ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

DOI: https://doi.org/10.32839/2304-5809/2019-2-66-49
УДК 519.11

Крутик А.В.
Сквирская государственная районная администрация

МАТЕМАТИЧЕСКОЕ ПРОГНОЗИРОВАНИЕ НА КОРОТКИЕ ПРОМЕЖУТКИ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ СКОЛЬЗЯЩЕГО МЕТОДА СГЛАЖИВАНИЯ ПЕРЕМЕННОЙ РАЗНОСТИ «АСИММЕТРИЧНОЙ ЧЕТЫРЕХТОЧЕЧНОЙ ФУНКЦИЕЙ»

Аннотация. В этой статье автор предлагает включить в нейронную сеть при прогнозировании следующего значения временного ряда методы математического прогнозирования путем скользящего сглаживания переменной разностью «асимметричной четырех – точечной функцией». Кроме того, в статье был предложен оригинальный и довольно простой метод коррекции параметров математического прогнозирования путем приведения параметров результата прогнозирования в соответствие с параметрами исходного временного ряда. На примере исследования выборки из числового ряда индексов Доу-Джонса автор показывает подробную реализацию предложенных методов прогнозирования и метод коррекции их характеристик. Конечно, основная идея представленной работы состоит в том, чтобы оптимизировать использование Земных ресурсов посредством их математического прогнозирования, что в конечном итоге, безусловно, приведет к улучшению качества нашей жизни.

Ключевые слова: математическое прогнозирование, переменные разности, нейронные сети, искусственный интеллект, индекс Доу-Джонса.

Krutik Aleksandr
Skvyra State District Administration

MATHEMATICAL PREDICTION FOR SHORT PERIODS OF TIME USING THE SLIDING METHOD OF SMOOTHING THE VARIABLE DIFFERENCE "ASYMMETRIC FOUR-POINT FUNCTION"

Summary. In this article, the author proposes to include in the neural network, when predicting the next value of the time series, mathematical prediction methods by sliding smoothing with a variable difference "asymmetric four-point function". In addition, the article proposed an original and fairly simple method for setting mathematical prediction parameters by adjusting the prediction result parameters in accordance with the parameters of the original time series. Using the example of studying a sample from a numerical series of Dow Jones indices, the author shows a detailed implementation of the proposed prediction methods and a method for adjusting their characteristics. Of course, the main idea of the presented work is to optimize the use of earth resources through their mathematical prediction, which ultimately will certainly lead to an improvement in the quality of our life. The author believes that many of the quantities characterizing certain phenomena (events) can not constantly increase or decrease. As practice shows, in nature, oscillatory changes in quantities are more characteristic; they are described by a sum of harmonics with modulated values of amplitudes and phases and with the presence of a random component that cannot be described by any known mathematical functions. In other words, in some time sections (mostly linear), the least squares method works well, in others – sliding smoothing (suppression of high-frequency bursts) works, and in the third, the simplest "closest neighbour (primitive method)" is best. Therefore, each new proposed effective method of mathematical prediction is a significant contribution to the study of mathematical prediction as a branch of mathematics in general. And, as shown by the analysis of the mathematical prediction methods proposed by the author, using the example of studying the predicted values of the Dow Jones index, the methods can be practically used along with other existing mathematical prediction methods in neural networks of artificial intelligence.

Keywords: mathematical prediction, variable differences, neural networks, artificial intelligence, Dow Jones index.

Постановка проблемы. Сегодня невозможно представить современную маркетинговую политику любой фирмы (тем более государства) без учета прогноза основных числовых экономических показателей от которых зависит увеличение их чистого дохода. Особое внимание, в последнее время, уделяется нейронным сетям искусственного интеллекта (и.и.), применяемым для математического прогнозирования. В основе этих сетей лежат известные методы математического прогнозирования. Тема исследований в этом направлении есть достаточно новой и не вычерпавшей всех возможных математических решений. Поэтому, считаю, что исследования в этой области есть особенно актуальными на данный час.

