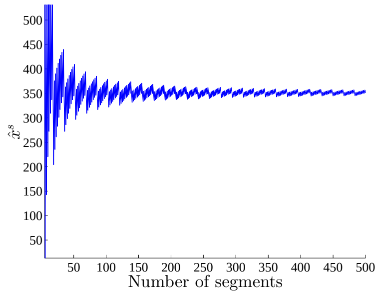
Целью проведенного вычислительного эксперимента является сравнение средних потерь прогнозирования различных временны́х рядов для различных функций потерь при использовании модели ARIMA и предложенного двухэтапного алгоритма ARIMA + Hist. Рассмотрены пять различных временны́х рядов [13–17], изображенных на рис. 1, и три



(*а*) Квадратичная функция потерь

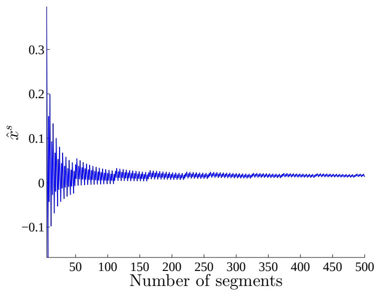


(*б*) Абсолютная функция потерь

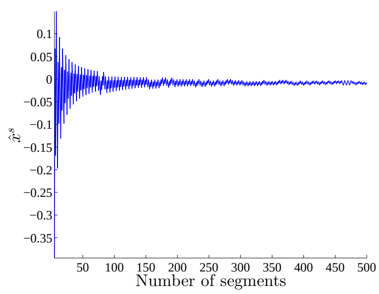


(*в*) Асимметричная функция потерь

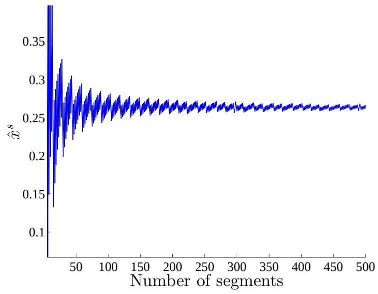
Рис. 3 Прогнозы алгоритма Hist для регрессионных остатков ряда Fraser River at hope



(*а*) Квадратичная функция потерь

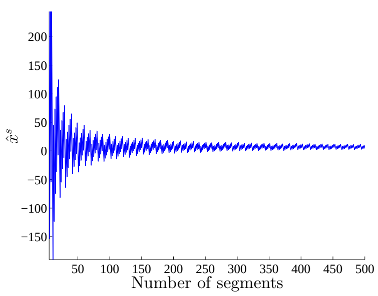


(*б*) Абсолютная функция потерь

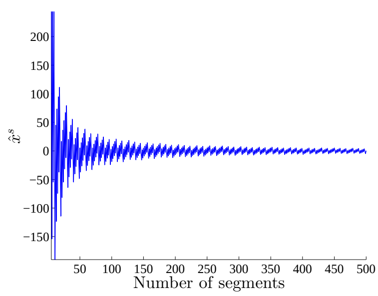


(*в*) Асимметричная функция потерь

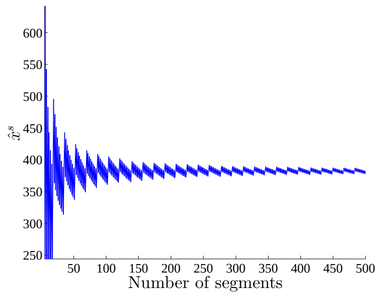
Рис. 4 Прогнозы алгоритма Hist для регрессионных остатков ряда Monthly Lake Erie Levels



(*а*) Квадратичная функция потерь

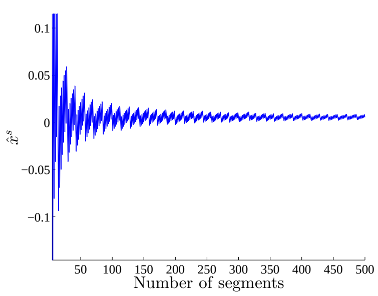


(*б*) Абсолютная функция потерь

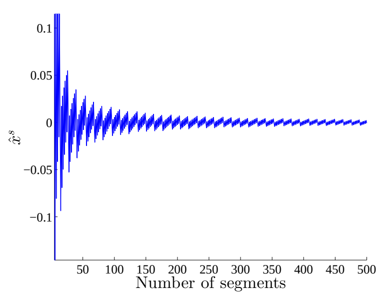


(*в*) Асимметричная функция потерь

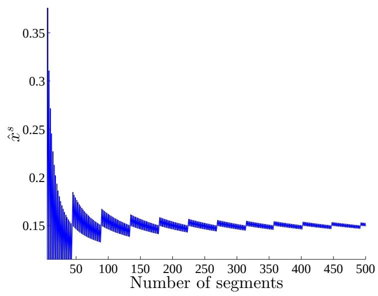
Рис. 5 Прогнозы алгоритма Hist для регрессионных остатков ряда Monthly production of chocolate confectionery in Australia



