Экономический журнал ВШЭ. 2016. Т. 20. № 4. С. 711–730.
HSE Economic Journal, 2016, vol. 20, no 4, pp. 711–730.

Эмпирическое сравнение математических методов построения динамических рядов системы таблиц «затраты – выпуск»[1]

Кузнецов С.Ю., Пионтковский Д.И., Соколов Д.Д., Старчикова О.С.

Исследуется сравнительная эффективность методов проекции таблиц «затраты – выпуск» применительно к таблицам использования. Эмпирической базой исследования служили таблицы использования 28 стран мира за 1995–2010 гг. по данным международного проекта WIOD. Из базы данных этого проекта для нашего исследования выбирались только базовые таблицы, построенные на основе агрегирования национальной статистики, а не посредством проекции предшествующих таблиц. Проведено сравнительное исследование трех математических методов, показавших себя наиболее эффективными при построении проекций таблиц использования Испании и Нидерландов в эмпирическом исследовании Темуршоева, Вебба и Ямано (2011). В этих методах таблица использования строится на основе аналогичной таб-

[1] Статья подготовлена в рамках проекта ПФИ НИУ ВШЭ «Факторы долгосрочного развития российской экономики в межстрановом контексте» (2016 г.).

Мы выражаем свою благодарность Э.Ф. Баранову за ценные советы и замечания, также благодарны коллегам, которые прояснили нам ситуацию с базовыми годами составления таблиц использования в различных странах, в том числе Mehran Kafaï (Statistics Portal Grand Duchy of Luxembourg), Eva Schwarz (Federal Statistical Office Germany) и Ylva Petersson Strid (Statistics Sweden). Мы благодарим анонимного рецензента за замечания, которые помогли улучшить статью.

Кузнецов Сергей Юрьевич – студент IV курса бакалавриата Отделения прикладной математики и информатики факультета компьютерных наук, стажер-исследователь лаборатории проблем инфляции и экономического роста НИУ ВШЭ. E-mail: sergeysmith1995@ya.ru

Пионтковский Дмитрий Игоревич – д.ф-м.н., профессор департамента математики на факультете экономических наук НИУ ВШЭ. E-mail: dpiontkovski@hse.ru

Соколов Денис Дмитриевич – выпускник бакалавриата Отделения прикладной математики и информатики Факультета компьютерных наук и стажер-исследователь лаборатории проблем инфляции и экономического роста НИУ ВШЭ; студент 1 курса программы «Международный докторат по экономическому анализу (IDEA)», Автономный университет Барселоны (UAB). E-mail: denisdsokolov@gmail.com

Старчикова Ольга Сергеевна – выпускник бакалавриата Отделения прикладной математики и информатики Факультета компьютерных наук и стажер-исследователь лаборатории проблем инфляции и экономического роста НИУ ВШЭ; студент 1 курса программы «Международный докторат по экономическому анализу (IDEA)», Автономный университет Барселоны (UAB). E-mail: starchikovaos@gmail.com

Статья поступила: 30.09.2016/Статья принята: 15.11.2016.

лицы за один из предшествующих периодов, а также заданных сумм строк и столбцов. Наиболее эффективным из этих методов в нашем исследовании оказался модифицированный бипропорциональный метод – алгоритм GRAS: построенные с его помощью таблицы по ряду критериев оказались ближе к опубликованным статистическим таблицам, чем результаты применения метода INSD и метода Куроды, основанных на квадратичном программировании. Это позволяет рассматривать метод GRAS как приоритетный при экстраполяции таблиц использования в экономике России. В то же время в ряде случаев таблица не может быть сбалансирована методом GRAS из-за существенных изменений в ее структуре по сравнению с предшествующим периодом. В 80% таких случаев таблица была успешно сбалансирована квадратичными методами, из которых более эффективным показал себя метод Куроды.

Ключевые слова: проект WIOD; RAS; метод Куроды; таблицы «затраты – выпуск».