Анализ последних исследований. Еще в 1927 году В.А. Базаров-Руднев предложил 3 метода прогноза: экстраполяция, аналитическая модель, экспертиза ([1] Константинопольская, 2019). В настоящее время существует около 220 методов прогнозирования. Чаще всего на практике используются не более 10, среди них: фактографические (экстраполяция, интерполяция, тренд-анализ), экспертные (в т.ч. опрос, анкетирование), публикационные (в т.ч. патентные), цитатные – индексные, сценарные, матричные, моделирование, аналогий, построение графов и другие методы.

Появляются новые методы прогнозирования, одним из таких наиболее перспективных методов, на мой взгляд, есть метод нейронных сетей. Предложенные в работе методы прогнозирования переменной разностью «асимметричной четырехточкой» на короткие интервалы времени могут эффективно дополнить уже существующие методы использующиеся в таких сетях и.и.

Выделение нерешенных ранее частей общей проблемы. На мой взгляд, так как мы живем в ограниченном жизненном пространстве, многие величины, характеризующие те или иные явления (события) не могут постоянно увеличиваться или уменьшаться. Как показывает практика в природе больше присущи колебательные изменения величин описывающиеся суммой гармоник с модулированными значениями амплитуд и фаз и с наличием случайной составляющей, которую вообще невозможно описать никакими известными математическими функциями. Поэтому, считаю, что на сегодняшний момент времени не существует универсальных методов математического прогнозирования. Все методы зависят как от промежутка времени на который делается прогноз, от точности представления (модулирования) исследуемого промежутка числового ряда той или иной известной функцией так и от самого исследуемого ряда чисел, его двумерной корреляционной функцией (зависимость коэффициента линейной корреляции Пирсона от частоты дискретизации и от времени запаздывания). Иными словами на одних временных участках (в основном линейных) хорошо работает метод наименьших квадратов, на других – методы скользящего сглаживания (подавления высокочастотных всплесков) а на третьих лучшим является простейший метод «ближайшего соседа (примитивный метод)» $Xpr_i = X_{i-1}$. Поэтому, на мой взгляд, каждый новый предложенный эффективный метод математического прогнозирования есть существенным вкладом в изучение математического прогнозирования как раздела математики вообще.

Формирование целей статьи. Можно ли решить задачу математического прогнозирования с высокими качественными характеристиками на короткий промежуток времени новыми, еще не изученными математическими методами? На мой взгляд такие методы существуют и чем больше мы их будем исследовать и применять тем более качественнее будет конечный продукт исследования – математический прогноз значения события (явления).

Постановка задания. И так, задача представленной вашему вниманию работы – создать новые универсальные достаточно эффективные методы математического прогнозирования на короткий промежуток времени и метод коррекции их качественных характеристик.

Изложение основного материала. 1. Математическое прогнозирование на короткие временные интервалы методами скользящего сглаживания переменной разности «асимметричной четырех – точкой». Коррекция параметров математического прогнозирования методом приведения параметров результата прогнозирования среднего математического отклонения (СКО) σ, математического ожидания (МАТ.ОЖ) м, к соответствующим параметрам исходного временного ряда.

1.1. Анализ формул приблизительного вычисления следующего члена числового ряда исходя из его переменных разностей ([2] Т. Андерсон, 1976).

1.1.1. Метод скользящего сглаживания переменной разности «асимметричной четырех – точкой», в основе которого лежит формула, образованная из промежуточной формулы приблизительного вычисления исходного числового ряда исходя из его переменных разностей ([3] А. Крутик, 2018).

Для расчета коэффициентов скользящей несимметричной «четырех – точки» представим уравнение вида:

$$Xpr_i = a * X_{i-1} + b * X_{i-2} + c * X_{i-3} + d * X_{i-4} \qquad (1)$$

где Xpr_i – прогнозированное значения X на время $ti = i * dt$;

где i – индекс (порядковый номер) числа ряда, dt значение дискрета времени равной $dt = t_i - t_{i-1}$;

$X_{i-1}, X_{i-2}, X_{i-3}, X_{i-4}$ – значения числового ряда X на соответствующие моменты времени ti;

коэффициенты a, b, c, d – коэффициенты при $X(i)$, обеспечивающие максимальное значение коэффициента линейной корреляции Пирсона r между исследуемым числовым рядом и числовым рядом образованным прогнозными значениями исследуемого ряда.