(*а*) Квадратичная функция потерь

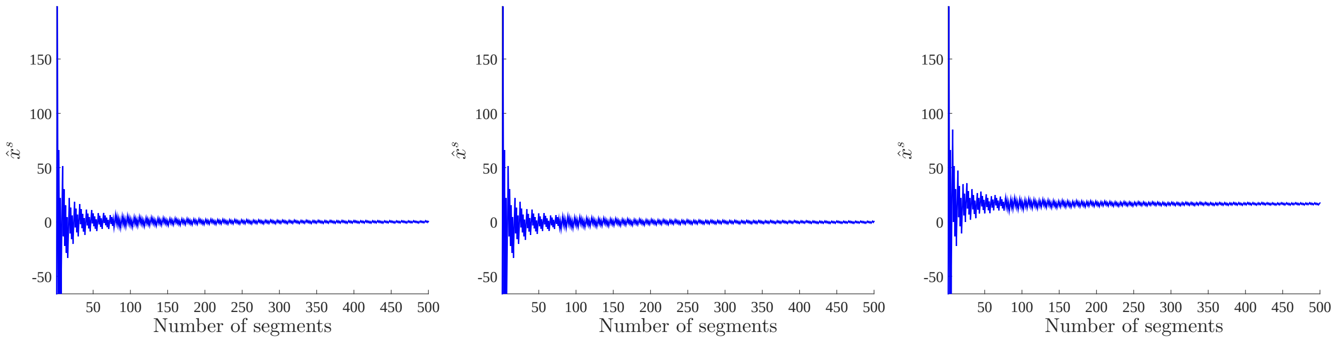


(*б*) Абсолютная функция потерь



(*в*) Асимметричная функция потерь

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



(*а*) Квадратичная функция потерь

(*б*) Абсолютная функция потерь

(*в*) Асимметричная функция потерь

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

функции потерь (7)–(9), одна из которых несимметричная. Получено экспериментальное подтверждение того, что при несимметричных потерях использование двухэтапного прогнозирования позволяет уменьшить средние потери.

Для каждого временного ряда, изображенного на рис. 1, подбиралась модель ARIMA по методологии Бокса–Дженкинса [2]. Выбранные модели для каждого временно́го ряда показаны в табл. 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 временно́го ряда использовались как контрольные. Для каждой контрольной точки по доступной истории временно́го ряда (все точки от первой до предшествующей рассматриваемой контрольной) обучалась выбранная для временно́го 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 (7)–(9) и вариантов алгоритма Hist с 20, 50, 300 и 500 столбцами в гистограмме. Средние потери для каждой функции потерь приведены для всех вариантов алгоритма в табл. 2.

Как видно из табл. 2, 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 составляют только средние потери прогнозирования для временно́го ряда Monthly production of chocolate confectionery in Australia. Для этого временно́го ряда

Таблица 1 Выбранные модели ARIMA для прогноза нестационарной компоненты

Временной ряд	Модель ARIMA
Fraser River at hope	ARIMA$(1,0,0) \times (1,0,1)_{12}$
Monthly Lake Erie Levels	ARIMA$(2,0,0) \times (1,0,1)_{12}$
Chocolate production in Australia	ARIMA$(1,1,1) \times (1,0,1)_{12}$
Sugar price	ARIMA$(1,0,0)$
Electricity consumption	ARIMA$(2,1,2) \times (1,0,1)_{24} \times (1,0,1)_{168}$

