

УДК 519.25

Научная статья

<https://doi.org/10.23947/2587-8999-2023-6-1-83-89>**Прогнозирование тенденций развития временных рядов с помощью оценки медианы Ходжеса-Лемана****В. В. Мисюра<sup>1</sup>  , Е. В. Мисюра<sup>2</sup> **<sup>1</sup>Донской государственный технический университет, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1<sup>2</sup>Российский экономический университет им. Плеханова, Институт цифровой экономики и информационных технологий, Российская Федерация, г. Москва, Стремянный переулок, 36 [vvmisyura2011@gmail.com](mailto:vvmisyura2011@gmail.com)**Аннотация**

**Введение.** Получение достоверных прогнозов, как в тенденции развития процесса, так и в его численном выражении является актуальной научной проблемой. Отсутствие стационарности в развитии временных рядов, малые выборки, наличие выбросов, скачков не позволяют найти оценки параметров модели, обладающих хорошими свойствами статистических оценок. Одним из средств, позволяющим решить указанные проблемы, является применение порядковых или робастных статистик. Научные монографии и специальные главы в книгах по математической статистике содержат глубокую и обширную теорию по исследованию свойств порядковых статистик, которые являются обоснованием их применения в методах прогнозирования. Целью настоящей работы является разработка и верификация метода получения одношаговых прогнозов тенденций развития временных рядов, основанного на устойчивых статистических оценках.

**Материалы и методы.** Разрабатывается метод получения одношаговых прогнозов тенденций развития временных рядов, основанных на построении доверительных интервалов выборочной устойчивой оценки Ходжеса-Лемана по средним Уолша. В частности, с помощью медианы Ходжеса-Лемана решается проблема малых выборок, полученных при выполнении процедуры сдвига окна временного ряда. Подробно рассматривается предложенный метод: даются основные определения, теоретическое обоснование метода, формулы расчета, подробное описание алгоритма, приводятся формулы расчета метрики качества прогнозов.

**Результаты исследования.** Метод получил реализацию в вычислительном эксперименте на примере прогнозирования спотовой цены на нефть марки Urals. Приведены результаты вычислительного эксперимента. Параметры предложенного метода могут быть настроены так, чтобы получать достоверные одношаговые прогнозы.

**Обсуждение и заключения.** Предложенный метод показал свою эффективность на экспериментальных данных и может быть использован как самостоятельный способ построения одношаговых прогнозов тенденций развития временных рядов. Дальнейшее развитие метода предполагает усовершенствование вычислительных процедур и верификацию в случае наличия скачков в динамике временного ряда

**Ключевые слова:** порядковая статистика, случайная величина, случайный процесс, медиана, медиана Ходжеса-Лемана, прогнозирование, тренд, средняя абсолютная ошибка.

**Для цитирования.** Мисюра, В. В. Прогнозирование тенденций развития временных рядов с помощью оценки медианы Ходжеса-Лемана / В. В. Мисюра, Е. В. Мисюра // Computational Mathematics and Information Technologies. — 2023. — Т. 6, № 1. — С. 83–89. <https://doi.org/10.23947/2587-8999-2023-6-1-83-89>

Original article

**Forecasting trends in the development of time series by estimating the Hodges-Lehman median****V. V. Misyura<sup>1</sup>  , E. V. Misyura<sup>2</sup> **<sup>1</sup>Don State Technical University, 1, Gagarin Sq., Rostov-on-Don, Russian Federation<sup>2</sup>Russian University of Economics named after Plekhanova, Institute of Digital Economy and Information Technologies, 36, Stremyanny Lane, Moscow, Russian Federation [vvmisyura2011@gmail.com](mailto:vvmisyura2011@gmail.com)

## Abstract

**Introduction.** The lack of stationarity in the time series' development, small samples, the outliers presence, jumps do not allow to find estimates of model parameters that have good properties of statistical estimates and, as a result, to find reliable forecasts, both in the development trend of the process and in its numerical expression. The means to solve these problems is the use of ordinal or robust statistics. Scientific monographs and special chapters in books on mathematical statistics contain a deep and extensive theory on the study of the properties of ordinal statistics, which are the justification for their application in forecasting methods. The aim of the work is to develop and verify method for obtaining one-step forecasts of trends in the development of time series based on stable statistical estimates.

