
СТАТИСТИКА И ЭКОНОМИЧЕСКОЕ ИЗМЕРЕНИЕ

УДК 311

О СПОСОБЕ ПОСТРОЕНИЯ ДИНАМИЧЕСКИ СОПОСТАВИМЫХ КОМПЗИТНЫХ ИНДЕКСОВ

Д.А. Борзых, К.К. Фурманов, И.К. Чернышева

Национальный исследовательский университет

Высшая школа экономики

E-mail: dborzykh@hse.ru

В статье предложен новый способ построения динамических композитных индексов, обладающих двумя важнейшими свойствами – свойством неотрицательности весов и свойством динамической сопоставимости. Апробирование данной методологии проведено на примере построения композитного индекса здоровья населения. Проведено сопоставление с существующими методами построения композитных индексов – с методом главных компонент и методом, предложенным Д.А. Борзых в работе 2016 г. Установлено, что метод главных компонент не обладает ни свойством неотрицательности весов, ни свойством динамической сопоставимости, а метод, предложенный ранее Д.А. Борзых, не обладает только свойством динамической сопоставимости. Таким образом, показано, что предлагаемый вниманию метод построения динамических композитных индексов обладает лучшими свойствами.

Ключевые слова: композитный индекс, динамическая устойчивость, динамическая сопоставимость, метод главных компонент, композитный индекс здоровья населения.

ON THE METHOD OF CONSTRUCTING DYNAMICALLY COMPARABLE COMPOSITE INDICES

D.A. Borzykh, K.K. Furmanov, I.K. Chernysheva

National Research University Higher School of Economics

E-mail: dborzykh@hse.ru

This paper proposes a new method for constructing dynamic composite indices that has two important properties, such as nonnegative weights and dynamic comparability. The method is applied to construct a composite index of population health, as an example. The current method of composite indices' construction is compared to existing approaches, namely the principal component method and the technique suggested by D. Borzykh in 2016. In contrast to the latter, the new approach meets both requirements, i.e. dynamic comparability and nonnegativity of weights. Thus, it is shown that the method suggested in this paper has the best properties.

Keywords: composite index, dynamical stability, dynamical comparability, method of principal components, composite index of population health.

Смысл понятия «композитный индекс» (composite index) интуитивно ясен. Тем не менее, если попытаться найти в литературе четкое определение этого понятия, можно обнаружить такое: «Композитный индекс – это объединение акций, индексов или других факторов, сгруппированных стандартизированным образом и представляющих собой статистическую меру общей ситуации на рынке или в экономическом секторе во времени...» [12]. В литературе встречаются также синонимичные понятия – интегральный индикатор, интегральный показатель, интегральная характеристика, синтетическая категория [1, 3, 4, 7].

Задача построения композитных индексов является одной из актуальных проблем ряда общественных наук – экономики, социологии, социально-экономической статистики. Это связано с тем, что композитные индексы, как правило, применяются для мониторинга ситуации в той или иной сфере деятельности человека. Так, например, в публикациях [1, 2, 4, 15] строится интегральный индекс качества жизни населения. В работе [13] определяется интегральная характеристика развития издательского дела, доступности и распространенности кино. В статье [8] изучаются возможности применения и расчета композитных индексов в области экономики знаний. При этом возможные применения композитных индексов данными публикациями не ограничиваются.

Несмотря на то, что суть и назначение композитных индексов ясны и понятны, конкретная реализация всегда сопряжена с рядом методологических трудностей. Среди этих трудностей можно выделить следующие:

- отбор исходных показателей, образующих композитный индекс;
- нормировка исходных показателей с целью их сопоставимости между собой;
- выбор способа расчета весовых коэффициентов;
- обеспечение сопоставимости получаемых индексов во времени.

На сегодняшний день проблема выбора исходных показателей, определяющих состав композитного индекса, решается исключительно субъективно. Авторы исследований выбирают исходные показатели, ориентируясь на внутреннее интуитивное ощущение важности того или иного показателя, а также на доступность соответствующих данных.

Проблема нормировки исходных показателей имеет два общепринятых решения: «стандартизация» (вычитание среднего и деление на стандартное отклонение) и линейное преобразование к единой шкале, например, к шкале от нуля до единицы. Другие варианты нормировки исходных показателей, такие как преобразование с помощью нелинейной функции, а также с использованием «эталонного» порогового значения, рассматриваются в работе [7]. Там же обсуждаются и недостатки некоторых способов нормировки.

Проблема выбора весов на данный момент также не имеет окончательного оптимального решения. В прикладных исследованиях веса зачастую определяются с использованием экспертных мнений. Такой подход обсуждается и критикуется в работе [7]. В научной литературе широко применяются метод главных компонент [10, 14] и модифицированный метод главных компонент [2–4]. Альтернативный подход (по отношению к обычному и модифицированному методу главных компонент) был предложен агент-

ством CFED в задаче построения рейтингов американских штатов [11]. В работах [5, 6] Д.А. Борзых модифицировал методику из публикации [11] и применил ее при решении задачи рейтингования регионов РФ по уровню здоровья населения и задачи количественного анализа динамики уровня здоровья населения России.

Проблема сопоставимости композитных индексов во времени состоит в том, что композитные индексы далеко не всегда могут быть корректно сравнимы в разные периоды времени. Данная проблема упоминается в статье [7]. Тем не менее подробного анализа этой проблемы в [7] нет. В отличие от работы [7], в данной статье концентрируется основное внимание именно на проблеме сопоставимости. В частности, в разделе 1 мы выделили и сформулировали формальное определение динамической сопоставимости и подробно исследовали это свойство в разделах 3 и 4.

Обратимся теперь к недостаткам, присущим существующим методам расчета композитных индексов. В данной публикации концентрируемся на двух таких недостатках – отсутствии свойства неотрицательности весов и отсутствии свойства динамической сопоставимости – и предлагаем способ их преодоления.

Теперь опишем структуру статьи и перечислим ее основные результаты.

В разделе 1 статьи предложен новый способ построения динамических композитных индексов, обладающих двумя важнейшими свойствами – свойством неотрицательности весов и свойством динамической сопоставимости.

В разделе 2 проведено апробирование предлагаемой методики на примере построения композитного индекса здоровья населения.

В разделе 3 проведено сопоставление предлагаемого метода с методом из работы [6]. Показано, что метод из работы [6], в отличие от описанного в данной статье, не обладает свойством динамической сопоставимости.

В разделе 4 выполнено сравнение предлагаемого метода с широко известным методом главных компонент. Установлено, что метод главных компонент не обладает ни свойством неотрицательности весов, ни свойством динамической сопоставимости.

