3. Музаев М.З., Денисова И.П., Самойлова К.Н., Ширшов В.Ю. Тренды развития страховых продуктов в условиях цифровизации экономики // Финансовые исследования. 2019. №3 (64).
4. Пашкова Е.Н., Генинг Д.С. Развитие интернет-страхования в условиях применения современных технологий // Вектор экономики. – 2020. – № 5 (47). – С. 69.
5. Пашкова Е.Н. Современные тенденции цифровизации ОСАГО // Вектор экономики. - 2019. – № 5 (35). – С. 122.
6. Сплетухов Ю. А. Направления расширения использования цифровых технологий в страховании //Корпоративная экономика. – 2020. – №. 1. – С. 31-37.
7. Щербакова Н. В., Ильиных Ю. М. Страхование в эпоху цифровых и интернет-технологий //Экономика Профессия Бизнес. – 2019. – №. 1.
8. Любарская О., Шкреба А., Янин А. Электронное страхование: вынужденное ускорение. [Электронный ресурс] – Режим доступа: https://raexpert.ru/docbank/b59/59c/078/3b865e5a2ec25eca9640554.pdf (дата обращения 15.03.2021)
9. Обзор ключевых показателей деятельности страховщиков за 2020 год, № 4 [Электронный ресурс] – Режим доступа: https://cbr.ru/Collection/Collection/File/32073/review_insure_20Q4.pdf (дата обращения 15.03.2021)
10. Развитие финансовых технологий [Электронный ресурс] – Режим доступа: https://cbr.ru/fintech/ (дата обращения 15.03.2021)

ПРОГНОЗИРОВАНИЕ РЕЙТИНГОВ УНИВЕРСИТЕТОВ С ПОМОЩЬЮ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

В.М. Московкин, Чжан Хэ
г. Белгород, Россия
Белгородский государственный национальный
исследовательский университет

В связи с проектами “5 – 100” и “Академическое превосходство” по введению ведущих российских университетов на определенные позиции мировых университетских рейтингов, встала задача прогнозирования этого процесса. В качестве метода прогнозирования был выбран анализ временных рядов, а в качестве глобального университетского рейтинга – шанхайский рейтинг ARWU, который имеет самый длинный временной ряд данных (18 лет). В качестве прогнозируемых временных рядов были выбраны интегральные показатели (Total Score) рейтингов Melbourne University (Австралия) и Vanderbilt University (США), которые имели четкие, соответственно, положительный и отрицательный тренд. Статья даёт чёткое представление, как последовательно нужно проводить анализ временного ряда с установлением его качества, а потом и прогнозировать на небольшой отрезок времени.

Ключевые слова: *прогнозирование, рейтинги университетов, временные ряды, ARWU, Total Score.*

PREDICTION UNIVERSITY RANKINGS USING TIME SERIES ANALYSIS

V.M. Moskovkin, Zhang He
Belgorod, Russia
Belgorod State National Research University

In connection with the projects “5 – 100” and “Academic Excellence” on the introduction of leading Russian universities in certain positions of the World University Rankings, the task of forecasting this process arose. Time series analysis was chosen as the forecasting method, and the Shanghai ARWU ranking, which has the longest time series of data

(18 years), was chosen as the Wlobal University Ranking. The integral indicators (Total Score) of the ratings of Melbourne University (Australia) and Vanderbilt University (USA) were chosen as the predicted time series, which had a clear, respectively, positive and negative trend. The article gives a clear idea of how to consistently analyze the time series with the establishment of its quality, and then predict for a short period of time.
Keywords: *prognosis, university rankings, time series, ARWU, Total Score.*

В связи с проектом “5 – 100” по введению к концу 2020 года, по крайней мере, пяти российских университетов в TOP-100 одного или нескольких ведущих глобальных университетских рейтингов (ARWU, THE, QS), встала задача прогнозирования вхождения ведущих российских университетов на заданные позиции в этих рейтингах. По завершению этого проекта началась подготовка к проекту “Академическое превосходство” по введению к 2023 году, по крайней мере, десяти ведущих российских университетов в TOP-500 вышеуказанных рейтингов. В этой связи приобрела актуальность разработка соответствующих методов прогнозирования или использования уже соответствующих.