Таблица 2 Средние потери прогнозирования

Ряд	Алгоритм	**Квадратичная**	**Абсолютная**	**Асимметричная**
River	No Hist	52400	498	616
	Hist(20)	53500	495	523
	Hist(50)	**52200**	496	516
	Hist(300)	52500	493	516
	Hist(500)	52400	**492**	**515**
Lake	No Hist	0,172	0,313	0,410
	Hist(20)	0,182	0,316	0,315
	Hist(50)	0,171	**0,313**	0,312
	Hist(300)	0,171	0,314	0,311
	Hist(500)	**0,171**	0,314	**0,311**
Chocolate	No Hist	71500000	8350	4180
	Hist(20)	66000	612	579
	Hist(50)	65800	609	575
	Hist(300)	65300	610	575
	Hist(500)	**65100**	**609**	**575**
Sugar	No Hist	0,127	0,265	0,340
	Hist(20)	0,128	0,267	**0,260**
	Hist(50)	0,127	0,266	0,267
	Hist(300)	0,127	0,265	0,266
	Hist(500)	**0,127**	**0,265**	0,266
Electricity	No Hist	500	16,9	19,9
	Hist(20)	717	19,1	**12,7**
	Hist(50)	589	**16,5**	13,4
	Hist(300)	**498**	17,1	13,0
	Hist(500)	502	16,9	13,0

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, *в*, этот временной ряд во второй половине истории имеет более высокий темп роста, чем в первой половине. Обученная преимущественно по 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 удается оценить это смещение и сделать более точный прогноз.

8 Заключение

Предложен двухэтапный алгоритм прогнозирования нестационарных временны́х рядов ARIMA + Hist, минимизирующий ожидаемые потери. Он не накладывает на вид функции потерь ограничений — она может быть симметричной или несимметричной, дифференцируемой или нет. На первом этапе строится прогноз нестационарной компоненты временно́го 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 временнóго 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 временнóго ряда.

Литература

[1] Patton A. J., Timmermann A. Properties of optimal forecasts under asymmetric loss and nonlinearity // J. Econometrics, 2007. Vol. 140. No. 2. P. 884–918. doi: http://dx.doi.org/10.1016/j.jeconom.2006.07.018

[2] Box G. E. P., Jenkins G. M., Reinsel G. C. Time series analysis: Forecasting and control. — 3rd ed. — Englewood Cliffs, NJ: Prentice Hall, 1994.

[3] Berk R. Asymmetric loss functions for forecasting in criminal justice settings // J. Quantitative Criminology, 2011. Vol. 27. No. 1. P. 107–123. doi: http://dx.doi.org/10.1007/s10940-010-9098-2

[4] Cipra, T. Asymmetric recursive methods for time series // Appl. Math., 1994. Vol. 39. No. 3. P. 203–214.

[5] Koenker R., Xiao Z. Quantile autoregression // J. Am. Stat. Association, 2006. Vol. 101. No. 475. P. 980–990. doi: http://dx.doi.org/10.1198/016214506000000672

[6] Koenker R. Quantile regression. — Cambridge University Press, 2005. doi: http://dx.doi.org/10.1017/CBO9780511754098

[7] Christoffersen P. F, Diebold F. X. Optimal prediction under asymmetric loss // Econometric Theory, 1997. Vol. 13. No. 06. P. 808–817. doi: http://dx.doi.org/10.1017/S0266466600006277

[8] Granger C. W. J. Prediction with a generalized cost of error function // OR, 1969. Vol. 20. No. 2. P. 199–207. doi: http://dx.doi.org/10.2307/3008559

[9] Christoffersen P. F, Diebold F. X. Further results on forecasting and model selection under asymmetric loss // J. Appl. Econometrics, 1996. Vol. 11. No. 5. P. 561–571. doi: http://doi.org/bs5mh8

[10] Diebold F. X., Gunther T., Tay A. Evaluating density forecasts // Int. Econ. Rev., 1998. Vol. 39. P. 863–883. doi: http://dx.doi.org/10.2307/2527342

[11] Biau G., Bleakley k., Györfi L., Ottucsák G. Nonparametric sequential prediction of time series // J. Nonparametric Stat., 2010. Vol. 22. No. 3. P. 297–317.

[12] Hyndman R. J., Athanasopoulos G. Forecasting: Principles and practice. — OTexts, 2006. https://www.otexts.org/book/fpp.

[13] Monthly production of chocolate confectionery in Australia. https://datamarket.com/data/set/22rl/monthly-production-of-chocolate-confectionery-in-australia-tonnes-july-1957-aug-1995#!ds=22rl&display=line.

[14] Fraser River at hope. https://datamarket.com/data/set/22nm/fraser-river-at-hope-1913-1990#!ds=22nm&display=line.