1. МЕТОДОЛОГИЯ ПОСТРОЕНИЯ ДИНАМИЧЕСКИ СОПОСТАВИМЫХ КОМПОЗИТНЫХ ИНДЕКСОВ

В этом разделе излагается методика построения композитных индексов, которая представляет собой модификацию подхода, предложенного в работе [6] Д.А. Борзых. Данная модификация сочетает в себе основные достоинства предыдущей методики, и при этом она лишена обнаруженного недостатка прежней методики – отсутствия свойства *динамической сопоставимости* (см. далее в этом разделе). Сравнение «старой» и «новой» методик, а также свойство динамической сопоставимости обсуждаются в разделе 3.

Перейдем теперь непосредственно к изложению «новой» методики построения композитных индексов. Обозначим через $X_{i,t}^{(j)}$ показатель с номером $j \in \{1, \dots, k\}$, относящийся к объекту $i \in \{1, \dots, n\}$ в момент времени $t \in \{1, \dots, T\}$.

Будем предполагать, что для каждого $j \in \{1, \dots, k\}$ выполнены следующие условия:

- 1) $\text{med}_{i,t} X_{i,t}^{(j)} > 0$,
- 2) $\min_{i,t} X_{i,t}^{(j)} \geq 0$,
- 3) $\min_{i,t} X_{i,t}^{(j)} < \max_{i,t} X_{i,t}^{(j)}$,

где через $\text{med}_{i,t} X_{i,t}^{(j)}$ обозначена выборочная медиана массива данных $\{X_{i,t}^{(j)}\}_{i=1, \dots, n}^{t=1, \dots, T}$.

Определим композитный индекс $I_{i,t}$, относящийся к объекту i в момент времени t :

$$I_{i,t} = w^{(1)} Z_{i,t}^{(1)} + \dots + w^{(k)} Z_{i,t}^{(k)}, \quad (1)$$

где веса $w^{(1)}, \dots, w^{(k)}$ вычисляются как отношения

$$w^{(j)} := \frac{v^{(j)}}{v^{(1)} + \dots + v^{(k)}}, \quad j = 1, \dots, k, \quad (2)$$

в которых рассчитываются величины $v^{(1)}, \dots, v^{(k)}$:

$$v^{(j)} = \frac{\text{med}_{i,t} |X_{i,t}^{(j)} - \text{med}_{i,t} X_{i,t}^{(j)}|}{\text{med}_{i,t} X_{i,t}^{(j)}}, \quad j = 1, \dots, k. \quad (3)$$

При этом нормализованные показатели $Z^{(1)}, \dots, Z^{(k)}$, участвующие в формуле (1), определяются следующим образом:

$$Z_{i,t}^{(j)} := \begin{cases} \frac{X_{i,t}^{(j)} - \min_{i,t} X_{i,t}^{(j)}}{\max_{i,t} X_{i,t}^{(j)} - \min_{i,t} X_{i,t}^{(j)}}, & \text{если чем больше показатель } X^{(j)}, \text{ тем лучше,} \\ \frac{\max_{i,t} X_{i,t}^{(j)} - X_{i,t}^{(j)}}{\max_{i,t} X_{i,t}^{(j)} - \min_{i,t} X_{i,t}^{(j)}}, & \text{если чем меньше показатель } X^{(j)}, \text{ тем лучше.} \end{cases} \quad (4)$$

Если композитный индекс I является многоярусным, т.е. строится на основе подындеков $I^{(1)}, \dots, I^{(k)}$, то для его вычисления применяется та же процедура с тем лишь отличием, что в качестве показателей $X^{(1)}, \dots, X^{(k)}$ теперь используются подындексы.

Из формулы (3) видно, что веса $w^{(j)}$, $j \in \{1, \dots, k\}$, участвующие в формуле (1), всегда неотрицательны. Это важное свойство позволяет легко интерпретировать результаты расчетов. Забегая вперед, отметим, что, например, широко известный метод главных компонент не обладает данным свойством.

Продолжим анализировать формулы (1)–(4). Заметим, что величина

$$\frac{\text{med}_{i,t} |X_{i,t}^{(j)} - \text{med}_{i,t} X_{i,t}^{(j)}|}{\text{med}_{i,t} X_{i,t}^{(j)}},$$

участвующая в формуле (2), представляет собой робастный (т.е. устойчивый к выбросам) аналог коэффициента вариации, который определяется как от-

ношение стандартного отклонения случайной величины и ее математического ожидания $\sqrt{D[X]} / E[X]$. В самом деле, числитель $\text{med}_{i,t} \left| X_{i,t}^{(j)} - \text{med}_{i,t} X_{i,t}^{(j)} \right|$ есть робастный аналог стандартного отклонения $\sqrt{D[X]}$, а знаменатель $\text{med}_{i,t} X_{i,t}^{(j)}$ – это робастный аналог математического ожидания $E[X]$.

Таким образом, собирая вместе только что сказанное и формулу (1), можно заключить, что при построении композитных индексов мы руководствуемся принципом максимума вариации: чем выше изменчивость показателя, тем бóльшую значимость мы придаем данному показателю в индексе.

Этот подход основан на следующем соображении: если значение некоторого показателя для всех объектов одно и то же, то по этому показателю мы не сможем предпочесть ни один из объектов по отношению к другим. Аналогичная логика применима и в случае, если рассматриваемый показатель не является константой, но меняется достаточно слабо.

Апробирование предлагаемой методики проведено в разделе 2 на примере построения композитного индекса здоровья населения.

Как уже отмечалось в работах [5, 6], между нашим подходом и методом главных компонент имеется некоторое идейное сходство: так же, как и метод главных компонент, наш подход основан на принципе максимизации изменчивости некоторых переменных. Однако в техническом плане реализации предлагаемого нами способа и метода главных компонент совершенно разные. Отсюда и столь существенные различия в свойствах получаемых композитных индексов. Подробное обсуждение метода главных компонент и его сопоставление с нашим подходом приведено в разделе 4 настоящей работы.

Перейдем теперь к свойству динамической сопоставимости.