Для прогнозирования динамики интегральных показателей (Total Score, Overall Score) и рангов глобальных университетских рейтингов целесообразно использовать, на наш взгляд, аппарат анализа временных рядов [1-3]. Наиболее длинные временные ряды этих показателей имеют место для рейтинга ARWU, который запущен в 2003 г. Британские университетские рейтинги для такого анализа пока не подходят, так как их ряды на один год меньше, и кроме того их методология расчёта несколько раз изменялась.

Рассматривая на сайте ARWU первые сто университетов, можно увидеть, что большинство из них имеют все 18 временных точек с рангами, Total Score и значениями индикаторов. Если рассматривать временные ряды относительно рангов, то можно увидеть, что первые 14 университетов имеют стабильные ранги, варьирующие в пределах четырёх единиц.

Рассматривая графики остальных временных рядов рангов университетов в TOP – 100 ARWU, можно увидеть временные ряды с большими и небольшими положительными и отрицательными трендами, осложненными нерегулярными колебаниями. Циклических колебаний с одинаковыми периодами практически не наблюдается. Всё это можно отнести к типизации временных рядов рассматриваемого рейтинга. Для целей прогнозирования мы выбрали несколько университетов с достаточно большими положительными и отрицательными трендами: University of Melbourne (Австралия), Vanderbilt University (США). Ниже проделаем анализ и прогнозирование временного ряда Total Score для University of Melbourne (рис. 1).

Когда мы имеем временной ряд, то, в первую очередь, мы должны проверить гипотезу о случайности ряда. Одним из возможных вариантов для этого служит проверка гипотезы H_0 о случайности (или её отсутствии) тренда временного ряда при конкурирующей гипотезе H_1 о не случайности (наличии случайности) тренда, основанная на сравнении средних значений первой и второй половины ряда по t – статистике (расчетному значению критерия, $t_{расч}$).

При расчете этого критерия, для нашего случая, $n_п$ =9 (количество членов первой половины ряда), $n_в$ =9 (количество членов второй половины ряда), получим следующие средние значения и дисперсии первой и второй половины: $\bar{y}_п$=27,3889, $\bar{y}_в$=33,6, $S^2_п$=1,736543, $S^2_в$=5,162222. На основе этих данных рассчитываем $t_{расч}$=-7,09421.

Значение | $t_{расч}$ | сравнивается с критическим значением распределения Стьюдента $t_{кр}(\alpha, df)$ с $df = n_п + n_в - 2$ степенями свободы и уровнем значимости α.

В случае, если | $t_{расч}$ | < $t_{кр}$ гипотеза H_0 о случайности тренда временного ряда принимается и ряд считается стационарным.

В противном случае, если | $t_{расч}$ | > $t_{кр}$ гипотеза H_0 отвергается, что свидетельствует о значимости различия средних первой и второй половины ряда и неслучайности (наличии) тренда. Другими словами, этом ряд имеет тенденцию изменения.

В нашем случае, когда df=16, α=0.05, | $t_{расч}$ |=7,09421 больше $t_{кр}$=2.1198, что свидетельствует о значимости различия средних первой и второй половины ряда. Наличие тренда неслучайно, то есть ряд не является стационарным.

Если при графическом изображении временного ряда тренд прослеживается недостаточно отчетливо, то ряд сглаживают, на график наносят сглаженные значения и, как правило, тенденция проявляется более четко. Кроме того, сглаживание позволяет устранить случайные колебания уровней. Метод простого скользящего среднего используется обычно в тех случаях, когда график временного ряда близок к прямой линии, поскольку при этом динамика исследуемого явления не искажается (p = 1). Определяется интервал сглаживания, то есть число входящих в него уровней m (m < n). В нашем случае, исходный и сглаженный ряд при m=5 и p = 1 (линейное сглаживание) показан на рис. 1.