**Materials and methods.** The article presents the results of the development of a method for obtaining one-step forecasts of trends in the development of time series based on the construction of confidence intervals of a selective stable Hodges-Lehman estimate based on Walsh averages. In particular, the Hodges-Lehman median is used to solve the problem of small samples obtained during the procedure of shifting the time series window. The proposed method is considered in detail in the article: the basic definitions, the theoretical justification of the method, calculation formulas, a detailed description of the algorithm, formulas for calculating the quality metric of forecasts are given.

**The results of the study.** The method was implemented in computational experiment using the example of forecasting the Urals crude oil's spot price. The article presents the computational experiment's results. The parameters of the proposed method can be configured to obtain reliable one-step forecasts.

**Discussion and conclusions.** The method proposed in the article has shown its effectiveness on experimental data and can be used as an independent method for constructing one-step forecasts of trends in the development of time series. Further development of the method involves the improvement of computational procedures, verification of the method in case of jumps in the dynamics of the time series.

**Keywords:** ordinal statistics, random variable, random process, median, Hodges-Lehman medians, forecasting, trend, average absolute error.

**For citation.** Misyura, V. V. Forecasting trends in the development of time series by estimating the Hodges-Lehman median / V. V. Misyura, E. V. Misyura. // Computational Mathematics and Information Technologies. — 2023. — Vol. 6, no. 1. — P. 83–89. <https://doi.org/10.23947/2587-8999-2023-6-1-83-89>

**Введение.** Существует большое количество методов и способов прогнозирования, в которых используются математические модели случайного процесса, и статистические оценки параметров модели, с помощью которых модель настраивается на конкретную реализацию. Однако отсутствие стационарности в развитии процесса, делает невозможным использование одной модели на протяжении всего периода прогнозирования. К нестационарности следует добавить малые выборки, наличие выбросов, недостаточность априорной информации. Все это не позволяет найти оценки параметров модели, обладающих хорошими свойствами статистических оценок. Такие модели, в которых параметры недостаточно хорошо определены, называют моделями с неопределенными параметрами. Одним из средств, позволяющим в некоторой степени снять неопределенность в случаях применения малых и засоренных выборок при настройке параметров модели, являются робастные статистики. Под робастными статистиками обычно понимаются порядковые и знаковые статистики.

Порядковые статистики широко используются в статистической практике. Для них разработана глубокая и обширная теория, как при конечных объемах выборки, так и в асимптотике. На использовании порядковых статистик основаны многие статистические процедуры оценивания параметров и проверки статистических гипотез. Порядковым статистикам посвящены научные монографии и специальные главы в книгах по математической статистике [1–3]. Предложенный подход к прогнозированию тенденции развития случайных процессов во времени обоснован свойствами порядковых статистик.

Исследование посвящено анализу временных рядов с целью предсказания тенденции их развития во времени. В статье развивается идея прогнозирования тенденций развития временных рядов, предложенная в [4] и основанная на применении оценок порядковых статистик. В работе [4] сделана успешная попытка прогнозирования тен-

денции поведения финансовых индексов с помощью робастных статистик, а именно использование медианных оценок для вычисления меры рассеяния вокруг центра выборки. В статье предлагается решить проблемы подбора весовых коэффициентов и малых выборок при выполнении процедуры сдвига окна временного ряда с помощью построения доверительных интервалов выборочной устойчивой оценки Ходжеса-Лемана по средним Уолша.

### Материалы и методы.

#### Метод прогнозирования тенденции развития случайного процесса во времени.

Будем рассматривать временной ряд как реализацию случайного процесса  $\{X_t, 0 \leq t \leq T\}$ . Случайную последовательность  $\{h_t \leq T\}$  получаем, используя формулу,  $h_t = \ln(X_t/X_{t-1})$ , где  $X_t$  — наблюдаемый уровень временного ряда. Предположим, что  $X_{i+1}$ ,  $i=1, 2, \dots, T$  может иметь три типа направления изменения числового значения: падение  $(-1)$ ; сохранение  $(0)$ ; рост тренда  $(1)$ .