Определение. Пусть индекс $I_{i,t}$, относящийся к объекту i в момент времени t вычислялся на основе показателей $X^{(1)} = (X_{i,t}^{(1)})_{i=1,\dots,n}^{t=1,\dots,T}, \dots, X^{(k)} = (X_{i,t}^{(k)})_{i=1,\dots,n}^{t=1,\dots,T}$, т.е. индекс $I_{i,t}$ есть функция аргументов $X^{(1)}, \dots, X^{(k)}$. Назовем индекс I динамически сопоставимым, если выполнено условие: для любого $i \in \{1, \dots, n\}$ и любых $s, t \in \{1, \dots, T\}$ из условия $(\forall j \in \{1, \dots, k\} : \delta^{(j)} \cdot X_{i,s}^{(j)} \leq \delta^{(j)} \cdot X_{i,t}^{(j)})$ следует неравенство $I_{i,s} \leq I_{i,t}$, где

$$\delta^{(j)} := \begin{cases} 1, & \text{если чем больше показатель } X^{(j)}, \text{ тем лучше,} \\ -1, & \text{если чем меньше показатель } X^{(j)}, \text{ тем лучше.} \end{cases}$$

Несложно заметить, что в силу монотонности функций

$$f(x) = \frac{x - \min_{i,t} X_{i,t}^{(j)}}{\max_{i,t} X_{i,t}^{(j)} - \min_{i,t} X_{i,t}^{(j)}} \quad \text{и} \quad g(x) = \frac{\max_{i,t} X_{i,t}^{(j)} - x}{\max_{i,t} X_{i,t}^{(j)} - \min_{i,t} X_{i,t}^{(j)}}$$

определяемый формулами (1)–(4) композитный индекс I обладает указанным свойством динамической сопоставимости. Более подробно свойство динамической сопоставимости обсуждается в разделе 3.

2. ПОСТРОЕНИЕ ДИНАМИЧЕСКИ УСТОЙЧИВОГО КОМПОЗИТНОГО ИНДЕКСА НА ПРИМЕРЕ КОМПОЗИТНОГО ИНДЕКСА ЗДОРОВЬЯ НАСЕЛЕНИЯ

В данном разделе продемонстрируем описанную в разделе 1 методику расчета композитных индексов на примере построения композитного индекса здоровья населения по региональной статистике РФ за 2005–2013 гг. Источник данных – статистический бюллетень [9].

Указанный композитный индекс здоровья населения рассчитывается на основе двух подындексов – индекса заболеваемости и индекса продолжительности жизни. В свою очередь, каждый из этих индексов вычисляется на основе ряда показателей.

Показатели, определяющие *индекс заболеваемости*:

- некоторые инфекционные и паразитарные болезни,
- новые образования,
- болезни крови, кроветворных органов и отдельные нарушения, вовлекающие иммунный механизм,
- болезни эндокринной системы, расстройства питания и нарушения обмена веществ,
- болезни нервной системы,
- болезни глаза и его придаточного аппарата,
- болезни уха и сосцевидного отростка,
- болезни системы кровообращения,
- болезни органов дыхания,
- болезни органов пищеварения,
- болезни кожи и подкожной клетчатки,
- болезни костно-мышечной системы и соединительной ткани,
- болезни мочеполовой системы,
- врожденные аномалии (пороки развития), деформации и хромосомные нарушения,
- травмы, отравления и некоторые другие последствия воздействия внешних причин.

Показатели, определяющие *индекс продолжительности жизни*:

- ожидаемая продолжительность жизни мужчин,
- ожидаемая продолжительность жизни женщин.

Используя эти данные, мы рассчитали все необходимые характеристики: нормализованные показатели, веса, подындексы и, наконец, композитный индекс здоровья населения.

Ввиду ограниченности объема публикации, не имеем возможности привести результаты всех проведенных расчетов. Поэтому ограничимся, с нашей точки зрения, только наиболее важными из них.

В приложении (табл. 1) приведены значения композитного индекса здоровья населения за 2005–2013 гг. для каждого из 79 рассматриваемых регионов.

Веса нормализованных показателей в индексе уровня заболеваемости населения:

- некоторые инфекционные и паразитарные болезни (0,06),

- новые образования (0,05),
- болезни крови, кроветворных органов и отдельные нарушения, вовлекающие иммунный механизм (0,11),
- болезни эндокринной системы, расстройства питания и нарушения обмена веществ (0,07),
- болезни нервной системы (0,08),
- болезни глаза и его придаточного аппарата (0,06),
- болезни уха и сосцевидного отростка (0,05),
- болезни системы кровообращения (0,06),
- болезни органов дыхания (0,04),
- болезни органов пищеварения (0,08),
- болезни кожи и подкожной клетчатки (0,05),
- болезни костно-мышечной системы и соединительной ткани (0,06),
- болезни мочеполовой системы (0,06),
- врожденные аномалии (пороки развития), деформации и хромосомные нарушения (0,12),
- травмы, отравления и некоторые другие последствия воздействия внешних причин (0,05).

Веса нормализованных показателей в индексе продолжительности жизни населения:

- ожидаемая продолжительность жизни мужчин (0,62),
- ожидаемая продолжительность жизни женщин (0,38).

Веса подындексов в композитном индексе здоровья населения:

- индекс заболеваемости населения (0,33),
- индекс продолжительности жизни (0,67).

Опять же в силу ограниченности объема публикации приводятся графики композитного индекса здоровья населения только для некоторой условно репрезентативной подвыборки регионов.

Для создания такой подвыборки по каждому из 79 регионов было проведено усреднение с 2005 по 2013 г. по композитному индексу здоровья населения, а потом регионы были разбиты на пять квантилей, представляющих собой группы регионов с высоким, средневысоким, средним, средненизким и низким уровнями здоровья населения. Из каждой полученной подгруппы регионов мы выбрали по одному представителю и построили график композитного индекса здоровья населения соответствующего региона (рис. 1).

Проанализировав графики всех 79 регионов, в частности и тех, что приведены на рис. 1, заключаем, что композитный индекс здоровья населения меняется достаточно плавно и имеет, как правило, возрастающую динамику. Впрочем, следует иметь в виду, что поведение композитного индекса всецело обусловлено поведением определяющих его показателей и в большей мере тех из них, чей вес в индексе больше. Это замечание следует непосредственно из формул (1)–(4).

В случае с композитным индексом здоровья населения его восходящая динамика объясняется восходящей динамикой подындкса продолжительности жизни населения, вес которого в композитном индексе здоровья населения составляет 0,67.