В конечном итоге метод несимметричной четырех – точки сводится к вычислению указанных коэффициентов. И так приступим к рассмотрению предложенного метода. В своей работе в формуле (8) автором рассчитана промежуточная формула вычисления приблизительного восстановления исходной функции исходя из ее переменных разностей для X_5 – 5-го элемента числового ряда и 4-й степени переменной разности. Рассмотрим более подробно эту формулу.

$$X_5 = 0.32(R_{4,1} - R_{5,1} + R_{3,2} - R_{4,2} + R_{5,2} - R_{2,3} + R_{3,3} - R_{4,3} + R_{5,3} + R_{1,4} - R_{2,4} + R_{3,4} - R_{4,4} + R_{5,4}) -$$
$$-0.03X_1 + 0.194X_2 - 0.52X_3 + 0.839X_4 + 0.032X_5 + 0.839X_6 - 0.52X_7 + 0.194X_8 - 0.03X_9$$

Необходимо заметить, что представленная формула не имеет физического смысла как такового вообще потому, что ее решение сводится к решению типа $X_5 = X_5$. Но как инструмент для оптимального расчета указанных выше коэффициентов формула есть определяющей при рассмотрении этого метода прогнозирования. И так разделим условно формулу на 3 части:

1-я часть $- 0.839X_4 - 0.52X_3 + 0.194X_2 - 0.03X_1$

– значения характеризующие «прошлое» исследуемого числового ряда

2-я часть $-0.032X_5 + 0.839X_6 - 0.52X_7 + 0.194X_8 - 0.03X_9$

– значения характеризующие «будущее» исследуемого числового ряда;

3-я часть $- 0,32(R_{41} - R_{51} +_{32} - R_{42} + R_{52} - R_{23} + R_{33} -$
$- R_{43} + R_{53} + R_{14} - R_{24} + R_{34} - R_{44} + R_{54})$

ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

– значения, характеризующие связь «прошлого с будущим» исследуемого ряда.

2-я и 3-я части исследуемой функции нам недоступны, поэтому предлагается провести исследование возможности применения для последующих расчетов 1-й части уравнения, припустив, что значения 2-й и 3-й части, при центрированном тренде (МАТ.ОЖ = 0) стремятся к «0».

И так, простую формулу сглаживания разностной «четырех – точкой», представим через коэффициенты 1-й части рассмотренного уравнения:

$$Xnp_5 = 0.839X_4 - 0.52X_3 + 0.194X_2 - 0.03X_1$$

В общем виде уравнение будет иметь следующий вид:

$$X_i = 0.839X_{i-1} - 0.52X_{i-2} + 0.194X_{i-3} + 0.03X_{i-4} \quad (2)$$

И, как показывают практические исследования прогнозирование по данной формуле при представлении исследуемого ряда чисел основными применяемыми для описания исследуемого явления функциями дает неплохой результат, который характеризуется хорошими качественными характеристиками. Назовем его **«несимметричной разностной четырех – точкой» № 1**.

1.1.2. Метод скользящего сглаживания переменной разности «несимметричной четырех – точкой» (№ 2).

В основе этого метода лежит формула вычисления очередного члена числового ряда исходя из его переменных разностей, положенная в основу алгоритма вычисления первой механической программированной вычислительной аналитической машины Чарльза Беббиджа.

Еще в позапрошлом веке при вычислении очередного значения ряда заданного полиномом н-го порядка широко применялся метод переменных разностей, вернее одно из его замечательных свойств:

Рассмотрим на простом примере это свойство. Припустим $Y_i = i^3$. Распишем следующий ряд переменными разностями:

i :	0	1	2	3	4	5
Y_i :	0	1	8	27	**64**	**125**
R_1 :		1	7	19	**37**	
R_2 :			6	12	**18**	
R_3 :				6	**6**	
R_4 :					**0**	

$Y_5 = 64 + 37 + 18 + 6 + 0 = 125$ (см. выделенные серым цветом числа) (нетрудно заметить, что каждая очередная переменная разность для числовых рядов также как и производная для функций понижает степень полинома на 1)

Выразим Y_5 через Y_i, получим: $Y_5 = 4Y_4 - 6Y_3 + 4Y_2 - Y_1$

В общем виде уравнение будет иметь следующий вид:

$$Y_i = 4Y_{i-1} - 6Y_{i-2} + 4Y_{i-3} - Y_{i-4} \quad (3)$$

Таким образом, полученное выражение в дальнейшем будем использовать для прогноза следующего члена числового ряда, назовем его также **«несимметричной разностной четырех – точкой» № 2**.