Функция, которая позволит прогнозировать тенденцию поведения временного ряда на один временной период, строится по следующей формуле:

$$\text{Trend}_i = \begin{cases} -1, \theta_+^i < 0, \\ 0, (\theta_+^i > 0) \& (\theta_-^i < 0), \\ 1, \theta_-^i > 0, \end{cases} \quad (1)$$

где  $\theta_-^i$  и  $\theta_+^i$  — пороговые переменные функции.

Пороговые переменные  $\theta_-^{i+\tau}$  и  $\theta_+^{i+\tau}$  для вычисления одношагового прогноза тенденции развития временного ряда предлагается вычислять как границы доверительного интервала выборочной порядковой статистики случайных последовательностей  $\{h_p, h_{i+1}, \dots, h_{i+\tau-1}\}$ ,  $i=1, \dots, T-\tau+1$ , где некоторое число  $1 < \tau < T$  определяет ширину окна процедуры сдвига временного ряда  $h_k$ ,  $k=1, \dots, T$ .

Значения случайных величин  $H_{(1)}, H_{(2)}, \dots, H_{(n)}$  называют порядковыми статистиками, если они соответствуют случайным величинам, входящим в выборку  $H_1, H_2, \dots, H_n$ , и расположенным в порядке возрастания их значений  $h_{(1)} \leq h_{(2)} \leq \dots \leq h_{(n)}$ . Статистические свойства порядковых статистик, как при конечных и малых объемах выборки, так и при больших объемах выборки (то есть в асимптотике), подробно изучены и описаны в литературе [3]. Обоснованием использования доверительных границ порядковых статистик является приведенное ниже Утверждение 1, доказательство которого следует из [4, с. 94].

Рассмотрим общий случай. Пусть  $\{H_{(1)}, H_{(2)}, \dots, H_{(n)}\}$  — порядковые статистики для выборки  $\{H_1, H_2, \dots, H_n\}$ . Обозначим квантиль уровня  $p$  через  $h_p = F^{-1}(p)$ ,  $0 < p < 1$ , где  $F_H(h)$  — неизвестная функция распределения наблюдаемой случайной величины. Справедлива следующая теорема.

**Утверждение 1.** Пусть имеется два числа  $r$  и  $s$  такие что,  $P(H_{(r)} < h_p < H_{(s)}) = 1 - 2\alpha$ . Заданная доверительная вероятность и интервал  $(H_{(r)}, H_{(s)})$  со случайными границами  $H_{(r)}$  и  $H_{(s)}$  включает неизвестный квантиль  $h_p = F^{-1}(p)$ ,  $0 < p < 1$ . Тогда вероятность  $P(H_{(r)} < h_p < H_{(s)})$  не зависит от неизвестной функции распределения  $F_H(h)$ .

Выберем в качестве порядковой статистики для выборки  $(h_i^{i+\tau-1} = \{h_i, h_{i+1}, \dots, h_{i+\tau-1}\})$  выборочную медиану. Тогда в качестве интервальной оценки для  $Me(h_i^{i+\tau-1})$  можно выбрать двумерную статистику вида  $(h_p, h_k)$ ,  $i \leq l < k \leq i + \tau - 1$ ,  $i=1, \dots, N-\tau+1$ , определяющую симметричный интервал с уровнем доверия  $1 - 2\alpha$ , полагая  $k = \tau - l - 1 + i$ .

Поскольку

$$P(h_i < Me(h_i^{i+\tau-1}) < h_k) = (1/2)^\tau \sum_{t=1}^{\tau-1} C_t^i = 1 - 2\alpha, \quad (2)$$

значение  $l$ , согласно теореме Муавра-Лапласа, можно рассчитать по формуле:

$$l = \left[ 0,5 \left\{ \tau + 1 - \sqrt{\tau} \Psi(1 - \alpha) \right\} + 1 \right] + i, \quad i=1, \dots, N - \tau + 1 \quad (3)$$

Заметим, что выборочная медиана вычисляется по формуле:

$$\text{med}(H_1, H_2, \dots, H_n) = \begin{cases} \frac{1}{2} \left( H_{(\frac{n}{2})} + H_{(\frac{n}{2}+1)} \right), & \text{если } n/2 \text{ — целое,} \\ H_{(\frac{n+1}{2})}, & \text{если } (n+1)/2 \text{ — целое,} \end{cases} \quad (4)$$

где  $H_{(i)}$  —  $i$ -я порядковая статистика, равная  $i$ -му по величине значению выборочного ряда  $h_1 \leq h_2 \leq \dots \leq h_n$ , ранжированного по возрастанию.