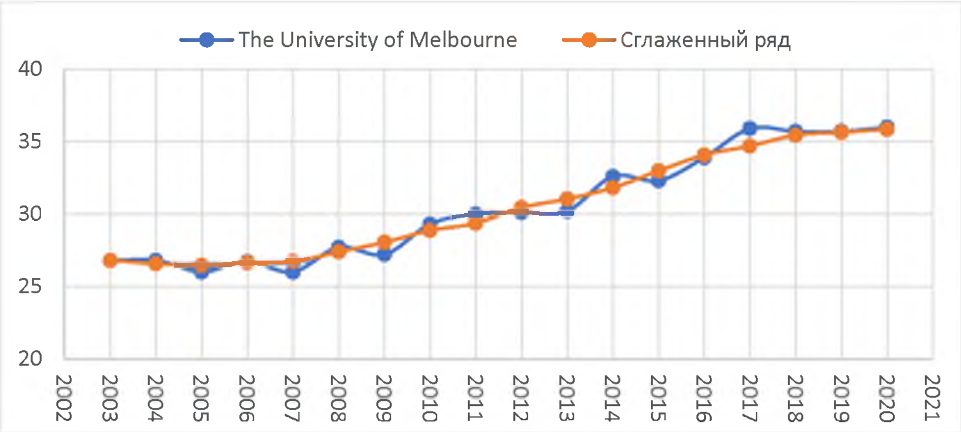

Рис. 1. Исходный и сглаженный ряд Total Score University of Melbourne

При наличии во временном ряду тенденции и циклических колебаний значение каждого последующего уровня ряда зависят от значений предыдущих уровней. В этом случае говорят, что ряд имеет автокорреляцию. Автокорреляционный анализ временного ряда позволяет установить степень зависимости последующих членов ряда от предыдущих и временной интервал, в течение которого эта зависимость статистически значима. После вычисления коэффициентов автокорреляции необходимо проверить их статистическую значимость сравнением с критическими значениями коэффициента корреляции $r^{l}_{кр}$(α, k). Критические значения берутся из таблицы критических значений корреляции по уровню значимости α и числу степеней свободы k = n – 1 – 2. Если | r_l | < $r^{l}_{кр}$,то коэффициент автокорреляции r_l статистически незначим и выводы, сделанные по его значению, имеют вероятность ошибки, равную 1 – α .

Последовательность коэффициентов автокорреляции называют автокорреляционной функцией временного ряда, а график зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется коррелограммой (корреограммой). Поскольку знаки коэффициентов автокорреляции при анализе не учитываются, коррелограмма строится по их абсолютным значениям.

Анализ коэффициентов автокорреляции производится на основании шкалы Чеддока.

Если абсолютное значение коэффициента автокорреляции первого порядка | r_1 | > 0,7 , ряд содержит линейную тенденцию, если | r_1 | < 0,5 – ряд содержит нелинейную тенденцию. В случае 0,5 < | r_1 | < 0,7 выбор вида уравнения тренда требует дополнительных исследований.

В случае, когда наибольшее абсолютное значение имеет коэффициент автокорреляции порядка $l = \tau > 1$ и при этом, $|r_\tau| > 0,7$, ряд содержит циклические колебания с периодом в τ моментов времени.

В нашем случае, графики функций автокорреляции и её критических значений приведены на рис. 2. Эти графики рассчитаны на основе встроенной функции “Автокорреляция” пакета SPSS.