1.1.3. Исследование качества прогнозирования очередного члена числового ряда (для оценки качества выбрав коэффициент линейной корреляции r) различными методами прогнозирования.

Для исследования возьмем следующие методы:

№ 1 «ближайшего соседа»: $Y_i = Y_{i-1}$;

№ 2 простым методом скользящего сглаживания: $Y_i = (Y_{i-1} + Y_{i-2} + Y_{i-3} + Y_{i-4}) / 4$;

№ 3 Методом скользящего сглаживания переменной разности «несимметричной четырех-точкой» № 2 (разд. 1.1.2): $Y_i = 4Y_{i-1} - 6Y_{i-2} + 4Y_{i-3} - Y_{i-4}$

№ 4 методом скользящего сглаживания «несимметричной четырех – точкой» № 1 (разд. 1.1.1): $X_i = 0.839X_{i-1} - 0.52X_{i-2} + 0.194X_{i-3} + 0.03X_{i-4}$

Для представления исследуемого числового ряда выберем следующие основные модели изменения значений ряда представленные функциями:

1. $Y_i = e^i$; **2.** $Y_i = 1 / (i^2 + 0.001)$; **3.** $Y_i = i - i^2 + i^3 - i^4$; **4.** $Y_i = \log(i + 6)$; **5.** $Y_i = \cos((6.28 / 12) * i)$; **6.** $Y_i = 2^i$, где i – индекс (номер числа в ряде)

Необходимо заметить что проведенные исследования были без учета случайной составляющей ошибки. Тем не менее, все исследуемые методы математического прогнозирования на короткий промежуток времени показали очень хорошие результаты. Результаты «победителей» в этом, так сказать соревновании, отмечены в таблице красным цветом. У многих может возникнуть закономерный вопрос – почему в качестве единого показателя качества прогнозирования выбран коэффициент линейной корреляции Пирсона? Ответ на этот вопрос постараюсь ар-

Таблица 1

Значения r – коэффициента линейной корреляции Пирсона между числовым рядом заданным одной из исследуемых функций и числовым рядом из прогнозируемых значений различными методами

Исследуемая функция	Коэф. корел. r между ислед. функцией и методом мат. прогн. № 1	Коэф. корел. r между ислед. функцией и методом мат. прогн. № 2	Коэф. корел. r между ислед. функцией и методом мат. прогн. № 3	Коэф. корел. r между ислед. функцией и методом мат. прогн. № 4
$Y_i = e^i$	1	1	1	1
$Y_i = 1 / (i^2 + 0.001)$	0,998	0,97	-0,517	0,9998
$Y_i = i - i^2 + i^3 - i^4$	0,999	0,998	1	0,9998
$Y_i = \log(i + 6)$	0,9999	0,9998	1	1
$Y_i = \cos((6.28 / 12) * i)$	0,74	-0,36	0,89	0,99
$Y_i = 2^i$	1	1	1	1

гументировано дать в следующих разделах данной работы.

1.2. Коррекция основных параметров математического прогнозирования методом приведения параметров результата прогнозирования (СКО, МАТ.ОЖ) к соответствующим параметрам исходного временного ряда.

Как уже отмечалось выше, метод достаточно прост и эффективен. И так приступим к рассмотрению указанного метода.

Имеются два числовых ряда;

1. $X(i)$ – исследуемый числовой ряд, характеризующийся значениями среднеквадратического отклонения СКО(X) и математическим ожиданием МАТ.ОЖ(X).

2. $Y(i)$ – числовой ряд образованный из прогнозных значений, полученных в результате математического прогнозирования одним из известных методов, который также характеризуется значениями СКО(Y) (σ) и МАТ.ОЖ(Y) (M), а также значением коэффициента линейной корреляции Пирсона r_{xy} между исследуемым и прогнозированным рядами.