Процедура сдвига временного ряда с шириной окна  $1 < \tau < T$  может привести к проблеме вычисления робастной оценки медианы, поскольку оценки вычисляются по малым выборкам. В этом случае оценка медианы может быть нестабильной и не очень точной.

Предлагается строить доверительный интервал медианы выборок с помощью устойчивой оценки Ходжеса-Лемана по средним Уолша.

Порядковая статистика Ходжеса-Лемана по средним Уолша рассчитывается по формуле [2, с. 103]:

$$H_{H-L} = \text{med}[(H_{(i)} + H_{(j)}) / 2], \quad 1 \leq i \leq j \leq \tau. \quad (5)$$

В формуле (5) вычисляется медиана по формуле (4) из набора  $n(n-1)/2$  средних Уолша  $(H_{(i)} + H_{(j)})/2$ ,  $1 \leq i \leq j \leq \tau$ . Свойства медианы Ходжеса-Лемана  $H_{H-L}$  хорошо изучены и подробно описаны в литературе [2, 5, 6]. Следует отметить высокую устойчивость этой оценки к отклонениям от нормальности распределения и засоренности выборки.

Таким образом, применение оценки Ходжеса-Лемана приводит к вычислению выборочной медианы на выборках, объем которых превышает 30, если ширина окна  $\tau \geq 9$ .

Для решения задачи прогнозирования тенденции развития случайных процессов во времени предлагается следующий алгоритм.

#### 1. Инициализация.

Получаем случайную последовательность  $\{h_t | t \leq T\}$ , используя формулу  $h_t = \ln(X_t / X_{t-1})$ , где  $X_t$ ,  $t = 1, \dots, T$  — уровень наблюдаемого временного ряда.

#### 2. Итерация.

Пусть некоторое четное число  $1 < \tau < T$  определяет ширину окна процедуры сдвига временного ряда  $h_k$ ,  $k = 1, \dots, T$ . Случайные последовательности  $(h)_i^{i+\tau-1} = \{h_i, h_{i+1}, \dots, h_{i+\tau-1}\}$ , при сдвиге  $\{h_k\}$  по  $i = 1, \dots, T - \tau + 1$  на один шаг упорядочиваем по возрастанию. Для последовательности  $(h)_i^{i+\tau-1}$  по формулам (2)–(5) строим доверительный интервал оценки Ходжеса-Лемана с заданным уровнем значимости  $\alpha$ . Пороговые переменные  $\theta_-^i$  и  $\theta_+^i$  получают значение левой и правой границы доверительных интервалов соответственно.

Вычисляем показатели тренда по формуле (1).

Для вычисления ошибки прогноза предложенного метода для  $i = 1, \dots, T - \tau + 1$  получаем фактические значения тренда динамики временного ряда  $X_t$ ,  $t = 1, \dots, T$  по следующим формулам:

$$\text{Trend}_i^{fk} = \begin{cases} -1, & X_i - X_{i-1} < -\sigma_i \\ 0, & (X_i - X_{i-1} \geq -\sigma_i) \& (X_i - X_{i-1} \leq \sigma_i) \\ 1, & X_i - X_{i-1} > \sigma_i, \end{cases} \quad (6)$$

где  $\sigma_i = \sqrt{Me(\sigma^2)_i^{i+\tau-1}}$ ,  $Me(\sigma^2)_i^{i+\tau-1}$  — медиана случайной последовательности

$$(\sigma^2)_i^{i+\tau-1} = \left\{ \left( X_i - Me(X)_i^{i+\tau-1} \right)^2, \dots, \left( X_{i+\tau-1} - Me(X)_i^{i+\tau-1} \right)^2 \right\}, \quad i = 1, \dots, T - \tau + 1.$$

#### 3. Остановка.

Таким образом, после реализации описанного выше алгоритма получаем две последовательности  $\{\text{Trend}_i\}$ ,  $\{\text{Trend}_i^{fk}\}$  размерности  $1, \dots, T - \tau + 1$ , состоящие из элементов множества:  $\{-1, 0, 1\}$ .