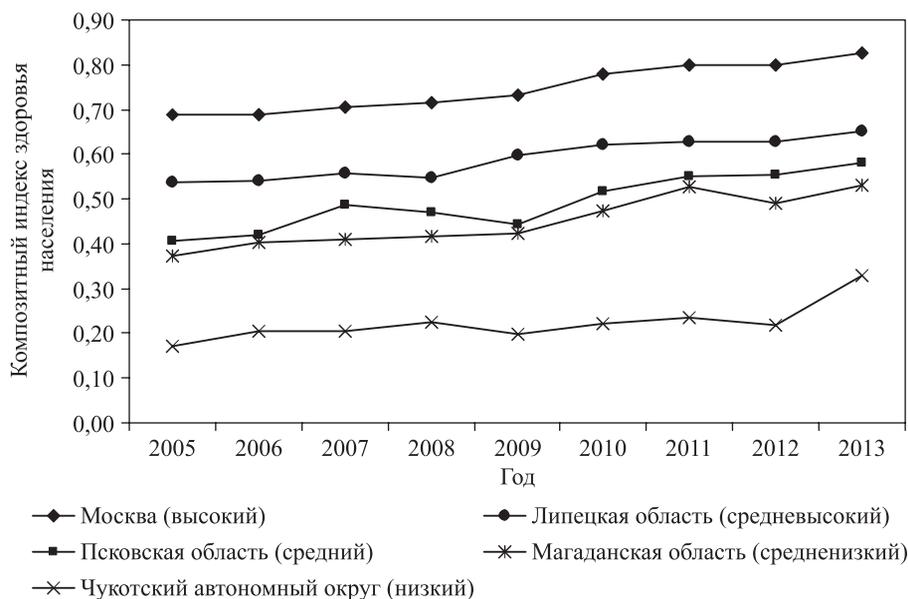


Рис. 1. Регионы с высоким, средневысоким, средним, средненизким и низким уровнем здоровья населения

3. СОПОСТАВЛЕНИЕ С МЕТОДОМ ИЗ РАБОТЫ [6]

В этом разделе проводится сопоставление предлагаемого в данной работе способа расчета композитных индексов и способа, описанного в статье [6]. Для краткости способ расчета композитных индексов из [6] будем называть «старым» способом, а способ, предлагаемый в данной статье, – «новым» способом. В частности, на синтетических данных будет показано, что «старый» способ, в отличие от «нового», не обладает свойством динамической сопоставимости.

Напомним, что в работе [6] композитный индекс $I_{i,t}$, относящийся к объекту i в момент времени t , вычислялся по формуле

$$I_{i,t} = w_t^{(1)} Z_{i,t}^{(1)} + \dots + w_t^{(k)} Z_{i,t}^{(k)}, \quad (1')$$

где веса $w_t^{(1)}, \dots, w_t^{(k)}$ находились как отношения

$$w_t^{(j)} := \frac{v_t^{(j)}}{v_t^{(1)} + \dots + v_t^{(k)}}, \quad j = 1, \dots, k, \quad (2')$$

величины $v_t^{(1)}, \dots, v_t^{(k)}$ рассчитывались по формулам

$$v_t^{(j)} = \frac{\text{med}_i |X_{i,t}^{(j)} - \text{med}_i X_{i,t}^{(j)}|}{\text{med}_i X_{i,t}^{(j)}}, \quad j = 1, \dots, k, \quad (3')$$

(здесь через $\text{med}_i X_{i,t}^{(j)}$ обозначена выборочная медиана массива данных $X_{i,t}^{(j)}$, где $i \in \{1, \dots, n\}$ меняется, а $t \in \{1, \dots, T\}$ фиксировано), а нормализованные показатели $Z^{(1)}, \dots, Z^{(k)}$ определялись как

$$Z_{i,t}^{(j)} := \begin{cases} \frac{X_{i,t}^{(j)} - \min_i X_{i,t}^{(j)}}{\max_i X_{i,t}^{(j)} - \min_i X_{i,t}^{(j)}}, & \text{если чем больше показатель } X^{(j)}, \text{ тем лучше,} \\ \frac{\max_i X_{i,t}^{(j)} - X_{i,t}^{(j)}}{\max_i X_{i,t}^{(j)} - \min_i X_{i,t}^{(j)}}, & \text{если чем меньше показатель } X^{(j)}, \text{ тем лучше.} \end{cases} \quad (4')$$

Сравнивая формулы (4) и (4'), видим, что оптимизация в формуле (4) производится как по i , так и по t , в то время как в формуле (4') – только по i (при фиксированном значении t). Такая модификация позволяет сделать сопоставимыми значения нормализованных показателей в разные моменты времени.

Теперь обратимся к формулам (3) и (3'). Как видно, в формуле (3) медиана находится по всему массиву данных $\{X_{i,t}^{(j)}\}_{i=1, \dots, n}^{j=1, \dots, T}$, в то время как в формуле (3') она находится по массиву $\{X_{i,t}^{(j)}\}_{i=1, \dots, n}$ (при каждом отдельном значении t). В результате при «новом» способе расчета весов $w^{(1)}, \dots, w^{(k)}$ с помощью формул (2) и (3) веса $w^{(1)}, \dots, w^{(k)}$ перестают быть переменными во времени. По-видимому, отказ от концепции переменных весов является необходимой «платой» для получения свойства динамической сопоставимости.

С целью сопоставления «старого» и «нового» способов расчета композитных индексов рассмотрим следующий искусственный пример.

Искусственный пример 1. Будем строить композитный индекс продолжительности жизни – I на основе двух показателей: ожидаемой продолжительности жизни мужчин (показатель $X^{(1)}$) и ожидаемой продолжительности жизни женщин (показатель $X^{(2)}$).

Предположим, что в 2005 г. ожидаемая продолжительность жизни как мужчин, так и женщин в нашем гипотетическом примере совпадает с той реальной ожидаемой продолжительностью жизни, которая была в регионах РФ в 2005 г. В рамках нашего примера предположим также, что с 2006 по 2013 г. ожидаемая продолжительность жизни мужчин с каждым годом росла на один год, а ожидаемая продолжительность жизни женщин за тот же период не менялась, оставаясь на уровне 2005 г. Таким образом, имеем следующую динамику показателей $X^{(1)}$ и $X^{(2)}$:

$$\begin{cases} X_{i,t}^{(1)} = X_{i,2005}^{(1)} + (t - 2005), \\ X_{i,t}^{(2)} = X_{i,2005}^{(2)}, \end{cases} \quad i = 1, \dots, 79, \quad t = 2006, \dots, 2013.$$

Используя в качестве данных искусственно созданные показатели $X^{(1)}$ и $X^{(2)}$, мы рассчитали композитный индекс продолжительности жизни «старым» и «новым» способами.

В результате, как и ожидалось, для каждого из регионов композитный индекс продолжительности жизни, рассчитанный «новым» способом, является не только неубывающей функцией, а более того – строго возрастающей (рис. 2).

Что касается «старого» способа расчета, то, например, для Чукотского автономного округа имеет место хотя и незначительное, но строгое снижение композитного индекса продолжительности жизни.

