**Введение**

Произошедшее в прошлом веке усложнение структуры науки, выражающееся в появлении большого числа междисциплинарных направлений исследования, увеличении количества исследователей привело к необходимости количественной оценке эффективности деятельности отдельных ученых и их коллективов. В это время сформировалась наукометрия как отдельное направление исследований. В настоящее время результаты наукометрии активно начинают использоваться не просто для оценки результаты работы конкретного ученого, но и как основание для принятия разного рода административных решений, включая принятие решений о выдаче грантов, присуждение научных степеней и т.д.

В настоящее время различные наукометрические показатели используются для:

- сравнения ученых в какой-либо области [alonso2009- Hirsch (2005), Alcaide, Гомес, Zurian и Benavent, (2008 ];

- оценки «экспертности» и влиятельности ученых [alonso2009- Оппенгейм ( 2007)]

- измерения научной эффективности исследовательских коллективов и целых отраслей [alonso2009- Egghe & Rao, 2008; J. Molinari & A. Molinari, 2008; A. Molinari & J. Molinari, 2008; Мугнайни, Пакер и Менегини (2008)];

- ранжирования научных журналов [alonso2009- Braun, Glanzel и Шуберт (2006), Vanclay ( 2008)];

- идентификации наиболее популярных тем исследований [alonso2009- Банка ( 2006)];

- оценки разницы в эффективности работы ученых разных стран [alonso2009- Гуань и Гао (2008); Гуань и Гао (2008)];

- оценки структуры и связей в социальных сетях исследователей [cimenler2014];

- создания показателей (TRIF), предсказывающие цитирования и уровень интереса научного сообщества к определенной работе [vanclay2013];

- оценки карьеры различных авторов, исследований в области современной истории науки [wu2011];

- оценки скорости цитирования и других различных характеристик цитирования [oravec2019];

- исследования вопросов о том, что влияет на цитирование статьи [peng2012].

Последний вопрос является очень важным в том числе и с философской точки зрения и связан с фундаментальными взглядами на науку. Согласно классическому взгляду [meyer2017-(Макробертс & MacRoberts, 1996], наука ищет истину, а научное знание основано на соответствующих фактах. Следствием этого является то, что научная обоснованность идеи не зависит от того, кто ее высказал, наука не должна зависеть от статуса или других личных качеств ученых. Это подразумевает, что в научной среде должна отсутствовать дискриминация, например, по признаку национальности или пола. Ученые, которые используют знания, установленные другими, должны отдать им должное, особенно ссылаясь на их работу. В идеале, эти нормы гарантируют объективность научного знания и эпистемологическую легитимность науки. Однако исследования в рамках теории социального конструктивизма, возникшей в результате развития социологии научных знаний [meyer2017 -Longino, 2016] показали, что научные знания и принятие результатов работы исследователей зависит также от социальных норм, институтов, ценностей, убеждений [meyer2017 - Barnes, 1977; Bloor, 1976]. Другие исследования использовали этнографические методы для исследования того, какие социальные факторы влияют на принятие чего-либо как научного факта [meyer2017 - Latour & Woolgar, 1986 ], некоторые работы обсуждали потенциальную гендерную предвзятость в научном сообществе [meyer2017 -Хардинг, 1992].

Передний край наукометрии как науки сегодня – исследования соотношений объективного и субъективного в науке как структуре. Эти исследования напрямую касаются базовых вопросов философии научного знания. Тот факт, что социальное измерение науки может вызвать уклон и искажения в результатах научных исследований потенциально приводит к вопросу о нужности и эффективности науки в ее современном виде [meyer2017].

В российском научном сообществе отношение к наукометрии двойственное – с одной стороны, существует большое количество исследователей, справедливо указывающих на некоторые недостатки применяемого наукометрией инструментария [6483, 6483-4, 7787, 0872, 8987], но существуют и отдельные интересные работы – например, посвященные вопросам «нулевого цитирования» [2444]. Однако, по нашему мнению, будущие исследования в данной области будут опираться на результаты и инструментарий наукометрии как на вспомогательный инструмент для проверки других гипотез.

**Материалы и методы**

Применение теории социального конструктивизма к научному сообществу привело к исследованию субъективных факторов в науке, т.е., по сути дела, к исследованию культуры научного сообщества, которая порождает те или иные субъективные факторы. При это единичным носителем такой культуры является определенная научная организация, поскольку внутри каждой такой организации существует определенная сформировавшаяся культура научных исследований.