– Далее пронормируем числовой ряд Y, в результате получим:

$$Ynor_i = (Y_i - \mathrm{M}(Y)) / \sigma(Y) \quad (4)$$

(Как известно, СКО (нормированного числового ряда равно 1, МАТ.ОЖ этого ряда равно 0)

– далее СКО этого нормированного ряда приведем к значению СКО исследуемого ряда путем умножения членов этого ряда на СКО(X). МАТ.ОЖ этого ряда приведем к МАТ.ОЖ исследуемого ряда путем смещения его значений (сумы) на величину МАТ.ОЖ(X). Запишем сказанное в виде формулы:

$$Ypr_i = ((Y_i - \mathrm{M}(Y)) / \sigma(Y)) * \sigma(X) + \mathrm{M}(X) \quad (5)$$

Проделав указанные преобразования с прогнозными значениями числового ряда $Y(i)$ мы получим СКО и МАТ.ОЖ равные соответствующим СКО и МАТ.ОЖ исследуемого ряда и как показывают исследования ошибка (разность значений между исследуемыми и прогнозированными значениями) в этом случае оптимизируется и зависит только от коэффициента линейной корреляции. При r = 1 ошибка равна 0.

1.3. Пример исследования выборки числового ряда индексов Доу-Джонса некоторыми методами математического прогнозирования. Коррекция параметров математического прогнозирования методом приведения параметров результата прогнозирования (СКО, МАТ.ОЖ) к соответствующим параметрам исходного временного ряда.

Для проведения исследования возьмем временной промежуток наблюдения за изменением индекса Доу-Джонса ([4] Минфин. Индекс Доу-Джонса, 2019) за период времени близкому к реальному и такому в котором произошел бы скачек индекса (его увеличение или понижение). Исходя из этих условий, выберем промежуток времени от 06.11.2018 г. и примерно до 26.01.2019 г. Как раз этот промежуток времени характеризуется незначительным временным понижением индекса 25 – 26.01.2018 г. от 25000 до 22000. Исследуемый период насчитывает 81 сутки, представив его с дискретностью равной 3 суткам (исходя из известного критерия – не менее 3-х измерений на период максимальной частоты оптимального спектра) получим следующий числовой ряд исследуемого индекса (см. рис. 1). Для упрощения расчетов (исключения дополнительных «0») значение индекса представим в сотнях (для получения реального значения необходимо указанные значения умножить на 100).

С периодом около 105 суток (35 отметок на графике) в этом промежутке происходили значительные понижения индекса и с периодом около 30 и 12 суток (соответственно 10 и 4 отметки на графике) происходили менее значительные его колебания. Таким образом, исследуемый числовой ряд попытаемся представить следующей математической моделью, что описывается суммой 3-х гармоник (значения углов гармоник представлено радианами):

$$Y_i = 30\cos((6.28/35)*i - 0.1) + 12\sin((6.28/10)* \\ *i + 0.5) - 0.5\sin((6.28/4)*i - 0.5) \quad (6)$$

Амплитуды, частоты и фазы гармоник формулы 6 подобраны исходя из критерия максимума коэффициента корреляции Пирсона r между исследуемым рядом и рядом заданным моделирующей функцией.

Далее проведем исследования значений коэффициента корреляции r между исследуемым рядом и рядом составленным из прогнозных значений рассчитанными по формулах представленных в предыдущих разделах этой работы и его влияние на основные показатели качества прогноза, такие как СКО ошибки и МАТОЖ ошибки. Также проведем исследования этих значений при применении рассмотренного метода коррекции.

