

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 8

Москва 2021

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 002:001:311.3

В.М. Московкин

Квартильный индекс в наукометрии

Первое упоминание о квартильном индексе восходит к работе Р. Кемптона и Л. Тейлора по популяционной биологии. Впервые для наукометрии этот индекс был предложен в 2019 г. историком А.В. Гринёвым из Санкт-Петербурга. Ряд усовершенствований этого индекса для мониторинга и оценки качественной структуры публикаций авторов и любых субъектов публикационной активности показал его значительное преимущество перед хиршеподобными метриками. Даются рекомендации по внедрению квартильного индекса в национальные и институциональные системы оценки качества публикационной активности.

Ключевые слова: *квартильный индекс, индекс Хирша, наукометрия, публикационная активность, качество публикаций, Scopus, Web of Science*

DOI: 10.36535/0548-0027-2021-08-1

ВВЕДЕНИЕ

Квартильные индексы хорошо известны в различных областях знания, особенно при изучении разнообразия каких-либо объектов. Расширенный поиск в *Google Scholar* даёт 209 откликов на запрос “*quartile index*” и 2 отклика на запрос «квартильный индекс»

(05.05.2021). Наиболее чётко суть этого показателя описана в докладе Чарльза Хултена из Национального бюро экономических исследований (1996): «Процедура, принятая в этом докладе, состоит в попытке иметь дело с агрегированием и нелинейностью в задачах путём взятия каждого индикатора Всемирного

банка и сортировке его по квартилям. Верхней квартиле придаётся весовое значение 1, второй квартиле – 0,75, следующей – 0,5 и нижней – 0,25. Это осуществляет ранжирование квартилей для каждой из четырёх систем отдельно, а затем конвертируется в агрегированный индекс с помощью нахождения простого среднего» (перевод – *ВМ*) [1]¹, где автор называет этот индекс агрегированным, не используя термин «квартильный индекс». Но по своей сути он квартильный, о чём чётко пишут последователи в работе [2]. Анализ публикаций, полученных при тестировании термина “*quartile index*” в Google Scholar, показал, что, с большой вероятностью, впервые *quartile index* был предложен Р. Кемптоном и Л. Тейлором в 1976 г. в статье «Модели и статистика для разнообразия видов», опубликованной в журнале “Nature” [3].

Помимо экономики и биологии, квартильный индекс наиболее широко используется в медицине, в меньшей степени в психологии и климатологии, что мы наблюдали при просмотре 209 публикаций, полученных с помощью поисковой системы Google Scholar.

В наукометрии этот индекс впервые получен российским историком, профессором А.В. Гринёвым [4, 5]. Он так же как и Ч. Хултен записывает этот индекс в виде:

$$Q_k = \left(\sum k_i n_i \right) / N \quad , \quad (1)$$

но делит дополнительно его на общее число объектов – в данном случае на общее количество публикаций N . В формуле (1) k_i – весовой коэффициент i -й квартили журнала, n_i – количество публикаций автора в журналах i -й квартили; номера квартилей i изменяются от 1 до 4. Весовой коэффициент первой квартили берётся равным 4, уменьшаясь на одну единицу при переходе от первой квартили к четвёртой. С точностью до константы такое изменение весовых коэффициентов с одинаковым шагом эквивалентно распределению аналогичных коэффициентов в работе [1].

Полагаем, что деление на общее количество объектов в выборке публикаций не имеет смысла. Зарубежные ученые при работе с индексом в своих областях знания этого не делают, да и сам А.В. Гринёв, выполнив тестовые расчеты этого индекса для 18 российских академиков историко-филологического отделения РАН, заметил, что индекс (1) «хорошо работает только с относительно большими статистическими выборками» [4].

МАТЕРИАЛЫ, МЕТОДЫ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Квартильный индекс публикационной активности, представляющий собой взвешенную сумму публикаций какого-либо субъекта публикационной активно-

сти, распределённых по квартилям, распространён и на безквартильную категорию. При заданной выборке публикаций нами рассчитаны возможные минимальные и максимальные значения этого индекса. Индекс апробирован на примере публикаций учёных НИУ “БелГУ”, подавших в 2020 г. заявки на стимулирующие выплаты по статьям, опубликованным в журналах, индексированных в базах данных Scopus и Web of Science – квартильный индекс публикационной активности распространён на цитирующие статьи этой выборки публикаций.