Теперь можно построить матрицу ошибок, которая показывает, насколько хорошо алгоритм определяет тренд динамики уровней временного ряда. Матрица ошибок используется для оценки качества работы алгоритма.

|         |    | -1        | 1        | 0        |
|---------|----|-----------|----------|----------|
| Прогноз | -1 | TP(-1,-1) | FN(-1,0) | FN(-1,1) |
|         | 0  | FN(0,-1)  | TP(0,0)  | FN(0,1)  |
|         | 1  | FN(1,-1)  | FN(1,0)  | TP(1,1)  |

Здесь TP  $(\hat{\xi}, \xi)$  — это количество прогнозов, которые верно определили оценку тенденции  $\xi$ , FN  $(\hat{\xi}, \xi)$  — это количество прогнозов, которые неверно определили направление динамики временного ряда.

Качество полученных прогнозов можно определить отношением:

$$\text{accuracy} = \sum_{\xi=-1}^1 TP(\hat{\xi}, \xi) / (T - \tau + 1). \quad (7)$$

Также можно рассчитать метрики качества предложенного метода для каждого из значений функции  $\{\text{Trend}_i^{fk}\}$ . Например, точность для прогнозирования падений и роста можно вычислить по формулам:

$$\text{accuracy}(1) = TP(1,1) / (TP(1,1) + FN(1,0) + FN(1,-1)). \quad (8)$$

$$\text{accuracy}(-1) = TP(-1,-1) / (TP(-1,-1) + FN(-1,0) + FN(-1,1)). \quad (9)$$

Средняя абсолютная ошибка (MAE, Mean Absolute Error) прогноза также может быть показателем качества предложенного метода получения одношаговых прогнозов тенденции развития временного ряда.

$$\text{MAE} = \frac{1}{T - \tau + 1} \sum_{i=1}^{T-\tau+1} |\text{Trend}_i - \text{Trend}_i^{fk}|. \quad (10)$$

**Результаты исследования.** Для анализа и апробации описанного метода был выбран логарифм изменения спотовой цены на нефть марки Urals. Для исследования были взяты цены закрытия торгов на Московской фондовой бирже за период с 12.07.2018 г. по 23.02.2022 г. Объем выборки составил 946 значений. Вычислительный эксперимент проводился с помощью среды статистического программирования R и средств прогнозирования и анализа данных MS Excel.

Пропуски данных в те дни, когда торги не ведутся (праздники и выходные), были восстановлены при соблюдении следующих правил:

- если был пропущен один день, то ему присваивается показатель закрытия торгов предыдущего дня;
- если пропущено два дня, то первый день восстанавливается согласно предыдущему пункту, второй согласно значению следующего дня;
- если пропущено три дня, первым двум дням присваивается значение предыдущего, а третий день восстанавливается по значению закрытия торгов следующего дня;
- если пропущено 4 дня, первые два дня заполняются значением за предыдущий день, третий и четвертый – значением за следующий день;
- если пропущено более 4 дней (в данном исследовании нет таких длинных пропусков) данные восстанавливались бы с помощью линейной (или нелинейной) регрессии от времени.

Получившийся ряд содержит 1323 значения. График временного ряда представлен на рис. 1:

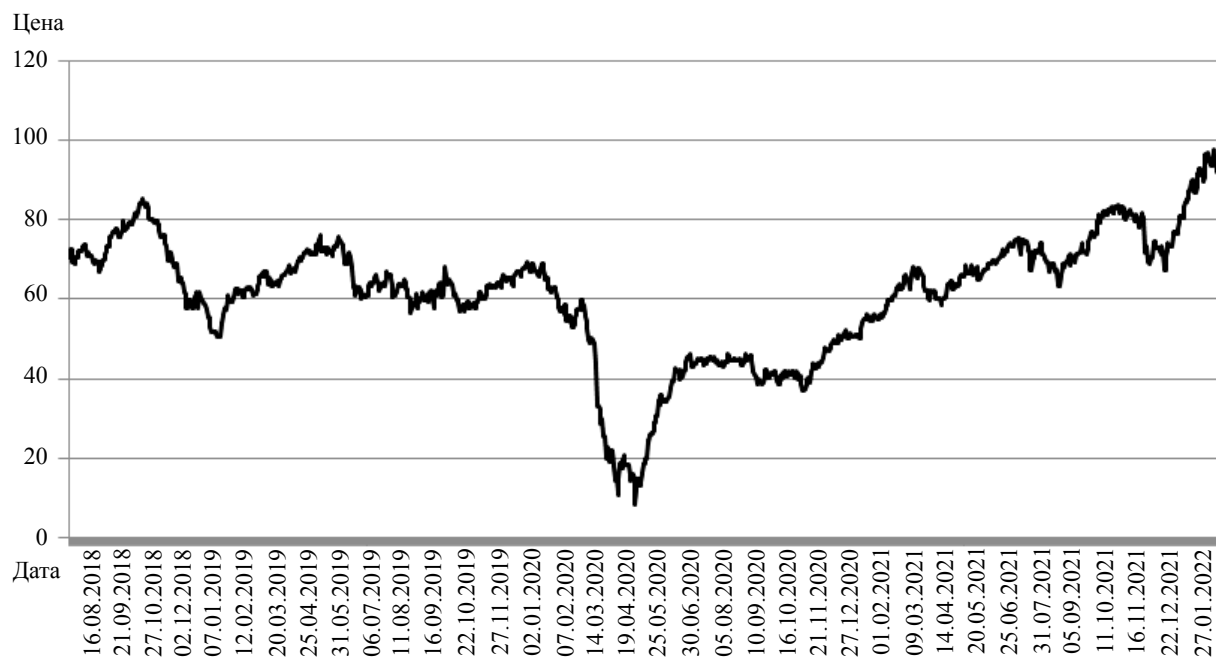


Рис. 1. Динамика спотовой цены на нефть марки Urals с 12.07.2018 г. по 23.02.2022 г.

Логарифм изменения спотовой цены на нефть марки Urals представлен на рис. 2.