1. Введение

Таблицы «затраты – выпуск», предложенные в 1930-е годы В.В. Леонтьевым, используются во множестве прикладных исследований и являются частью системы национальных счетов (СНС) многих стран мира [Miller, Blair, 2009]. Первостепенное значение имеют таблицы использования и таблицы ресурсов, в которых отражается объем использования и выпуска отдельных категорий продуктов в различных отраслях экономики: на их основе строятся, в частности, симметричные таблицы «отрасль-отрасль» и «продукт-продукт». В настоящей работе мы будем иметь дело с таблицами использования.

Таблицы использования строятся либо в основных (базовых) ценах, либо в ценах покупателей: последние таблицы вычисляются как суммы таблиц в основных ценах и таблиц транспортных и торговых наценок, а также таблицы чистых налогов на продукты. Все эти таблицы использования разбиваются на четыре квадранта, устроенных следующим образом: в I квадранте отражено потребление каждой задействованной в экономике отраслью всех видов производимых продуктов (промежуточное потребление), во II квадранте представлен конечный спрос в разрезе его элементов, а в III квадранте – добавленная стоимость каждой отрасли. IV квадрант чаще всего не заполняется, но иногда в нем (частично) отражается перераспределение национального дохода. Таблицы «затраты – выпуск» строятся статистическими службами отдельных стран и могут объединяться в таблицы для групп стран (например, для Евросоюза). В терминологии руководства [Eurostat, 2008] квадранты I и II, с которыми мы имеем дело ниже, называются таблицами промежуточного потребления и конечного спроса соответственно (*table of intermediate use & table of final uses*). В таблицах использования в ценах покупателей для первого квадранта используется также наименование квадранта промежуточного спроса (*intermediate use quadrant*), а для второго квадранта – квадранта конечного спроса (*final use quadrant*).

Построение системы таблиц «затраты – выпуск» требует масштабного обследования предприятий всех значимых отраслей, которое обычно проводится национальными статистическими службами не чаще, чем раз в несколько лет (как правило – раз в пять лет). Поскольку прикладные задачи анализа и прогноза требуют временного ряда ежегодных таблиц, то на основе таблиц, построенных в результате упомянутого обследования (так называемых базовых таблиц), и ежегодных данных национальных счетов строятся таблицы за годы, когда такого обследования не проводилось. Эта процедура носит название проекции, или экстраполяции, таблиц[2].

Для этой процедуры – проекции таблиц «затраты – выпуск» – разработан ряд математических методов, среди которых наиболее известным является классический метод RAS. Эти методы постоянно совершенствуются, разрабатываются новые, и для выявления наиболее эффективных методов регулярно проводятся сравнительные эмпирические исследования. Одно из последних таких исследований на материале таблиц «затраты – выпуск» Испании и Нидерландов предпринято в работе [Temurshoev, Webb, Yamano, 2011]. Задача выбора наиболее эффективного метода проекции особенно актуальна сейчас в России, поскольку на основе недавно построенной Росстатом базовой системы таблиц за 2011 г. предстоит построить ряд аналогичных таблиц за последующие и предшествующие годы.

Цель настоящей статьи – более полное сравнительное эмпирическое исследование методов проекции таблиц использования, дополняющее результаты Темуршоева, Вебба и Ямано [Temurshoev, Webb, Yamano, 2011]. Отталкиваясь от работы [Temurshoev, Webb, Yamano, 2011], мы полагаем установленным, что наиболее эффективными из распространенных методов можно считать методы GRAS, INSD и варианты метода Куроды (см. о них раздел 2 ниже), и исследуем только эти методы. Поскольку, к тому же, мы концентрируемся на таблицах использования, такая концентрация позволила расширить эмпирическую базу исследования: мы использовали данные международного проекта WIOD, в рамках которого собраны таблицы использования по 40 странам мира за 1995–2012 гг.[3] Были предприняты усилия, чтобы отделить первичные (базовые) таблицы, составленные на основе агрегации статистических данных, от построенных путем проекции. На основе анализа имеющейся информации было выбрано 52 базовые таблицы использования из 29 стран, которые сравнивались с таблицами, построенными нами на основе различных математических методов проекции.