Многие университеты сейчас вводят стимулирующие публикационные схемы, в которых вознаграждения за статьи в журналах, индексированных в базах данных Scopus и Web of Science, зависят от квартилей журналов, в которых опубликованы эти статьи. Для мониторинга этого процесса и ранжирования учёных университета или НИИ, в которых внедрены такие схемы, очень удобно применять такой индекс. Он может быть использован в системе мониторинга и оценки качества публикаций на любом уровне – от публикационной активности автора до публикационной активности страны. Но сам индекс требует некоторого развития. Рассмотрим для этого три аспекта.

1. Очевидно, что расчёты по этому индексу можно вести или отдельно по двум вышеуказанным базам данных, или одновременно по этим базам данных. Учитывая, что 80 – 90% журналов, индексируемых в БД Web of Science (естественно, без учёта безквартильного сегмента этой базы данных под названием Emerging Sources Citation Index) входят в БД Scopus, а квартили одних и тех журналов в этих базах данных могут различаться, поэтому необходимо договориться, какую квартиль следует выбирать для расчёта квартильного индекса. Эта проблема возникает при расчёте вознаграждений при стимулирующих публикационных схемах. Обычно из двух разных квартилей выбирается более высокая.

2. Помимо квартильных журналов, в рассматриваемых базах данных, имеется большое множество безквартильных журналов, поэтому их тоже надо обязательно учитывать, как это делается в стимулирующих публикационных схемах.

3. Так же, как импакт-фактор журналов позволяет вычислять средний импакт-фактор журналов, в которых опубликованы статьи автора, и средний импакт-фактор журналов, из которых шли ссылки на эти статьи (эти процедуры реализованы в РИНЦ), аналогичным образом квартильный индекс можно рассчитывать как для всех публикаций автора (коллектива авторов, университета, региона, страны) за определённый период времени, так и для всех его цитирований.

Остановимся подробно на втором пункте усовершенствования квартильного индекса. Для этого, очевидно, вес 5 следует дать статьям первой квартили, а вес 1 – статьям, опубликованным в безквартильных журналах. В работах [1, 4, 5] не использовалась стандартная процедура приведения суммы весов к единице. В нашем случае, даже без нормировки квартильного индекса на общее число публикаций, эта процедура позволит уменьшить разброс его зна-

¹ “The procedure adopted in this paper attempts to deal with aggregation and non-linearity problems by taking each World Bank performance indicator and sorting it into quartiles. The top quartile is assigned a value of 1.00, the second, 0.75, the next 0.50, and the bottom assigned a value 0.25. This produces a quartile ranking for each of the four systems separately, and this is then converted into aggregate index by simple averaging?”

чений. Обозначая веса через w_i , как чаще всего в литературе, запишем вышеуказанное условие в виде

$$\sum_{i=1}^5 w_i = 1, \quad (2)$$

а сам квартильный индекс в виде

$$I_Q = \sum w_i N_i, \quad (3)$$

где N_i – количество публикаций автора в журналах i -й квартили при $i = 1, 2, 3, 4$, а при $i = 5$ количество безквартильных статей равняется N_5 .

Если балльные веса, изменяющиеся от 1 до 5, перевести в весовые коэффициенты w_i с использованием формулы (2), то мы получим: $w_1 = 5/15$, $w_2 = 4/15$, $w_3 = 3/15$, $w_4 = 2/15$, $w_5 = 1/15$. Подставляя эти значения w_i в формулу (3), получим следующую конкретную формулу для расчёта квартильного индекса публикационной активности автора (коллектива авторов и др.):

$$I_Q = (5N_1 + 4N_2 + 3N_3 + 2N_4 + N_5)/15. \quad (4)$$

Посмотрим, в каком интервале будет изменяться квартильный индекс при произвольном количестве публикаций: $N = N_1 + N_2 + N_3 + N_4 + N_5$. Из формулы (4) следует, что

$$\min\{I_Q\} = N/15, \max\{I_Q\} = (1/15)5N = N/3.$$

Таким образом, при произвольном количестве публикаций максимально возможное значение квартильного индекса I_Q превышает его минимальное значение в 5 раз, что следует из соотношения балльных весов квартилей.

Покажем, что нормировка квартильного индекса на общее число публикаций N смысла не имеет. Действительно, пусть первый автор публикует одну статью в журнале Q_1 , тогда при $N = 1$ получим $I_Q = 1/3$, а второй автор публикует две статьи в журналах Q_1 , тогда при $N = 2$ получим $I_Q = 10/15$. Нормировка двух найденных квартильных индексов на общее количество публикаций приводит к одному и тому же результату $1/3$, хотя очевидно, что второй случай более значимый.