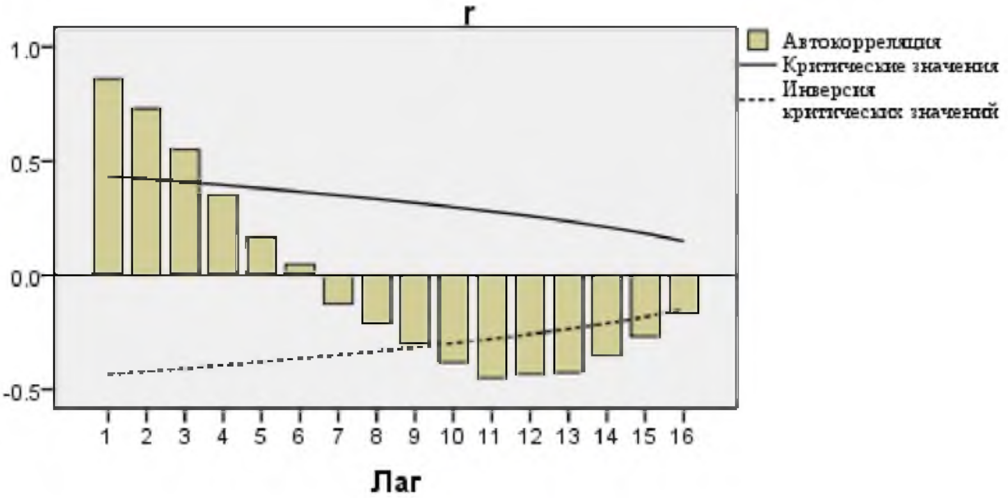

Рис. 2. Графики функций автокорреляции, и её критических значений для рангов University of Melbourne

Поскольку $r_1 = 0,86 \geq 0,70$ (рис. 2), изучаемый ряд имеет линейную тенденцию, уравнение тренда представляет собой прямую линию $\hat{y} = b_0 + b_1 \cdot t$. Из этого же рисунка видим, что $r_2 = 0,729 \geq 0,70$, поэтому изучаемый ряд содержит циклические колебания с периодом в 2 года. Уравнение линейного тренда находим с помощью метода наименьших квадратов $y = 0,6711x - 1319,4$.

Проделаем оценку качества уравнения линейного тренда. Если $\bar{y_t} \approx \tilde{y_t}$, то расчеты выполнены верно. В нашем случае $\bar{y_t}$=30,49444 $\tilde{y_t}$=30,51765, то есть $\bar{y_t} \approx \tilde{y_t}$

Для оценки математической точности уравнения тренда можно воспользоваться средней относительной ошибкой аппроксимации

$$\bar{A} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \tilde{y_t}}{y_t}\right| \cdot 100\%.$$

Из этой формулы получим $\bar{A}$=2,6221%. Принятие решения о точности уравнения основывается на специальной таблице и если в ней $\bar{A}$ меньше 10, то точность уравнения считается высокой, что и имеет место в нашем случае. Проверку статистической значимости уравнения тренда произведем с помощью F–критерия Фишера, при этом его табличное (критическое) значение $F_{табл}$ находится по таблице критических значений распределения Фишера- Снедекора (F-распределения) по уровню значимости α и двум числам степеней свободы df1 = m =1 (линейный тренд) и df2 = n – m – 1 = 18 – 2 =16, или с помощью встроенной функции Excel «FРАСПОБР».

Если $F_{расч} > F_{табл}$, то с вероятностью ошибки α уравнение линейного тренда в целом статистически значимо (адекватно). В противном случае ($F_{расч} < F_{табл}$) уравнение в целом статистически незначимо. В нашем $F_{расч}$=199.7352 > $F_{табл}$ =4.49, то есть приходим к статистически значимому тренду при α =0,05.

Для проверки статистической значимости оценок параметров b_0, b_1 воспользуемся t-критерием Стъюдента. Теоретическое значение критерия $t_{табл}$ находится по таблице критических значений распределения Стъюдента по уровню значимости α и числу степеней свободы df = n – m – 1 = 18 – 2 = 16, или с помощью встроенной функции Excel «СТЬЮДРАСПОБР».

Если | t_{bj} | > $t_{табл}$, то с вероятностью ошибки α оценка параметра уравнения тренда b_j (j = 0, 1) статистически значима. В противном случае (| t_{bj} | < $t_{табл}$) b_j статистически незначима. В нашем случае t_{b0}=-13,81347, t_{b1}=14,13277, $t_{табл}$=2,1199, то есть | t_{b0} | > $t_{табл}$, | t_{b1} | > $t_{табл.}$ Следовательно, мы приходим к статистически значимым параметров линейного уравнения тренда при α =0,05.