Целью данного исследования является изучение различных факторов, влияющих на качественные и количественные показатели эффективности научной деятельности ученых в ГБОУ ВО «Нижегородский государственный инженерно-экономический университет».

Вопрос о показателях, объективно оценивающих результаты научной деятельности, остается открытым до сих пор. Один из самых успешных показателей был предложен Хиршем в работе [Hirsch(2005)] и получил название h-индекса. Основные преимущества h-индекса заключаются в том, что его легко вычислить, и он учитывает как количество, так и цитируемость (что многими воспринимается как адекватная оценка качества) публикаций исследователя. Практика использования данного показателя показала, что он работает лучше, другие классические показатели, использовавшиеся в наукометрии ранее (импакт-фактор, общее количество статей, общее количество цитирований, количество цитирований в расчете на статью и количество высоко цитируемых статей), кроме того, трудность увеличения h-индекса растет экспоненциально, что существенно затрудняет его искусственный рост [alonso2009].

Все эти соображения обосновали выбор семейства показателей, основанных на h-индексе, как зависимых факторов (поскольку они оценивают качественные и количественные аспекты научной работы ученого).

Вопросы состава факторов, влияющих на показатели факторов, обсуждались в большом количестве статей. Общепринятыми являются такие факторы как репутация автора (mingers2010-Allison and Long, 1990; Long и др., 1998), репутация учреждения (mingers2010-Stahl et al., 1988; Trieschmann) и др., 2000), репутация журнала (mingers2010-Franke et al., 1990; Podsakoff et al., 2005). В [mingers2010 - Недерхоф и ван Раан (1987)] было доказано предположение о существование эффекта Мэтью (большое количество цитат приводит к хорошей репутации, и эта хорошая репутация привлекает еще больше цитат).

В некоторых областях последние работы цитируются более часто, чем старые. Влияние таких показателей как язык, количество таблиц и рисунков, было рассмотрено в статье (mingers2010 - Stremersch et al., 2007 ). Также в исследовании [mingers2010] упомянуто наличие следующих влияющих факторов: наличие статей в свободном доступе, тираж журналов, наличие в них рецензирования, количество авторов в статье, национальность и пол автора, его возраст, членство в редакциях журналов, применяемые эмпирические методы, длина заголовка статьи и самой статьи, количество ссылок в самой статье и др.

В [onodera2014] в качестве влияющих переменных рассматривались: количество авторов; количество учреждений, которые представляют авторы; количество стран, которые представляют авторы, количество статей, опубликованных первым автором до публикации работы; количество ссылок на статьи, полученные к моменту публикации целевой статьи, количество времени работы первого автора до публикации статьи, общее количество ссылок, длина статьи, количество математических формул в статье [onodera2014].

Вопрос о том, влияет ли пол на цитируемость статей, детально изучался в некоторых исследованиях, например, в [thelwall2018]. Согласно результатам данной работы, в Испании, Великобритании и США нет существенных предубеждений против статей, написанных женщинами (сделана оговорка, что возможно существование некого предубеждения, но ему противодействуют другие факторы, которые делают исследования, проводимые женщинами, более эффективными), но этот эффект отмечается, например, в Индии и Турции.

В работе [copiello2019] в качестве предикторов использовались общее количество лет деятельности автора, его возраст, количество статей, количество докладов на конференции, количество соавторов

В исследовании [meyer2017 - Stremersch и соавт. (2007)] изучали, какие факторы влияют на частоту цитирования в пяти основных маркетинговых журналах, обнаружив, что в числе таковых выступают длина

В работе [peng2012], посвященном предикторам цитирования в интернет-журналах, проверялись следующие гипотезы:

* чем популярнее тема работы, тем больше ссылок получает исследование.
* чем объемнее исследование, тем больше ссылок исследование получает.
* чем более авторов в работе, тем больше цитаты, которые получает исследование
* исследования с первыми / единственными авторами из США. цитируются чаще, чем те, у кого первый / единственный автор из других стран.
* чем более авторитетные работы цитируют исследование, тем больше оно получает цитат.
* чем престижнее журнал, тем больше цитат получает исследование в нем.
* чем больший процент статей журнала публикуются в интернете, тем больше цитат получают соответствующие исследования
* чем междисциплинарнее журнал, тем больше цитат получают статьи в нем
* чем выше уровень самоцитирования в журнале, тем больше цитат получают статьи в нем
* связь между временем публикации и цитированием определяется факторами уровня журнала, включая престиж журнала, видимость в интернет-исследованиях, степень сконцнтрированности на одной дисциплине, уровень самоцитирования.