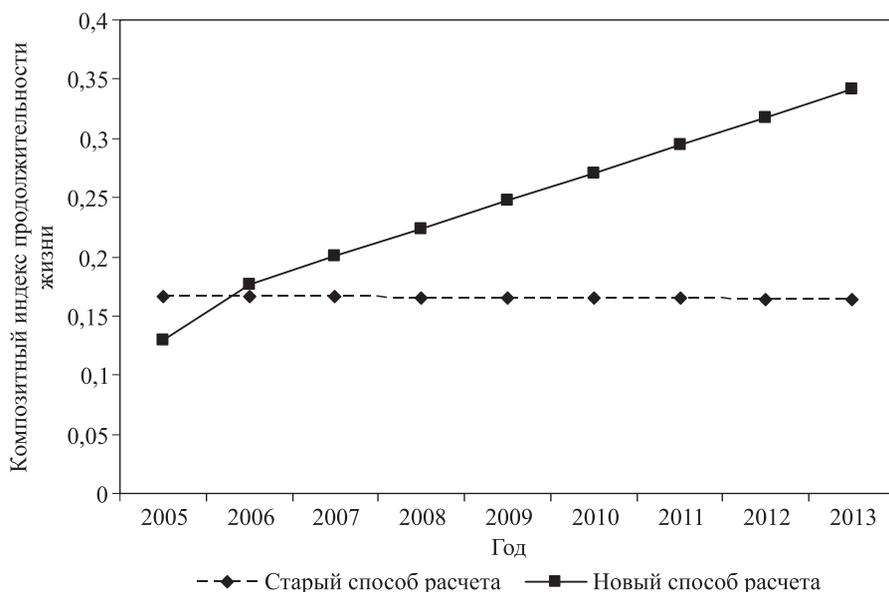


Рис. 2. Динамика композитного индекса продолжительности жизни в гипотетическом примере для Чукотского автономного округа

Итак, мы получили, что для i , соответствующего Чукотскому автономному округу, имеют место неравенства $X_{i,2005}^{(1)} \leq X_{i,2013}^{(1)}$ и $X_{i,2005}^{(2)} \leq X_{i,2013}^{(2)}$, но при этом $I_{i,2005} > I_{i,2013}$. Стало быть, «старый» способ расчета композитных индексов не обладает свойством динамической сопоставимости (определение см. раздел 1).

Теперь проанализируем описанную ситуацию с позиций интуиции и здравого смысла. Совершенно ясно, что в рассматриваемом гипотетическом примере с каждым годом продолжительность жизни населения (в целом) регионов растет, поэтому логично требовать от любого разумного способа расчета композитного индекса именно такого поведения – роста композитного индекса продолжительности жизни для каждого из регионов.

Как видно из рис. 2, композитный индекс продолжительности жизни, рассчитанный «старым» способом, ведет себя «не желаемым образом» – вместо того, чтобы расти, он снижается (хотя и незначительно).

Таким образом, можно констатировать, что композитный индекс продолжительности жизни, рассчитанный «старым» способом в динамике ведет себя неудовлетворительно, так как не обладает свойством динамической сопоставимости.

4. СРАВНЕНИЕ С МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

Метод главных компонент, как и рассмотренные выше методы, основан на представлении об изменчивости показателя как о мере его информативности. Первой главной компонентой показателей $Z^{(1)}, \dots, Z^{(k)}$ называется их линейная комбинация $I = v^{(1)}Z^{(1)} + \dots + v^{(k)}Z^{(k)}$ с наибольшей дисперсией,

коэффициенты в которой определяются как решение следующей экстремальной задачи:

$$\begin{cases} D(v^{(1)}Z^{(1)} + \dots + v^{(k)}Z^{(k)}) \rightarrow \max_{v^{(1)}, \dots, v^{(k)} \in \mathbf{R}}, \\ (v^{(1)})^2 + \dots + (v^{(k)})^2 = 1. \end{cases} \quad (5)$$

Ограничение на коэффициенты в формуле (5) необходимо, поскольку неограниченное увеличение любого из них позволяет достичь сколь угодно большой дисперсии всей линейной комбинации.

В качестве итогового композитного индекса используется значение первой главной компоненты. Другие главные компоненты (помимо первой) определяются иным образом [3] и используются в случаях, когда поведение первичных показателей допускает описание несколькими индексами (в настоящей статье такие ситуации не рассматриваются).

В формуле (5) показатели $Z^{(1)}, \dots, Z^{(k)}$ предполагаются уже нормализованными. Для нормализации, как правило, используется либо «стандартизация» (вычитание среднего и деление на стандартное отклонение), либо сведение к шкале с заданными наименьшим и наибольшим значением аналогично формуле (4). Стандартизация реализуется по умолчанию во многих статистических пакетах, сведение к фиксированной шкале используется в методологии построения индексов, предложенной С.А. Айвзяном [1]. В настоящей работе используется приведение к шкале с заданными наибольшим и наименьшим значениями, так как это способствует большей интерпретируемости нормализованных показателей. В отсутствие условия нормировки на показатели $Z^{(1)}, \dots, Z^{(k)}$ коэффициенты $v^{(1)}, \dots, v^{(k)}$ начинают зависеть от единиц измерения показателей $Z^{(1)}, \dots, Z^{(k)}$, что крайне нежелательно.

Отметим также, что индекс, рассчитанный методом главных компонент, может выходить за пределы от 0 до 1. С целью унификации, значения индекса можно привести к той же шкале от 0 до 1, что и участвующие в нем нормализованные показатели $Z^{(1)}, \dots, Z^{(k)}$, по формуле (4).

Основное идейное отличие метода главных компонент от метода, описанного в разделе 2, заключается в том, что при максимизации дисперсии линейной комбинации в задаче (5) в расчет принимается не только степень разброса показателей, но также и взаимосвязи между ними – ковариации.

Интуитивно достаточно ясно, что при наличии отрицательных ковариаций между показателями отдельные нормализованные показатели могут входить в главную компоненту с отрицательными коэффициентами. В результате этого возникает эффект, противоречащий здравому смыслу: при увеличении показателей, нормируемых по принципу «чем больше, тем лучше», происходит не увеличение, а, наоборот, уменьшение значения композитного индекса, в результате чего нарушается свойство динамической сопоставимости.

Следующий пример подтверждает высказанные выше интуитивные соображения. В данном примере показано, что метод главных компонент не обладает ни свойством неотрицательности весов (коэффициентов), ни свойством динамической сопоставимости.

Искусственный пример 2. Предположим, что состояние единственного исследуемого объекта описывается двумя показателями $X^{(1)}$ и $X^{(2)}$, динамика которых приведена в табл. 2 (см. приложение) и графически изображена на рис. 3. Чтобы при нормализации показателей $X^{(1)}$ и $X^{(2)}$ нормализованные показатели $Z^{(1)}$ и $Z^{(2)}$ совпадали с исходными, мы изначально взяли показатели $X^{(1)}$ и $X^{(2)}$, удовлетворяющие условиям $\min_t X_t^{(1)} = \min_t X_t^{(2)} = 0$ и $\max_t X_t^{(1)} = \max_t X_t^{(2)} = 1$. При этом сами значения показателей $X^{(1)}$ и $X^{(2)}$ подобраны с тем, чтобы ряд $X^{(2)}$ имел восходящий тренд, ряд $X^{(1)}$ – нисходящий (за исключением промежутка времени $t \in [4, 5]$, когда показатель $X^{(1)}$ рос) и при этом между показателями $X^{(1)}$ и $X^{(2)}$ наблюдалась отрицательная корреляция¹. Отметим, что ключевым моментом приведенного примера является то, что на промежутке времени $t \in [4, 5]$ оба ряда имеют одну и ту же положительную динамику.