Теперь необходимо проделать анализ отклонений от тренда. Согласно аддитивной модели временного ряда

$$y_t = \widehat{y_t} + \varepsilon_t$$

остатки (отклонения от тренда) ε_t находятся как разность между фактическими (наблюдаемыми) и теоретическими значениями уровней ряда:

$$\varepsilon_t = y_t - \widehat{y_t}$$

В линейной модели остатки должны удовлетворять требованиям теоремы Гаусса-Маркова. Для этого следует проверить требование M (ε_t) = 0. Числовой оценкой математического ожидания M(ε_t) является среднее значение M(ε_t) = $\overline{\varepsilon_t}$. Для выполнения требования M (ε_t) = 0 необходимо, чтобы

$$\overline{\varepsilon_t} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \approx 0,$$

а также, чтобы дисперсия остатков ε_t была одинаковой для всех значений t (свойство гомоскедастичности). Если это условие не соблюдается, то имеет место гетероскедастичность остатков. В нашем случае первое требование имеет место.

Чтобы проверить второе требование, нужно показать, что для разных наблюдений остатки ε_t были некоррелированы (независимы). Для этого используется критерий Дарбина-Уотсона (о наличии в остатках автокорреляции первого порядка), в котором рассчитывается статистика $d_{расч}$, удовлетворяющая условию 0 ≤ $d_{расч}$ ≤ 4.

Для линейного уравнения тренда теоретические значения критерия Дарбина-Уотсона d_U и d_L находятся по таблице критических значений по объему выборки n, числу степеней свободы df = 1 и уровню значимости α. С помощью критических значений числовой промежуток (0; 4) разбивается на пять отрезков и в зависимости от попадания в тот или иной отрезок принимается решение о независимости остатков. В нашем случае: $d_{расч}$=1.002204, d_L=1.16, d_u=1.39, в результате чего мы попадаем в зону с положительной автокорреляцией остатков. То есть второе условие для теоремы Гаусса – Маркова о независимости остатков не выполняется.

Если уравнение линейного тренда признано качественным, а остатки удовлетворяют требованиям теоремы Гаусса-Маркова, то аддитивная модель временного ряда считается качественной, то есть она адекватно описывает исходный ряд. К сожалению, остатки нашего ряда не удовлетворяют второму требованию теоремы Гаусса-Маркова. Сведём все полученные результаты в таблицу.

Таблица

Параметры аддитивной линейной модели временного ряда для рангов рейтинга ARWU University of Melbourne

Аддитивная модель y_t= $\widehat{y_t} + \varepsilon_t$	
Тренд	Отклонения от тренда
Линейный $\widehat{y_t} = b_0 + b_1 \cdot t$	$\overline{\varepsilon_t} = y_t - \widehat{y_t}$
$\bar{A}$=2,6221%	$M(\varepsilon_t) = \overline{\varepsilon_t}$ ≈0
$F_{расч}$=199,7352 ; $F_{табл}$ =4,49 α=0,05	$d_{расч}$=1,002204
$t_{расч}$=-7,09421	d_L=1,16; d_u=1,39
t_{b0}=-13,81347; t_{b1}=14,13277	Автокорреляция
$t_{табл}$=2,1199 b_0= 1319,4; b_1= 0,6711 R^2=0,9258; r=0,962203	r_1 = 0,86 ≥ 0,70 r_2=0,729>0,70 Согласно второго неравенства, изучаемый ряд содержит циклические колебания с периодом в 2 года.

Таким образом, построенная модель временного ряда $y = 0{,}6711x - 1319{,}4 + \varepsilon_t$ имеет высокую точность ($\bar{A}$ = 2,6221%), на уровне 5%-й ошибки статистически значима в целом, параметры тренда также статистически значимы. Отклонения от тренда имеют нулевое среднее, но зависимы. В целом модель обладает хорошим качеством и может быть использована для прогнозирования. Однако она имеет один недостаток – непостоянство дисперсии остатков, который объясняется тем, что в модели не учитываются циклические колебания, имеющие наибольшую величину отклонения от тренда.