Результаты расчетов показали, что из трех рассматриваемых методов проекции наиболее эффективным показал себя модифицированный бипропорциональный – метод GRAS. Преимущество этого метода абсолютно при проекции первого квадранта таблицы использования (матрица промежуточного использования), на котором метод сводится к методу RAS, а также по результатам проекции первого и второго квадрантов при известных суммарных окаймляющих итогах. Однако в достаточно большом количестве случаев вообще таблица не может быть сбалансирована этим методом (например, если расположение нулевых элементов значительно изменилось между базовым годом и прогнозным). Наши вычислительные эксперименты показывают, что в 80% таких случаев квадратичные методы проекции Kuroda 1 и INSD дают удовлетворительный результат.

Работа имеет следующую структуру. В разделе 2 даны математические описания некоторых современных методов проекции таблиц «затраты – выпуск». В третьем разделе работы описаны исходные данные нашего исследования. Существенную часть здесь составляет информация о том, за какие годы включенные в WIOD страны выпускали базовые таблицы использования. На основе этой информации мы выбираем таблицы ис-

[2] В англоязычной литературе для описания этой процедуры кроме термина «projection» используются также термины «updating», «adjusting» и «balancing».

[3] См.: www.wiod.org

пользования по 28 странам из базы данных WIOD для вычислительных экспериментов. В четвертом разделе описаны несколько традиционных «метрик», позволяющих сравнивать построенные нашими методами таблицы и с реальными данными. И здесь, и в разделе 2 наше изложение следует, в основном, работе [Temurshoev, Webb, Yamano, 2011]. Результаты расчетов представлены в разделе 5. На первом шаге расчетов выделены ряды таблиц, для которых все рассматриваемые методы дают сбалансированный результат: для таких таблиц затем статистические итоги расчетов подведены отдельно. Заключительный раздел 6 посвящен краткому анализу результатов расчетов.

2. Некоторые методы проекции таблиц «затраты – выпуск»

Рассмотрим сначала класс так называемых методов, которые мы условно называем пропорциональными. Они представляют собой модификации классического метода RAS. Мы обозначаем через $A_0 = \left(a^0{}_{ij}\right)$ матрицу, соответствующую базовой таблице, через A – неизвестную матрицу, которую требуется построить. При этом предполагается, что известны суммы элементов по строкам и по столбцам матрицы $A = \left(a_{ij}\right)$: получившиеся векторы мы обозначаем соотвественно как u и v.

RAS. Метод RAS (классический алгоритм реализации бипропорционального метода, предложенный Р. Стоуном [Stone, 1961]) заключается в минимизации целевой функции

$$f\left(A, A_0\right) = \sum_{i,j} a_{ij} \ln \frac{a_{ij}}{e a_{ij}^0}, \tag{1}$$

где A_0 – базисная матрица; A – новая матрица при линейных ограничениях, которые фиксируют суммы всех строк и всех столбцов матрицы A:

$$\sum_j a_{ij} = u_i, \tag{2}$$

$$\sum_i a_{ij} = v_j. \tag{3}$$

Эту задачу минимизации решают итеративно по формулам $A_{n+1} = RA_n$ в случае четного n и $A_{n+1} = A_n S$ в случае нечетного n, где

$$R = \begin{pmatrix} \frac{u_1}{\sum_j a_{1j}^n} & 0 & \cdots & 0 \\ 0 & \frac{u_2}{\sum_j a_{2j}^n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{u_n}{\sum_j a_{nj}^n} \end{pmatrix} \text{ и } S = \begin{pmatrix} \frac{v_1}{\sum_i a_{i1}^n} & 0 & \cdots & 0 \\ 0 & \frac{v_2}{\sum_i a_{i2}^n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{v_n}{\sum_i a_{in}^n} \end{pmatrix}.$$

Как хорошо известно (см. [Miller, Blair, 2009]), можно гарантировать сходимость этого метода только для матрицы A_0 с неотрицательными значениями (за исключением некоторых разреженных A_0 (см. [Miller, Blair, 2009, 7.4.9]). В этом случае матрица A также получается неотрицательной.