Прерывистая линия на рис. 3 отражает значения индекса, построенного методом главных компонент (МГК) и приведенного к единой шкале с исходными показателями. Из-за отрицательной корреляции между показателями $X^{(1)}$ и $X^{(2)}$ вес величины $X^{(1)}$ в первой главной компоненте отрицательный (коэффициент $v^{(1)} = -0,68$), в результате чего возникает противоречивая ситуация: оба показателя на промежутке времени $t \in [4, 5]$ растут, но при этом композитный индекс, построенный по методу главных компонент, – падает. Таким образом, свойство динамической сопоставимости нарушается.

Наоборот, для индекса, построенного «новым» способом, описанным в разделе 2, этот же период времени ($t \in [4, 5]$) является периодом наибольшего роста, так же как и для показателей $X^{(1)}$ и $X^{(2)}$ (Приложение, табл. 2).

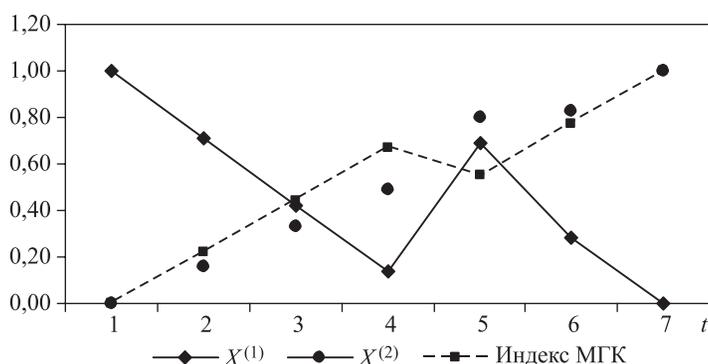


Рис. 3. Динамика индекса, построенного методом главных компонент, и его составляющих для искусственного примера 2

¹ Отметим, что отрицательная корреляция между показателями, входящими в индекс, не является чем-то надуманным или экзотическим. Так, например, если обратиться к данным о здоровье в регионах России, рассмотренным в разделе 2, оказывается, что показатель болезней крови, кровяной системы и отдельных заболеваний, вовлекающих иммунный механизм, отрицательно коррелирует с рядом других составляющих индекса заболеваемости – наиболее тесная отрицательная связь обнаруживается с показателем новообразований (коэффициент корреляции равен $-0,31$).

5. СОПОСТАВЛЕНИЕ РЕЗУЛЬТАТОВ И ВЫВОДЫ

В статье предложен новый способ построения динамических композитных индексов, обладающих двумя важнейшими свойствами, – свойством неотрицательности весов и свойством динамической сопоставимости.

Апробирование данной методологии проведено на примере построения композитного индекса здоровья населения. Проведено сопоставление с существующими методами построения композитных индексов – с методом главных компонент и методом, предложенным в работе [6].

Установлено, что метод главных компонент не обладает ни свойством неотрицательности весов, ни свойством динамической сопоставимости, а метод, предложенный ранее первым из авторов, не обладает только свойством динамической сопоставимости.

В рамках предложенной в данной работе методики расчета индексов нам удалось получить композитный индекс, обладающий желаемым свойством динамической сопоставимости. Таким образом, показано, что предлагаемый вниманию метод построения динамических композитных индексов обладает лучшими свойствами по сравнению с предложенными ранее методами.

Литература

1. Айвазян С.А. К методологии измерения синтетических категорий качества жизни населения // Экономика и математические методы. Т. 39. 2003. № 2.
2. Айвазян С.А. Россия в межстрановом анализе синтетических категорий качества жизни населения: анализ российской траектории на стыке XX–XXI вв. (1995–2004) // Мир России. 2005. № 1.
3. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статика. Т. 1 / 2-е изд. М.: ЮНИТИ-ДАНА, 2001.
4. Айвазян С.А., Степанов В.С., Козлова М.И. Измерение синтетических категорий качества жизни региона и выявление ключевых направлений совершенствования социально-экономической политики (на примере Самарской области и ее муниципальных образований) // Прикладная эконометрика. 2006. № 2.
5. Борзых Д.А. Динамическое рейтингование здоровья населения регионов РФ по данным Росстат за 2005–2012 гг. // Учет и статистика. 2015. № 3.
6. Борзых Д.А. Количественный анализ динамики уровня здоровья населения РФ // Вестник НГУЭУ. 2016. № 1.
7. Митяков Е.С., Корнилов Д.А. К вопросу о выборе весов при нахождении интегральных показателей экономической динамики // Экономика, инновации и менеджмент. 2011. № 2.
8. Пупия Л.К. Потребности и возможности измерения экономики знаний // Инновации. 2006. № 10 (97).
9. Регионы России. Социально-экономические показатели. 2013: стат. сб. / Росстат. М., 2013.
10. Dalton-Greyling T., Tregenna F. Construction and analysis of composite quality of life index for a region of South Africa. ERSА (Economic Research Southern Africa) Working Paper 481, 2014.
11. Development Report Card For The States CFED. [Электронный ресурс]. URL: http://cfed.org/assets/pdfs/2007_DRC.pdf
12. Investopedia. [Электронный ресурс]. URL: <http://www.investopedia.com/terms/c/compositeindex.asp#ixzz4OZLBdyHR>

13. *Козлова М. А.* Тенденции развития и пространственной организации региональной социально-культурной инфраструктуры: дис. ... канд. экон. наук. Екатеринбург, 2014. URL: <http://science.usue.ru/download/kozlova.pdf>
14. *Николаев И.А., Марченко Т.Е., Титова М.В.* Индекс антикризисной эффективности. Аналитический доклад ФБК. URL: http://www.fbk.ru/upload/images/effectiveness_index.pdf
15. Рейтинг качества жизни в регионах РФ. [Электронный ресурс]. URL: http://vid1.rian.ru/ig/ratings/life_2013.pdf.