Используем теперь построенную модель для краткосрочного прогнозирования временного ряда. Используем два метода: метод прогнозирования по среднему приросту и метод прогнозирования по уравнению тренда.

Для прогнозирования по первому методу необходимо определить средний прирост и делать прогноз по формуле

$$\widehat{y_{n+T}} = y_n + \bar{\Delta} \cdot T$$

где $\widehat{y_{n+T}}$ – прогнозируемый уровень ряда; T – срок прогноза (период упреждения); y_n – последний уровень ряда, за который рассчитан средний прирост $\bar{\Delta}$:

$$\bar{\Delta} = \frac{1}{n-1}\sum_{t=1}^{n-1}(y_{t+1} - y_t)$$

В нашем случае получено следующее прогнозное уравнение

$\widehat{y_{n+T}} = y_n + 0{,}541176T$ с точностью $\bar{A}$=4,1663%. По этому уравнению прогнозные значения Total Score на последовательные три года равны: y_{2021} = 36,54118 (соответствует, приблизительно, 33 – 34 местам или рангам); y_{2022} =37,08235 (соответствует, приблизительно, 32 – 33 местам); y_{2023} =37,62353 (соответствует, приблизительно, 30 – 31 местам). Прогнозное уравнение на фоне исходного временного ряда показано на рис. 3.

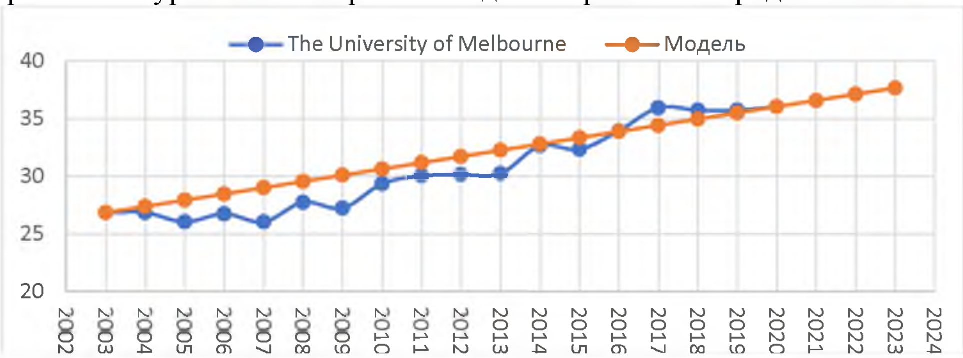

Рис. 3. Краткосрочное прогнозирование Total Score University of Melbourne по методу среднего прироста

Для краткосрочного прогноза по линейному уравнению тренда необходимо, очевидно, посчитать значения этого уравнения для нескольких прогнозных лет (рис. 3). Но в дополнение к точечному прогнозу необходимо определить границы возможного изменения прогнозируемого показателя. Их называют доверительными интервалами или интервальными оценками прогноза. Ширина доверительных интервалов зависит от степени колеблемости исследуемого процесса, от периода упреждения, от количества наблюдений в исходном временном ряду и других факторов.

В случае линейного тренда доверительный интервал имеет вид

$$(\widehat{y_{n+T}} - t_{1-a} \cdot S_{\hat{y}};\ \widehat{y_{n+T}} + t_{1-a} \cdot S_{\hat{y}})$$

и с вероятностью $\boldsymbol{\gamma} = 1 - \boldsymbol{\alpha}$ покрывает предсказываемое значение y_{n+T} . Здесь $t_{1-\alpha}$ – критическое значение распределения Стьюдента, найденное по уровню значимости α и числу степеней свободы df = n – m – 1, n – длина ряда, m – число параметров уравнения тренда, T – период упреждения.