В качестве зависимой переменной предлагается использовать классический h-индекс, но он не лишен недостатков, в числе которых выделяют существенные различия в величине индекса для исследований в разных науках, зависимость от продолжительности карьеры каждого ученого (alonso2009 - Kelly & Jennions, 2006), учет самоцитирования, отсутствие наличия контекста цитирования, неспособность измерения качественных признаков количественными методами [0872].

Для устранения этих недостатков разными учеными предлагалось использование разных модификаций индекса Хирша [3065, 0951]. Мы считаем, что в целях нашего исследования необходимо использовать так называемый m-индекс, который рассчитывается как отношение h-индекса исследователя к числу лет, прошедших после первой публикации исследователя [9492-7].

Таким образом, мы исследуем предикторы индекса Хирша и m-индекса среди исследователей Нижегородского государственного инженерно-экономического университета на основании данных российского индекса научного цитирования. В качестве независимых переменных нами были взяты:

- ученая степень исследователя

- место работы исследователя (структурное подразделение университета)

- статус исследователя (совместитель или штатный сотрудник)

- период публикаций

- число публикаций в РИНЦ

- процент публикаций автора, процитированных хотя бы раз

- средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи

- область науки, в которой автором написано наибольшее количество статей.

При этом существует важный методологический вопрос – поскольку распределение показателей цитирования / индекса Хирша обычно смещено и не является нормальным [Thelwall14,16 - de Solla Price, 1965; Seglen, 1992, 7787, 1453], классический метод наименьших квадратов не должен использоваться в регрессии. В научной литературе использовались несколько методов:

- использовалась модель обобщенной линейной регрессии [mingers2010, Thelwall14 - Dobson & Barnett, 2008]

- использовалась отрицательная биноминальная множественная регрессия [onedera2014 - (Bornmann & Daniel, 2008; Chen, 2012; Дэвис и др., 2008; Didegah & Thelwall, 2013; Он, 2009; фургон Dalen & Henkens, 2001, 2005; Уолтерс, 2006, Chen, 2012; Didegah & Thelwall, 2013a, 2013b; Макдональд, 2007].

- использовалась логистическая регрессия [Thelwall14 - Baldi, 1998; Bornmann & Williams, 2013; Kutlar, Kabasakal & Ekici, 2013; Грех, 2011; Уиллис, Бахлер, Нойбергер и Дам, 2011; Ся & Наканиси, 2012; Yu, Yu, & Wang, 2014. Baumgartner & Leydesdorff, 2014].

При этом исследование вопросов о распределении показателей цитирования показало, что оно соответствует дискретному логнормальному распределению (Thelwall14 - Эванс, Хопкинс и Каубе, 2012; Radicchi, Fortunato, & Castellano, 2008; Тельвол и Уилсон, 2014 ), степенному закону для хвоста распределения [Thelwall14 - Яо, Пэн, Чжан и Сюй, 2014].

При этом в литературе указывается, что использование пакета GAMLSS в R может использоваться для расчета коэффициентов обобщенной линейной модели на основе при непрерывном логнормальном распределении [Thelwall14,16 - Stasinopoulos & Rigby, 2007].

Таким образом, на первом шаге необходимо проанализировать распределение h и m – индексов, на основе чего принимать решения о применяемых методах.

**Результаты**

Рассмотрим распределение показателей (рисунок 1).