Bibliography

1. *Ajvazjan S.A.* K metodologii izmerenija sinteticheskikh kategorij kachestva zhizni naselenija // *Jekonomika i matematicheskie metody*. T. 39. 2003. № 2.
2. *Ajvazjan S.A.* Rossija v mezhstranovom analize sinteticheskikh kategorij kachestva zhizni naselenija: analiz rossijskoj traektorii na styke XX–XXI vv. (1995–2004) // *Mir Rossii*. 2005. № 1.
3. *Ajvazjan S.A., Mhitarjan V.S.* Teorija verojatnostej i prikladnaja statika. T. 1 / 2-e izd. M.: JuNITI-DANA, 2001.
4. *Ajvazjan S.A., Stepanov V.S., Kozlova M.I.* Izmerenie sinteticheskikh kategorij kachestva zhizni regiona i vyjavlenie ključevyh napravlenij sovershenstvovanija social'no-jekonomicheskoj politiki (na primere Samarskoj oblasti i ejo municipal'nyh obrazovanij) // *Prikladnaja jekonometrika*. 2006. № 2.
5. *Borzyh D.A.* Dinamicheskoe rejtingovanie zdorov'ja naselenija regionov RF po dannym Rosstat za 2005–2012 gg. // *Uchet i statistika*. 2015. № 3.
6. *Borzyh D.A.* Kolichestvennyj analiz dinamiki urovnja zdorov'ja naselenija RF // *Vestnik NGUJeU*. 2016. № 1.
7. *Mitjakov E.S., Kornilov D.A.* K voprosu o vybore vesov pri nahozhdenii integral'nyh pokazatelej jekonomicheskoj dinamiki // *Jekonomika, innovacii i menedzhment*. 2011. № 2.
8. *Pipija L.K.* Potrebnosti i vozmožnosti izmerenija jekonomiki znanij // *Innovacii*. 2006. № 10 (97).
9. Regiony Rossii. Social'no-jekonomicheskie pokazateli. 2013: stat. sb. / Rosstat. M., 2013.
10. *Dalton-Greyling T., Tregenna F.* Construction and analysis of composite quality of life index for a region of South Africa. ERSA (Economic Research Southern Africa) Working Paper 481, 2014.
11. Development Report Card For The States CFED. [Jelektronnyj resurs]. URL: http://cfed.org/assets/pdfs/2007_DRC.pdf
12. Investopedia. [Jelektronnyj resurs]. URL: <http://www.investopedia.com/terms/c/compositeindex.asp#ixzz4OZLBdyHR>
13. *Kozlova M.A.* Tendencii razvitija i prostranstvennoj organizacii regional'noj social'no-kul'turnoj infrastruktury: dis. ... kand. jekon. nauk. Ekaterinburg, 2014. URL: <http://science.usue.ru/download/kozlova.pdf>
14. *Nikolaev I.A., Marchenko T.E., Titova M.V.* Indeks antikrizisnoj jeffektivnosti. Analiticheskij doklad FBK. URL: http://www.fbk.ru/upload/images/effectiveness_index.pdf
15. Rejting kachestva zhizni v regionah RF. [Jelektronnyj resurs]. URL: http://vid1.rian.ru/ig/ratings/life_2013.pdf

ПРИЛОЖЕНИЕ

Таблица 1

Динамика композитного индекса здоровья населения регионов РФ за 2005–2013 гг.
(начало)

Название региона	2005 г.	2006 г.	2007 г.	2008 г.	2009 г.	2010 г.	2011 г.	2012 г.	2013 г.	Сред- нее
1	2	3	4	5	6	7	8	9	10	11
Москва	0,69	0,69	0,70	0,71	0,73	0,78	0,80	0,80	0,83	0,75
Краснодарский край	0,58	0,77	0,82	0,83	0,82	0,78	0,66	0,67	0,67	0,73
Кабардино-Балкарская Республика	0,67	0,64	0,67	0,68	0,70	0,75	0,78	0,81	0,80	0,72
Астраханская область	0,48	0,71	0,74	0,78	0,77	0,62	0,59	0,61	0,64	0,66
Республика Ингушетия	0,70	0,62	0,63	0,64	0,67	0,56	0,62	0,69	0,71	0,65
Воронежская область	0,58	0,59	0,60	0,61	0,64	0,66	0,68	0,69	0,70	0,64
Республика Дагестан	0,59	0,63	0,65	0,64	0,66	0,60	0,63	0,65	0,66	0,63
Ростовская область	0,50	0,67	0,70	0,69	0,69	0,60	0,59	0,60	0,62	0,63
Республика Северная Осетия – Алания	0,62	0,54	0,59	0,59	0,61	0,65	0,68	0,68	0,70	0,63
Республика Калмыкия	0,56	0,58	0,60	0,60	0,60	0,71	0,67	0,66	0,67	0,63
Ставропольский край	0,64	0,50	0,52	0,54	0,56	0,70	0,71	0,73	0,72	0,62
Карачаево-Черкесская Республика	0,66	0,47	0,51	0,54	0,53	0,68	0,73	0,71	0,70	0,62
Республика Адыгея	0,56	0,58	0,60	0,59	0,63	0,61	0,62	0,64	0,66	0,61
Санкт-Петербург	0,55	0,55	0,57	0,57	0,59	0,63	0,65	0,67	0,70	0,61
Московская область	0,57	0,57	0,57	0,58	0,60	0,63	0,65	0,64	0,66	0,61
Волгоградская область	0,54	0,55	0,56	0,59	0,59	0,61	0,64	0,65	0,66	0,60
Липецкая область	0,54	0,54	0,56	0,55	0,60	0,62	0,63	0,63	0,65	0,59
Тамбовская область	0,50	0,52	0,55	0,56	0,56	0,61	0,64	0,66	0,67	0,59
Курская область	0,50	0,52	0,56	0,57	0,59	0,61	0,62	0,63	0,66	0,59
Ленинградская область	0,48	0,49	0,53	0,54	0,58	0,62	0,66	0,67	0,69	0,58
Республика Татарстан	0,52	0,52	0,54	0,57	0,59	0,59	0,60	0,62	0,63	0,58
Белгородская область	0,51	0,52	0,55	0,57	0,59	0,60	0,60	0,60	0,61	0,57
Республика Мордовия	0,51	0,52	0,54	0,55	0,56	0,58	0,61	0,63	0,61	0,57
Костромская область	0,45	0,50	0,54	0,53	0,55	0,57	0,59	0,63	0,64	0,56
Свердловская область	0,46	0,51	0,53	0,55	0,57	0,58	0,59	0,59	0,59	0,55
Новосибирская область	0,49	0,45	0,45	0,56	0,58	0,60	0,60	0,61	0,62	0,55
Кировская область	0,46	0,50	0,53	0,53	0,55	0,55	0,60	0,61	0,62	0,55
Пензенская область	0,43	0,48	0,50	0,52	0,55	0,58	0,60	0,62	0,65	0,55
Саратовская область	0,48	0,52	0,53	0,54	0,56	0,56	0,57	0,58	0,59	0,55
Тюменская область	0,46	0,49	0,51	0,52	0,55	0,58	0,60	0,60	0,61	0,55
Рязанская область	0,46	0,48	0,49	0,51	0,53	0,56	0,58	0,60	0,63	0,54
Калужская область	0,44	0,49	0,50	0,51	0,52	0,56	0,61	0,60	0,62	0,54
Томская область	0,44	0,39	0,44	0,56	0,58	0,59	0,56	0,61	0,60	0,53
Тульская область	0,41	0,44	0,47	0,48	0,51	0,54	0,57	0,58	0,59	0,51
Мурманская область	0,43	0,46	0,49	0,49	0,51	0,51	0,55	0,57	0,57	0,51
Удмуртская Республика	0,43	0,46	0,47	0,49	0,51	0,51	0,53	0,56	0,57	0,50
Калининградская область	0,37	0,43	0,47	0,48	0,51	0,53	0,57	0,57	0,58	0,50