Остаточная средняя квадратическая ошибка прогноза находится по формуле

$$S_{\hat{y}} = \sqrt{\frac{\sum_{t=1}^{n}(\widehat{y_{n+T}} - y_t)^2}{n - m - 1}}$$

Для прогнозных 2021, 2022, 2023 годов $S_{\hat{y}}$ были получены, соответственно, в виде: 7,796873; 8,423759: 9,063197, при этом критерий Стьюдента равен $t_{1-\alpha}$=0,064, α=0,05. Отсюда доверительные прогнозные интервалы для Total Score будут иметь, соответственно, вид: (36,3941, 37,3921); (37,02508, 38,10332); (37,65526, 38,81534). Первый интервал соответствует, приблизительно, 32 – 34 местам или рангам, второй – 29 – 33 местам, третий – 29 – 30 местам. Доверительные интервалы для Total Score показаны на рис. 4. Как можно видеть предыдущий метод прогноза дал близкие результаты.

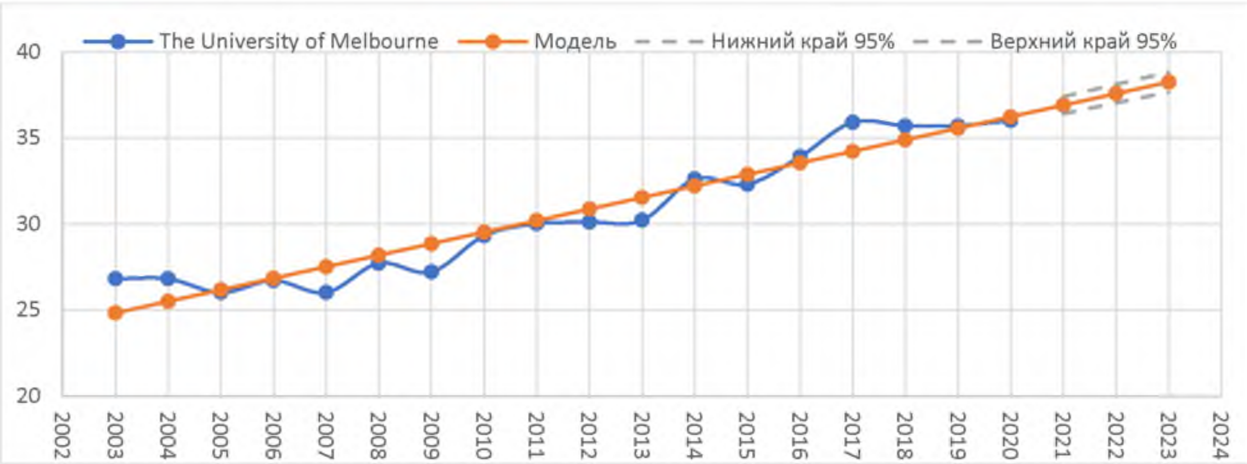

Рис. 4. Краткосрочное прогнозирование Total Score University of Melbourne по методу уравнения тренда

Аналогичный анализ временного ряда проделан для Vanderbilt University, который имел отрицательный тренд, и он в рейтинге ARWU перешёл с 32 места в 2013 году на 62 в 2020 году. Это говорит о том, что менеджеры этого университета должны принять срочные меры по повышению конкурентоспособности своего университета.

ЛИТЕРАТУРА

1. Карпенко Н.В. Эконометрика. Анализ и прогнозирование временного ряда: Учебное пособие. – М.: РУТ (МИИТ), 2018. – 132 с.

2. Бриллинджер Д. Временные ряды. Обработка данных и теория. М.: Мир, 1980 . – 536 с.

3. Носко В. П. Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов М. : Ин-т экономики переход. Периода, 2004. – 501 с.

КОНЦЕПЦИИ МЕТОДИЧЕСКОГО РАЗВИТИЯ МАЛОГО ПРЕДПРИНИМАТЕЛЬСТВА: ИННОВАЦИИ-ИНВЕСТИРОВАНИЕ

Д. И. Рўзиева
г. Ташкент, Узбекистан
Ташкентский государственный экономический университет

Развитие производственной деятельности субъектов малого предпринимательства в определенной степени зависит от уровня имеющихся