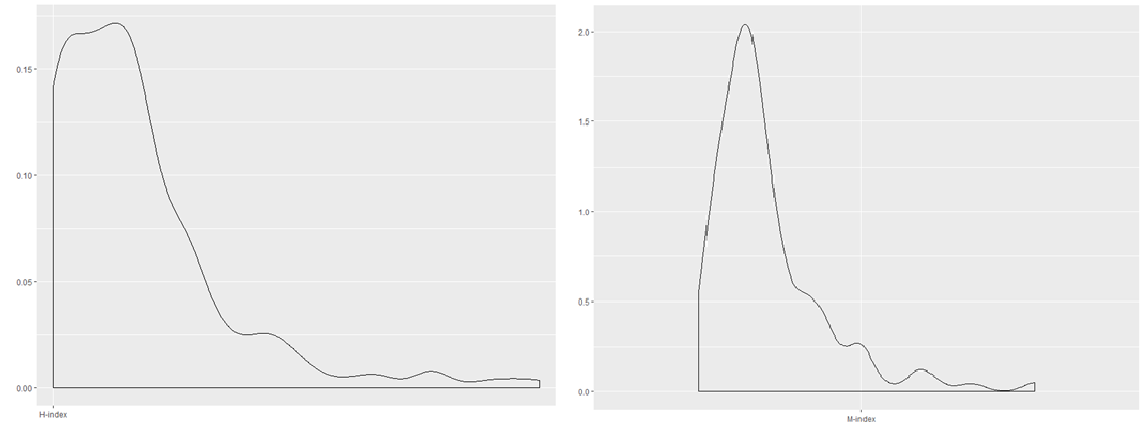


Рис. 1 – Распределение значений h-индекса и m-индекса среди ученых ГБОУ ВО «Нижегородский государственный инженерно-экономический университет»

На рисунке видно, что распределение обоих показателей смещено сильно влево - в этих условиях нецелесообразно использовать обычный метод наименьших квадратов. Кроме того, индекс Хирша по своим свойствам является счетной переменной (принимает только неотрицательные целые значения). Дальнейшие исследования показали, что распределение индекса Хирша по изучаемой совокупности лучше всего описывается пуассоновским распределением с параметром λ=4.0851 (статистика Колмогорова-Смирнова 0.11878). Следовательно, при решении задачи построения уравнений регрессии H-индекса логичнее всего использовать пуассоновскую регрессию.

Расчеты производились с использованием функции glm статистической среды R. Поскольку зависимые переменные представляют собой смесь непрерывных и факторных переменных, и существует большое количество различных комбинаций влияния, сначала была построена полная модель, включающая в себя все возможные варианты эффектов от факторных переменных (в том числе и возможное влияние на изменение коэффициентов регрессии), а далее с помощью step-wise регрессии была найдена оптимальная модель, не содержащая статистически не значащих эффектов. В качестве критерия, оценивающего оптимальность построенных моделей, выступал информационный критерий Акаике (AIC).

Однако, поскольку данные содержали избыточную дисперсию, потом полученная модель была пересчитана с предположением о том, что распределение является квазипуассоновким [Wetherill, G.B. and Brown, D.W. (1991) Statistical Process Control, New York, Chapman and Hall, pp. 216–218, Кабакофф], что позволяет получить более точные оценки значимости коэффициентов. Таким образом, с учетом того, что пуссановская регрессия моделирует логарифм условного среднего зависимой переменной, было получено следующее уравнение регрессии:

, (1)

где N – количество статей в РИНЦ

S – доля статей, процитированных хотя бы 1 раз

I – средний импакт-фактор журнала, в котором были опубликованы статьи

d0 – бинарная переменная, равная 1, если у исследователя нет ученой степени

d1 – бинарная переменная, равная 1, если у исследователя есть степень кандидата наук

d2 – бинарная переменная, равная 1, если у исследователя есть степень доктора наук

Потенцированные коэффициенты модели, а также информация об их статистической значимости представлена в таблице 1.

Таблица 1 – Информация о статистической значимости модели (1)

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр | Коэффициент в уравнении модели | Значение после потенцирования модели | p-значение |
| Свободный член | -1.592 | 0.204 | < 2e-16 |
| ln(Количество статей) | 0.605 | 1.831 | < 2e-16 |
| Доля статей, процитированных хотя бы 1 раз | 0.019 | 1.02 | < 2e-16 |
| Средний импакт-фактор журнала, в котором были опубликованы статьи (нет ученой степени) | 0.426 | 1.53 | 7.33e-11 |
| Средний импакт-фактор журнала, в котором были опубликованы статьи (есть ученая степень кандидата наук) | -0.22 | 0.8 | 0.000337 |
| Средний импакт-фактор журнала, в котором были опубликованы статьи (есть ученая степень доктора наук) | -0.3 | 0.74 | 0.022829 |

В отличии от h-индекса, m-индекс не является счетной переменной, но его распределение сильно скошено влево. Статистический анализ показал, что 25%-ый квантиль m-индекса находится на уровне 0.25, медиана – на уровне 0.375, 75%-ый квантиль m-индекса находится на уровне 0.55. Для анализа факторов, влияющих на m-индекс, мы применили инструментарий квантильной регрессии для каждого из квантилей. Сложностью оценки достоверности квантильной регрессии является отсутствие общепринятых стандартных формул для вычисления p-значений, поэтому критерием достоверности в данном случае являлось отсутствие 0 в доверительном интервале значений параметров медианной регрессии. Таким образом, были получены 3 модели квантильной регрессии:

(2)

(3)

(4)