Окончание табл. 1

1	2	3	4	5	6	7	8	9	10	11
Омская область	0,43	0,49	0,56	0,47	0,52	0,53	0,53	0,48	0,49	0,50
Челябинская область	0,45	0,47	0,49	0,48	0,51	0,51	0,51	0,52	0,54	0,50
Республика Бурятия	0,40	0,43	0,47	0,48	0,50	0,52	0,53	0,55	0,58	0,49
Вологодская область	0,41	0,44	0,48	0,48	0,49	0,50	0,52	0,55	0,56	0,49
Псковская область	0,40	0,42	0,49	0,47	0,44	0,52	0,55	0,55	0,58	0,49
Республика Башкортостан	0,46	0,45	0,44	0,46	0,51	0,51	0,52	0,53	0,55	0,49
Хабаровский край	0,39	0,39	0,47	0,49	0,52	0,52	0,52	0,55	0,57	0,49
Ярославская область	0,37	0,42	0,43	0,45	0,47	0,53	0,56	0,56	0,59	0,49
Приморский край	0,41	0,45	0,45	0,46	0,49	0,50	0,52	0,55	0,55	0,49
Орловская область	0,41	0,45	0,47	0,47	0,50	0,51	0,52	0,49	0,54	0,48
Брянская область	0,41	0,44	0,45	0,46	0,50	0,51	0,52	0,52	0,54	0,48
Курганская область	0,41	0,44	0,47	0,47	0,48	0,50	0,52	0,55	0,52	0,48
Камчатский край	0,40	0,45	0,47	0,49	0,48	0,47	0,48	0,52	0,55	0,48
Нижегородская область	0,42	0,44	0,47	0,47	0,48	0,49	0,50	0,51	0,52	0,48
Республика Марий Эл	0,41	0,43	0,45	0,45	0,47	0,49	0,51	0,52	0,54	0,47
Самарская область	0,45	0,45	0,48	0,45	0,47	0,48	0,48	0,50	0,48	0,47
Оренбургская область	0,38	0,41	0,44	0,45	0,48	0,49	0,51	0,53	0,54	0,47
Смоленская область	0,38	0,40	0,43	0,42	0,45	0,50	0,54	0,55	0,55	0,47
Забайкальский край	0,34	0,44	0,45	0,46	0,48	0,47	0,50	0,51	0,54	0,47
Ульяновская область	0,40	0,41	0,43	0,45	0,46	0,49	0,48	0,51	0,55	0,46
Красноярский край	0,39	0,35	0,39	0,47	0,49	0,50	0,50	0,51	0,53	0,46
Тверская область	0,39	0,40	0,43	0,45	0,46	0,47	0,49	0,49	0,51	0,45
Ивановская область	0,37	0,38	0,41	0,44	0,45	0,48	0,48	0,50	0,54	0,45
Магаданская область	0,37	0,40	0,41	0,42	0,42	0,47	0,53	0,49	0,53	0,45
Кемеровская область	0,36	0,52	0,54	0,39	0,42	0,44	0,44	0,45	0,44	0,44
Владимирская область	0,38	0,39	0,40	0,40	0,42	0,47	0,50	0,51	0,53	0,44
Республика Хакасия	0,30	0,40	0,43	0,44	0,47	0,49	0,48	0,47	0,48	0,44
Республика Саха (Якутия)	0,40	0,39	0,42	0,41	0,43	0,44	0,46	0,45	0,45	0,43
Республика Коми	0,33	0,39	0,42	0,43	0,44	0,44	0,45	0,46	0,50	0,43
Архангельская область	0,35	0,37	0,40	0,42	0,44	0,44	0,45	0,48	0,51	0,43
Амурская область	0,36	0,46	0,43	0,41	0,43	0,42	0,43	0,43	0,45	0,43
Новгородская область	0,34	0,35	0,39	0,39	0,41	0,43	0,46	0,49	0,46	0,41
Сахалинская область	0,30	0,36	0,40	0,41	0,41	0,41	0,44	0,45	0,52	0,41
Пермский край	0,31	0,34	0,37	0,39	0,41	0,44	0,45	0,47	0,49	0,41
Еврейская автономная область	0,33	0,39	0,40	0,42	0,43	0,42	0,40	0,42	0,44	0,41
Чувашская Республика	0,40	0,36	0,34	0,33	0,41	0,44	0,44	0,46	0,46	0,41
Алтайский край	0,34	0,37	0,39	0,39	0,42	0,43	0,44	0,43	0,44	0,41
Иркутская область	0,29	0,35	0,37	0,41	0,42	0,42	0,41	0,42	0,42	0,39
Республика Алтай	0,24	0,30	0,36	0,36	0,41	0,41	0,40	0,45	0,47	0,38
Республика Карелия	0,27	0,34	0,36	0,37	0,40	0,40	0,41	0,41	0,43	0,38
Республика Тыва	0,20	0,31	0,34	0,38	0,38	0,39	0,41	0,40	0,40	0,36
Чукотский автономный округ	0,17	0,20	0,21	0,23	0,20	0,22	0,23	0,22	0,33	0,22

Таблица 2

Данные для искусственного примера 2

t	$X^{(1)}$	$X^{(2)}$	«Новый» индекс, приведенный к шкале от 0 до 1 по формуле (4)	Индекс МГК, приведенный к шкале от 0 до 1 по формуле (4)
1	1,00	0,00	0,42	0,00
2	0,71	0,16	0,28	0,22
3	0,42	0,33	0,14	0,44
4	0,14	0,49	0,00	0,67
5	0,69	0,80	1,00	0,55
6	0,28	0,83	0,56	0,77
7	0,00	1,00	0,43	1,00