GRAS. Классический метод RAS неприменим к матрицам, в которых некоторые элементы отрицательные. Для таких матриц аналогичную роль играет метод Generalized RAS (GRAS) [Günlük-Şenesen, Bates, 1988; Temurshoev, Miller, Bouwmeester, 2013]. Мы будем минимизировать ту же самую функцию (1) при тех же линейных ограничениях и при условии сохранения знаков элементов. Представим матрицу A в виде

$$A = P - N,$$

где P и N – матрицы тех же размеров, что и A, причем P содержит все положительные числа из матрицы A и нули на месте отрицательных, а N содержит числа, равные по модулю отрицательным числам из матрицы A, и нули на месте положительных.

Тогда элементы матрицы следующей итерации вычисляются по следующим формулам: в случае четного n $a_{ij}^{n+1} = r_i^{(n)} a_{ij}^{n}$ при $a_{ij}^{n} > 0$ и $a_{ij}^{n+1} = \left(r_i^{(n)}\right)^{-1} a_{ij}^{n}$ при $a_{ij}^{n} < 0$, а в случае нечетного n $a_{ij}^{n+1} = a_{ij}^{n} s_j^{(n)}$ при $a_{ij}^{n} > 0$ и $a_{ij}^{n+1} = a_{ij}^{n} \left(s_j^{(n)}\right)^{-1}$ при $a_{ij}^{n} < 0$.

Приведем также рекуррентные формулы для вычисления множителей $r_i^{(n)}$ и $s_j^{(n)}$.

Используя текущие значения множителей $r_i = r_i^{(n)}$ и $s_j = s_j^{(n)}$ и матриц $P = \left(p_{ij}\right)$, $N = \left(n_{ij}\right)$, введем обозначения $p_i\left(s\right) = \sum_j p_{ij} \cdot s_j$, $p_j\left(r\right) = \sum_i r_i \cdot p_{ij}$, $n_i\left(s\right) = \sum_j n_{ij} \cdot s_j^{-1}$ и $n_j\left(r\right) = \sum_j n_{ij} r_i^{-1}$. Тогда рекуррентные формулы для вычисления следующих значений $r_i = r_i^{(n+1)}$ и $s_j = s_j^{(n+1)}$ принимают вид

$$(4) \qquad r_i = \begin{cases} \dfrac{u_i + \sqrt{u_i^2 + 4p_i(s)n_i(s)}}{2p_i(s)}, \ p_i(s) > 0 \\ -\dfrac{n_i(s)}{u_i}, \ p_i(s) = 0 \end{cases}$$

и

$$s_j = \begin{cases} \dfrac{v_j + \sqrt{v_j^2 + 4p_j(r)n_j(r)}}{2p_j(r)}, \ p_j(r) > 0 \\ -\dfrac{n_j(r)}{u_j}, \ p_j(r) = 0. \end{cases}$$

Следующие два метода непосредственно не исследуются эмпирически в нашей статье и приводятся для сведения читателей. Дело в том, что они являются дальнейшими обобщениями метода GRAS и применительно к тем таблицам и той информации, которые используются в нашей статье, результаты их применения не отличимы от применения GRAS.

Sign preserved RAS. Данный метод [Lenzen, Gallego, Wood, 2009] – вариант метода GRAS, позволяющий использовать линейные ограничения любого вида, а не только ограничения по окаймляющим итогам. Он формулируется для задачи в обобщенном векторизованном виде, когда вместо линейных ограничений вида (2) и (3) записывается общая система линейных ограничений

$$(5) \qquad Ga = c,$$

где a – векторизация матрицы A (его элементы обозначаются как a_j); $G = \left(g_{ij}\right)$ – матрица коэффициентов линейных ограничений (фактически она состоит из нулей и единиц); c – вектор правых частей c_i линейных ограничений.

В такой постановке выражения (4) преобразуются в

$$(6) \qquad r^{(n)} = \frac{c_i + \sqrt{c_i^2 + 4 \sum\limits_{j, a_j^{(n-1)} g_{ij} > 0} g_{ij} a_j^{(n-1)} \sum\limits_{j, a_j^{(n-1)} g_{ij} < 0} g_{ij} a_j^{(n-1)}}}{2 \sum\limits_{j, a_j^{(n-1)} g_{ij} > 0} g_{ij} a_j^{(n-1)}}$$

и $a_j^{(n)} = a_j^{(n-1)} \left[r^{(n)} \right]^{\operatorname{sgn}\left(a_j^{(n-1)} g_{ij}\right)}$ при $i = n \bmod q$, где q — общее количество уравнений в системе (5).