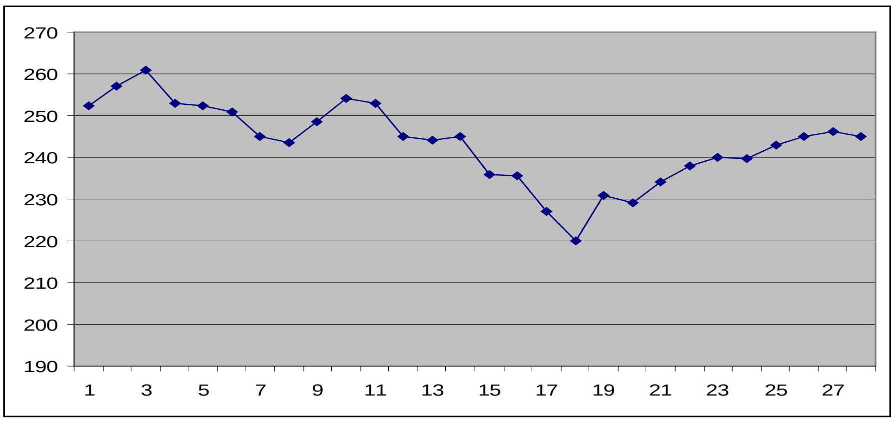

Рис. 1. Числовой ряд исследуемого индекса Доу-Джонса

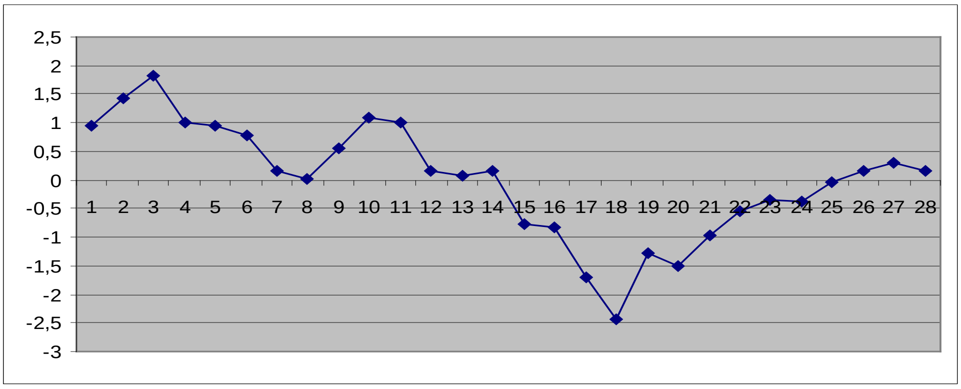

Рис. 2. Нормированные значения исследуемого ряда индексов Доу-Джонса

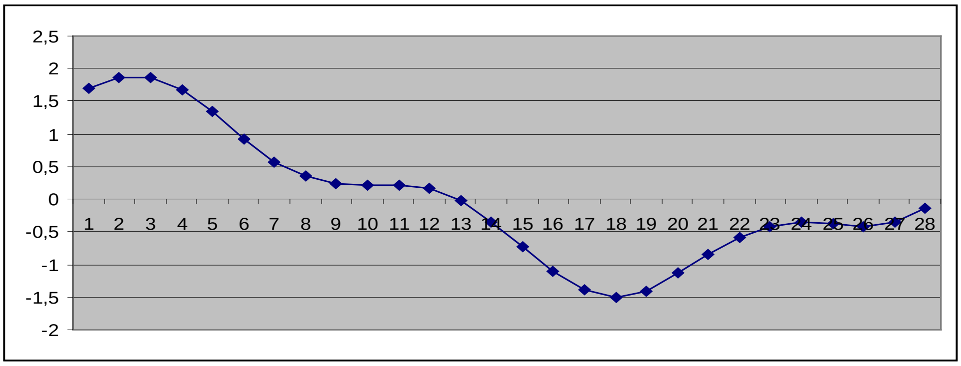

Рис. 3. Значения модулированного нормированного ряда функцией Y(i) (6)

Таблица 2

Результаты исследования очередного члена ряда индексов Доу-Джонса различными методами прогнозирования.