[15] Monthly Lake Erie Levels. https://datamarket.com/data/set/22pw/monthly-lake-erie-levels-1921-1970#!ds=22pw&display=line.

[16] Sugar price. https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSForecasting/TimeSeries/Sources/tsSugarPrice.csv.

[17] Electricity consumption. https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSForecasting/TimeSeries/Sources/tsEnergyConsumption.csv.

Поступила в редакцию 30.08.15

Forecasting nonstationary time series under asymmetric loss*

***V. Y. Chernykh*[1] *and M. M. Stenina*[1,2]**

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

[1]Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, Russia

[2]Higher School of Economics, 22 Myasnitskaya st., Moscow, Russia

The problem of forecasting time series under asymmetric loss functions is considered. A new two-step forecasting algorithm ARIMA + Hist is presented. At the first step, autoregression integrated moving average algorithm ARIMA with seasonal components is used. The parameters of the model are selected according to Box–Jenkins methodology. At the second step, the residuals are analyzed 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 the histogram of regression residuals. The performance of the algorithm is demonstrated 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*

DOI: 10.21469/22233792.1.14.01

References

[1] Patton, A. J., and A. Timmermann. 2007. Properties of optimal forecasts under asymmetric loss and nonlinearity. *J. Econometrics* 140(2):884–918. doi: http://dx.doi.org/10.1016/j.jeconom.2006.07.018

[2] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control.* 3rd ed. Englewood Cliffs: NJ: Prentice Hall.

[3] Berk, R. 2011. Asymmetric loss functions for forecasting in criminal justice settings. *J. Quantitative Criminology* 27(1):107–123. doi: http://dx.doi.org/10.1007/s10940-010-9098-2

[4] Cipra, T. 1994. Asymmetric recursive methods for time series. *Appl. Math.* 39(3):203–214.

[5] Koenker, R., and Z. Xiao. 2006. Quantile autoregression. *J. Am. Stat. Association* 101(475):980–990. doi: http://dx.doi.org/10.1198/016214506000000672

[6] Koenker, R. 2005. *Quantile regression.* Cambridge University Press. doi: http://dx.doi.org/10.1017/CBO9780511754098

*This work was done under financial support of the Russian Foundation for Basic Research (grant 14-07-31046)

[7] Christoffersen, P. F, and F. X. Diebold. 1997. Optimal prediction under asymmetric loss. *Econometric Theory* 13(6):808–817. doi: http://dx.doi.org/10.1017/S0266466600006277

[8] Granger, C. W. J. 1969. Prediction with a generalized cost of error function. *OR* 20(2):199–207. doi: http://dx.doi.org/10.2307/3008559

[9] Christoffersen, P. F, and F. X. Diebold. 1996. Further results on forecasting and model selection under asymmetric loss. *J. Appl. Econometrics* 11(5):561–571. doi: http://doi.org/bs5mh8

[10] Diebold, F. X., T. Gunther, and A. Tay. 1998. Evaluating density forecasts. *Int. Econ. Rev.* 39:863–883. doi: http://dx.doi.org/10.2307/2527342

[11] Biau, G., K. Bleakley, L. Györfi, and G. Ottucsák. 2010. Nonparametric sequential prediction of time series. *J. Nonparametric Stat.* 22(3):297–317.

[12] Hyndman, R. J., and G. Athanasopoulos. 2006. *Forecasting: Principles and practice*. OTexts. Available at: https://www.otexts.org/book/fpp (accessed December 29, 2015).

[13] Monthly production of chocolate confectionery in Australia. Available at: https://datamarket.com/data/set/22rl/monthly-production-of-chocolate-confectionery-in-australia-tonnes-july-1957-aug-1995#!ds=22rl&display=line (accessed December 29, 2015).

[14] Fraser River at hope. Available at: https://datamarket.com/data/set/22nm/fraser-river-at-hope-1913-1990#!ds=22nm&display=line (accessed December 29, 2015).

[15] Monthly Lake Erie Levels. Available at: https://datamarket.com/data/set/22pw/monthly-lake-erie-levels-1921-1970#!ds=22pw&display=line (accessed December 29, 2015).

[16] Sugar price. Available at: https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSForecasting/TimeSeries/Sources/tsSugarPrice.csv (accessed December 29, 2015).

[17] Electricity consumption. Available at: https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSForecasting/TimeSeries/Sources/tsEnergyConsumption.csv (accessed December 29, 2015).

Received August 30, 2015