где E – время, прошедшее с выхода первой публикации

N – количество статей в РИНЦ

S – доля статей, процитированных хотя бы 1 раз

s0 – бинарная переменная, равная 1, если исследователь работает только в данной организации

s1 – бинарная переменная, равная 1, если исследователь работает также в другой организации

Доверительные интервалы коэффициентов представлены в таблице 2

Таблица 2 – Значения коэффициентов и их доверительных интервалов в моделях (2)-(4)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Показатель | Модель 1 | | | Модель 2 | | | Модель 3 | | |
| coefficients | lower bd | upper bd | coefficients | lower bd | upper bd | coefficients | lower bd | upper bd |
| Стаж публикационной деятельности | -0.023 | -0.037 | -0.018 | -0.033 | -0.044 | -0.027 | -0.047 | -0.057 | -0.035 |
| ln(Количество статей) | 0.143 | 0.117 | 0.188 | 0.158 | 0.144 | 0.198 | 0.188 | 0.148 | 0.204 |
| Доля статей, процитированных хотя бы 1 раз | 0.007 | 0.005 | 0.007 | 0.008 | 0.007 | 0.009 | 0.01 | 0.009 | 0.013 |
| Исследователь работает только в университете | -0.154 | -0.24 | -0.055 | -0.077 | -0.138 | -0.018 | - | | |
| Исследователь работает также в другой организации | -0.12 | -0.316 | -0.012 | -0.082 | -0.128 | 0.068 |

**Обсуждение**

На основании полученного уравнения (1) можно констатировать следующее:

- значение индекса Хирша ученых исследуемой организации в большей степени зависит от количества их статей и доли их процитированных статей – это говорит о наличии эффекта Мэтью;

- для ученых без ученой степени показатели импакт-фактора журнала, в котором они публикуются, намного больше влияет на их индекс Хирша, нежели для исследователей с ученой степенью. Это подтверждает наличие внешних общих субъективных факторов, выделенных социологией науки; кроме того, это может свидетельствовать о наличии сложных репутационных эффектов, объясняемых тем фактом, что после защиты диссертационной работы у автора складывается определенная репутация в научных кругах, и начинают другие факторы влиять на показатели его цитируемости;

- не выявлено наличие внутренних субъективных организационных факторов, прямо влияющие на показатель индекса Хирша, например – структурное подразделение, ученое звание, предпочитаемая область исследования.

На основании полученного уравнений (2-4) можно констатировать следующее:

- стаж публикационной активности положительно влияет на m-индекс (уменьшается вероятность того, что m-индекс будет низким), причем сила этого влияния увеличивается пропорционально квантидю – это говорит о том, что существует тенденция увеличения общего m-индекса при увеличении времени нахождения исследователя в науке;

- количество статей отрицательно влияет на m-индекс (увеличивается вероятность того, что m-индекс будет низким), причем сила этого влияния увеличивается пропорционально квантидю – это говорит о том, что существует тенденция снижения m-индекса при увеличении количества исследований;

- доля статей, на которые есть цитаты, отрицательно влияет на m-индекс (увеличивается вероятность того, что m-индекс будет низким), причем сила этого влияния увеличивается пропорционально квантидю;

- для исследователей с низким m-индексом существует внутренний организационный эффект, заключающийся в том, что для совместителей ниже вероятность большего значения m-индекса, что говорит о наличии в организации соответствующей публикационной культуры.

Таким образом, можно сделать следующий вывод: несмотря на наличие классических эффектов, выявленных исследованиями в области социологии науки, в ГБОУ ВО НГИЭУ существует сбалансированная публикационная культура, обеспечивающая отсутствие дискриминации какой-то группы исследователей по количественным и качественным показателям цитирования (несколько удивительным может быть отсутствие разницы в данных показателях для исследователей в разных областях знания).

**Заключение**

Достаточно большое количество разнородных результатов исследований, дифференцированно описывающих влияние различных факторов на показатели цитирования, говорит о необходимости продолжения исследований в области квалиметрии наукометрии и разработке системы показателей, максимально полно описывающих количественные и качественные результаты научных работ исследований.

При этом использованный нами подход восприятия наукометрических индексов исследователей в отдельной организации с целью описания внутренней публикационной культуры является, по мнению авторов, перспективным с точки зрения исследования организаций и научных коллективов, и может быть использован другими организациями. Также, по нашему мнению, в ближайшее время получит распространение идея использования показателей цитирования и др. как промежуточных, являющихся средством, а не объектом изучения.