Приведём конкретный расчёт квартильного индекса на примере распределения публикаций учёных НИУ “БелГУ”, которые были поданы в 2020 г. на выплаты согласно стимулирующей публикационной схемы, по квартилям. Получено следующее распределение общего числа публикаций в количестве $N = 238$ по квартилям и по безквартильной категории: $N_1 = 36$, $N_2 = 21$, $N_3 = 61$, $N_4 = 52$, $N_5 = 68$, тогда $I_Q = (5 \times 36 + 4 \times 21 + 3 \times 61 + 2 \times 52 + 68)/15 = 619/15 = 41,27$. Для данного количества публикаций N минимально возможное значение квартильного индекса I_Q равняется $238/15 = 15,87$, а максимально возможное – $238/3 = 79,33$. Среднее между ними значение будет 47,6, т. е. фактически квартильный индекс для публикаций НИУ “БелГУ”, поданных на стимулирующие выплаты в 2020 г., ниже его среднего значения.

Рассмотрим в заключение, как рассчитывается квартильный индекс цитируемости статей автора или других субъектов публикационной активности. Пусть N статей имеют C цитирований, которые распределяются следующим образом:

N_1 статей имеют C_1 цитирований из журналов первой квартили (Q_1);

N_2 статей имеют C_2 цитирований из журналов второй квартили (Q_2);

N_3 статей имеют C_3 цитирований из журналов третьей квартили (Q_3);

N_4 статей имеют C_4 цитирований из журналов четвёртой квартили (Q_4);

N_5 статей имеют C_5 цитирований из журналов с отсутствием квартили.

Тогда N_1 статей имеют взвешенное (балльное) цитирование $5C_1$; N_2 статей – $4C_2$; N_3 статей – $3C_3$; N_4 статей – $2C_4$; N_5 статей – C_5 , а все N статей имеют суммарное балльное цитирование $I_{CQ} = 5C_1 + 4C_2 + 3C_3 + 2C_4 + C_5$. При переходе к весовым коэффициентам, сумма которых равна единице, получим квартильный индекс

$$I_{CQ} = (5C_1 + 4C_2 + 3C_3 + 2C_4 + C_5)/15. \quad (5)$$

Эта формула, по своей структуре, аналогична формуле (4).

ЗАКЛЮЧЕНИЕ

Таким образом, очевидно, что в наукометрии квартильный индекс имеет большие преимущества перед индексом Хирша и другими хиршеподобными индексами, так как учитывает весь спектр публикаций автора или другого субъекта публикационной активности, а также качественную структуру публикаций на основе распределения их по квартилям.

В настоящее время квартильный индекс особенно актуален, так как Минобрнауки РФ и, соответственно, руководители научно-исследовательских и научно-образовательных учреждений ориентируют ученых публиковаться в высокочастотных журналах.

Полагаем, что агентства, генерирующие базы данных наукометрических индикаторов (Clarivate (WoS), Elsevier (Scopus), Scimago Journal&Country Rank и другие), должны обратить внимание на этот индекс, а Минобрнауки РФ и РАН могли бы внедрить его в практику мониторинга публикационной активности и оценки её качества.

СПИСОК ЛИТЕРАТУРЫ

1. Hulten C.R. Infrastructure Capital and Economic Growth: How Well You Use It May Be More Important than How Much You Have. NBER Working Paper 5874. – Cambridge, United States: National Bureau of Economic Research, 1966. – 37 p.
2. Chong A., Calderon C. Volume and Quality of Infrastructure and the Distribution of Income: An Empirical Investigation // Income and Wealth. – 2004. – Vol. 50, №1. – P. 87-106.
3. Kempton R.A., Taylor L.R. Models and statistics for species diversity // Nature. – 1976. – Vol. 262. – P. 818-820.

4. Гринёв А.В. Публикационная активность ведущих российских историков в ББД Scopus и квартильный индекс // КЛИО. – 2019. – №11(155). – С. 36-47.
5. Grinev A.V. The Disadvantages of Using Scientometric Indicators in the Digital Age // IOP Conference Series: Material Science and Engineering. – 2020. – Vol. 940, article №012149.

Сведения об авторе

МОСКОВКИН Владимир Михайлович – доктор географических наук, директор Центра развития публикационной активности, профессор кафедры мировой экономики Белгородского государственного национального исследовательского университета
e-mail: moskovkin@bsu.edu.ru

Материал поступил в редакцию 21.05.21.