Метод прогнозирования	Коэффициент корреляции, r	Без коррекции СКО и МАТОЖ		С коррекцией СКО и МАТОЖ	
		СКО ошибки 1	МАТОЖ ошибки 1	СКО ошибки 2	МАТОЖ ошибки 2
$Y_i = Y_{i-1}$;	0.84	4,9	-0,33	5,22	-2,6
$Y_i = (Y_{i-1} + Y_{i-2} + Y_{i-3} + Y_{i-4}) / 4$;	0.67	7,1	-1,1	7,6	-2,7
$Y_i = 4Y_{i-1} - 6Y_{i-2} + 4Y_{i-3} - Y_{i-4}$	0.36	22,3	0,38	10,2	-1,7
$X_i = 0.839X_{i-1} - 0.52X_{i-2} + 0.194X_{i-3} + 0.03X_{i-4}$	0.79	5,6	124	5,91	-2,3
$Y_i = 30\cos((6.28/35)*i - 0.1)$ $+12\sin((6.28/10)*i + 0.5)$ $-0.5\sin((6.28/4)*i - 0.5)$	0.78	14,8	250	6,05	-2,9

Итак, проведем анализ полученных результатов исследования (табл. 2). Из проведенного исследования видно, что при увеличении коэффициента линейной корреляции r СКО ошибки уменьшается. МАТОЖ ошибки тоже значительно уменьшается, но требует осуществления коррекции предложенным методом. Прогнозирования кроме методов «скользящего сглаживания» и «ближайшего соседа» так они изначально имеют приблизительно одинаковые значения СКО и МАТОЖ с исследуемым рядом, требуют коррекции СКО и МАТОЖ предложенным методом.

Выводы. Исследуемые методы математического прогнозирования на «короткие дистанции» значений индекса Доу-Джонса за предложенный временной период дали средние результаты качества прогнозирования. Поэтому, в данном случае, практичнее использовать метод «ближайшего соседа» как более стабильный не требующий подстроек метод.

Не исключено, что при изменении индекса с течением времени рассмотренные методы могут улучшить свои качественные показатели (в первую очередь значения коэффициента, r) и стать основными методами для его прогноза. Поэтому при прогнозе индекса целесообразно вести постоянный анализ качества прогнозирования по всем известным методам и выбирать лучший на конкретном участке времени по значению коэффициента линейной корреляции Пирсона r, создав, таким образом, простейшую нейронную сеть искусственного интеллекта.

В предложенной работе детально изложены предложенные автором два методы скользящего сглаживания переменной разностью «асимметричными четырех точечными функциями» математического прогнозирования на короткие временные интервалы. Также в работе предложено проводить коррекцию параметров математического прогнозирования методом приведения СКО и МАТОЖ членов ряда прогнозных значений к соответствующим параметрам исходного временного ряда. И как показал проведенный в работе их анализ, на примере исследования прогнозируемых значений индекса Доу-Джонса, методы могут быть практически использованы наряду с другими уже существующими методами математического прогнозирования в нейронных сетях искусственного интеллекта. Конечной целью рассмотренных методов, безусловно, есть стремление оптимизации расходования существующих ресурсов планеты.

Список литературы:

1. Константиновская Л.В. Методы и приемы прогнозирования. URL: https://www.astronom2000.info/прогнозирование/mip/scin (дата обращения: 22.02.2019).
2. Андерсон Т. Статистический анализ временных рядов. Перевод с английского И.Г. Журбенко и В.П. Носко. Москва, 1976. С. 76–94.
3. Крутик А. Статистическое распознавание числовой последовательности методом анализа знаковых признаков переменных разностей в поле анализа системами с искусственным интеллектом. *Молодий вчений*. 2018. № 9(61). С. 195–201.
4. Индекс Доу-Джонса (Dow Jones). Индексы. Минфин. URL: https://index.minfin.com.ua/markets/stock/dji/scin (дата обращения: 11.02.2019).

References:

1. Konstantinovskaya L.V. Metody i priemy prognozirovaniya. URL: https://www.astronom2000.info/прогнозирование/mip/scin (22.02.2019).
2. Anderson T. (1976). Statisticheskiy analiz vremennykh ryadov. Perevod s angliyskogo I.G. Zhurbenko i V.P. Nosko. Moskva. S. 76–94.
3. Krutik A. (2018). Statisticheskoe raspoznavanie chislovoy posledovatel'nosti metodom analiza znakovykh priznakov peremennykh raznostey v pole analiza sistemami s iskusstvennym intellektom. *Molodiy vcheniy*. № 9(61). S. 195–201.
4. Indeks Dou-Dzhonsa (Dow Jones). Indeksy. Minfin. URL: https://index.minfin.com.ua/markets/stock/dji/scin (11.02.2019).