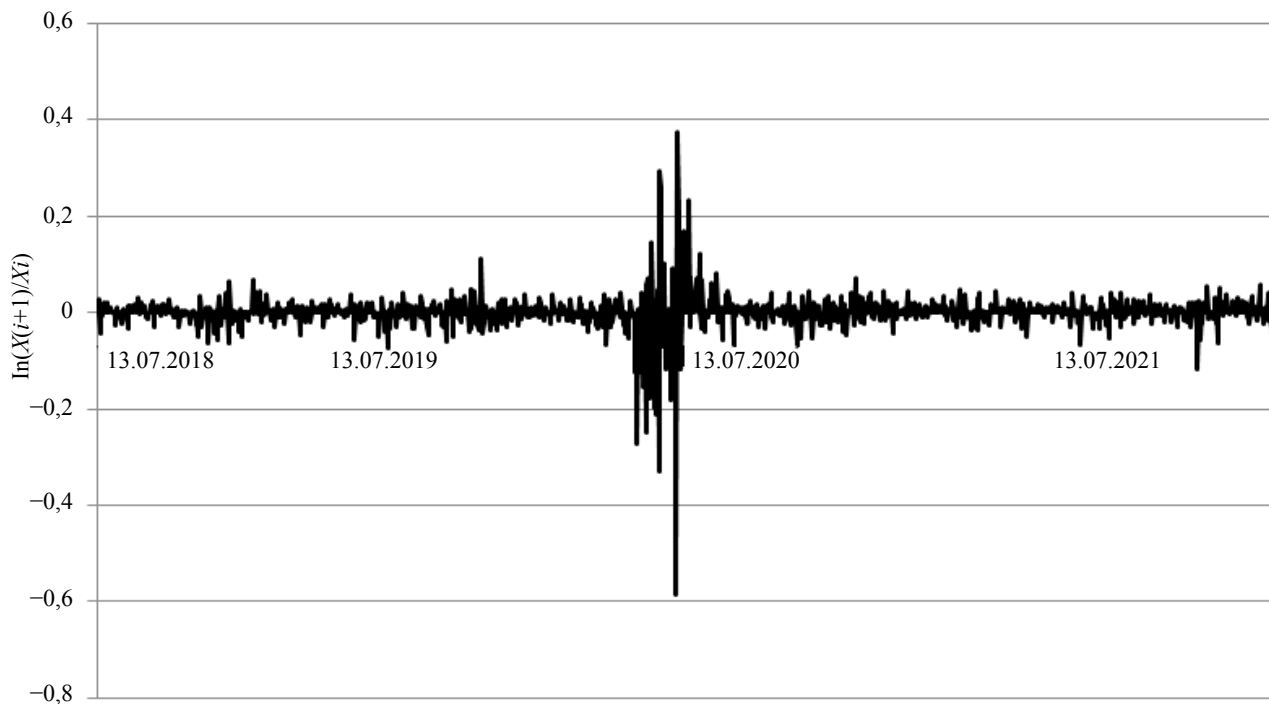


Рис. 2. Логарифм изменения спотовой цены на нефть марки Urals с 13.07.2018 по 23.02.2022

Чтобы получить одношаговый прогноз тенденции развития временного ряда, выделим временной период с 23.05.2020 г. по 23.02.2022 г., соответствующий стационарному случайному процессу  $h_k$ ,  $k=1, \dots, 643$ .

Реализация алгоритма возможна при различных значениях  $\tau \geq 9$ .

На рис. 3 показан график зависимости ошибки MAE от ширины окна.

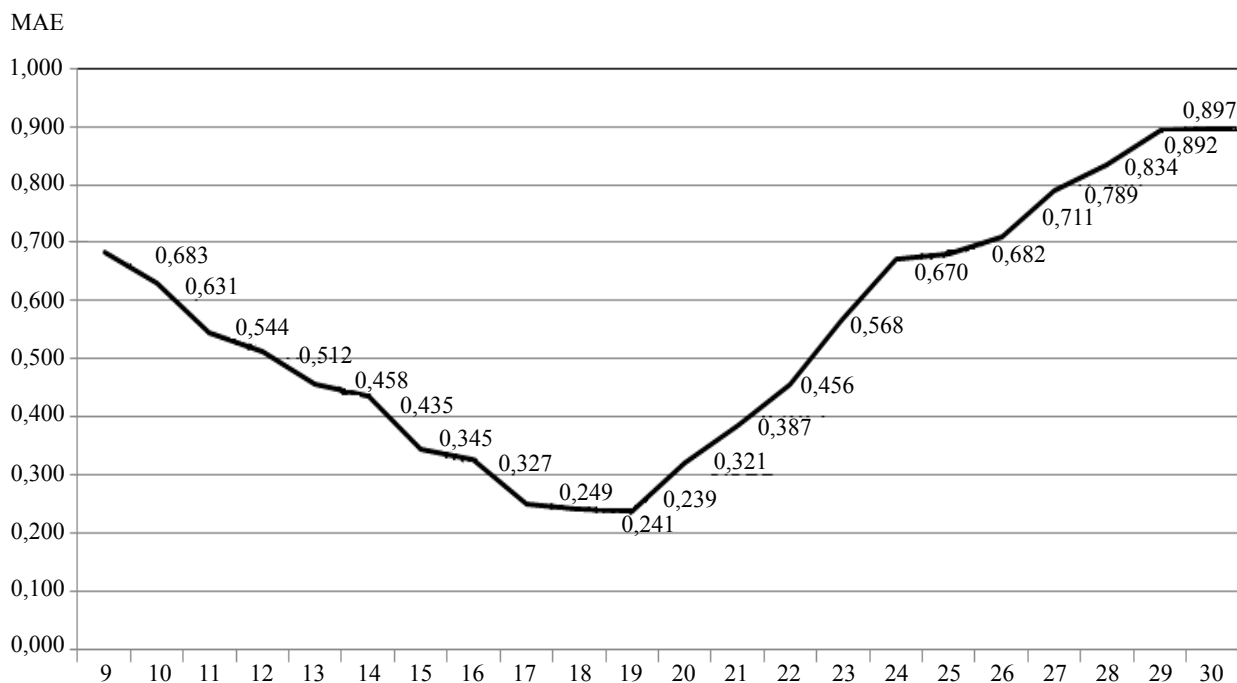


Рис. 3. График зависимости средней абсолютной ошибки прогноза от ширины окна

Матрица ошибок для  $\tau = 9$  выглядит следующим образом.

|         |    |    |    |     |
|---------|----|----|----|-----|
| Прогноз |    | -1 | 0  | 1   |
|         | -1 | 71 | 35 | 46  |
|         | 0  | 50 | 97 | 56  |
|         | 1  | 70 | 60 | 149 |

Тогда нетрудно вычислить точность построенного прогноза  $\text{ассигассу} = 317/634=0,5$  и точность для прогнозов различных направлений динамики цены нефти:  $\text{ассигассу}(-1)=71/152=0,467$ ,  $\text{ассигассу}(0)=97/203=0,478$ ,  $\text{ассигассу}(1)=149/279=0,53$ . Самая высокая точность алгоритма для выбранного набора данных соответствует прогнозу роста цены нефти, хотя существенной разницы для падения и сохранения тенденции не наблюдается.

**Обсуждение и заключения.** Из результатов вычислительного эксперимента видно, что параметры предложенного метода могут быть настроены так, чтобы получать достоверные одношаговые прогнозы направления развития временных рядов. С прохождением ширины окна некоторого порогового значения ошибка прогноза начинает расти. Основная вычислительная сложность алгоритма заключалась как в получении непосредственных оценок Хорджеса-Лемана, так и определения границ доверительных интервалов. Предложенный в статье метод прогноза основанный на устойчивых порядковых статистиках показал свою эффективность на экспериментальных данных и может быть использован как самостоятельный метод построения одношаговых прогнозов. Дальнейшее развитие предложенного метода предполагает усовершенствования вычислительных процедур, верификацию метода в случае наличия в случайной последовательности выбросов и скачков.

### Список литературы

1. Дейвид, Г. Порядковые статистики / Г. Дейвид. — Москва : Наука. Главная редакция физико-математической литературы, 1979. — 336 с.
2. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. — Москва : ФИЗМАТЛИТ, 2006. — 816 с.
3. Сархан, А. Е. Введение в теорию порядковых статистик / А. Е. Сархан, В. Гринберг / пер. с англ. под ред. Ф.Я. Боярского. — Москва : Статистика, 1970 г. — 414 с.
4. Мисюра, В. В. Прогнозирование тенденций развития случайных процессов с помощью порядковой статистики/ В. В. Мисюра // Инженерный журнал Дона. — 2018, № 4(51). — с. 1155.
7. Шуленин, В. П. Математическая статистика: Непараметрическая статистика / В. П. Шуленин / Ч. 2. — Томск : Изд-во НТЛ, 2012 г.
6. Hodges, J. L. Estimation of location based on rank tests / J. L. Hodges, E. L. Lehmann // The Annals of Mathematical Statistics. — 1963. — No. 4. — P. 598–611.

Поступила в редакцию 14.02.2023.

Поступила после рецензирования 07.03.2023.

Принята к публикации 09.03.2023.

*Об авторах:*

**Мисюра Валентина Владимировна**, кандидат физико-математических наук, доцент, кафедра математики и информатики, Донской государственный технический университет (РФ, 344003, Ростов-на-Дону, пл. Гагарина, 1), [ORCID](https://orcid.org/0000-0001-9151-2011), [vvmisyura2011@gmail.com](mailto:vvmisyura2011@gmail.com)

**Мисюра Елена Владимировна**, Российский экономический университет им. Плеханова, Институт цифровой экономики и информационных технологий (Российская Федерация, 117997, г. Москва, Стремянный переулок, 36), [ORCID](https://orcid.org/0000-0001-9151-2011), [misurahelen@gmail.com](mailto:misurahelen@gmail.com)

*Конфликт интересов:*

Авторы заявляют об отсутствии конфликта интересов.

*Все авторы прочитали и одобрили окончательный вариант рукописи.*