В отличие от описанных методов RAS и GRAS, рассматриваемый вариант метода применим в случае произвольных линейных ограничений. Единственный его недостаток – то, что знаки всех элементов аппроксимируемой матрицы должны, по определению, сохраняться. К сожалению, на практике это свойство выполняется не всегда.

От этого недостатка свободна следующая модификация метода GRAS.

NRAS. Метод, предложенный в работе [Lenzen, Gallego, Wood, 2009], носит название **non sign preserved RAS**; будем обозначать его как NRAS. В этом методе алгоритм из предыдущего параграфа модифицируется таким образом, что если матрицу A невозможно сбалансировать, знак каких-то ее элементов может поменяться по сравнению с A_0. Замена знака происходит, если среди элементов, входящих в линейное ограничение, нет отрицательных, а правая часть отрицательна, или наоборот. Таким образом, выражения (6) преобразуются в указанных выше случаях в

$$r^{(n)} = \frac{\operatorname{sgn}\left(\sum\limits_i G_{ij} a_j^{(n-1)}\right) c_i}{\sum\limits_{j, a_j^{(n-1)} G_{ij} > 0} G_{ij} a_j^{(n-1)} - \sum\limits_{j, a_j^{(n-1)} G_{ij} < 0} G_{ij} a_j^{(n-1)}} \text{ и } a_j^{(n)} = a_j^{(n-1)} \left[r^{(n)} \right]^{\operatorname{sgn}\left(a_j^{(n-1)} g_{ij}\right)}.$$

Другой класс составляют так называемые квадратичные методы, в описании которых мы даем только формулы соответствующих целевых функций. Для приведенных ниже квадратичных функций нахождение условного минимума осуществлялось стандартными средствами пакета GUROBI с использованием MATLAB. Отметим, что зависимость найденного приближенно минимума от выбранного метода оптимизации пока не исследовалась.

INSD. Это часто используемый метод с целевой функцией (**improved normalized squared difference**) [Friedlander, 1961].

$$f = \sum_i \sum_j \frac{\left(a_{ij}^0 - a_{ij}\right)^2}{a_{ij}^0}. \tag{7}$$

Здесь суммирование ведется по всем i и j, для которых $a_{ij} \neq 0$. В отличае от метода наименьших квадратов (МНК), он лучше балансирует малые потоки, в то время как МНК их практически игнорирует.

Метод Куроды. Метод, предложенный в работе [Kuroda, 1988], заключается в минимизации целевой функции

$$f\left(A, A_0\right) = \frac{1}{2}\sum_i \sum_j \left(\frac{a_{ij}}{u_i} - \frac{a_{ij}^0}{u_i^0}\right)^2 w_{ij} + \frac{1}{2}\sum_i \sum_j \left(\frac{a_{ij}}{v_i} - \frac{a_{ij}^0}{v_i^0}\right)^2 v_{ij},$$

где w_{ij} и v_{ij} – веса, различные в различных вариантах метода (нумерация вариантов совпадает с их нумерацией из статьи [Temurshoev, Webb, Yamano, 2011]):

- Kuroda 1: $w_{ij} = \frac{\left(u_i^0\right)^2}{a_{ij}^2}, v_{ij} = \frac{\left(v_i^0\right)^2}{\left(a_{ij}^0\right)^2}$ [Kuroda, 1988, Case (2)];
- Kuroda 2: $w_{ij} = \frac{u_i^2}{2}, v_{ij} = \frac{v_i^2}{2}$ [Wilcoxen, 1989];
- Kuroda 3: $w_{ij} = v_{ij} = 1$.

При этом суммируются только слагаемые с ненулевыми знаменателями.

Вопрос о проекции таблиц «затраты – выпуск» актуален сейчас для России, поскольку на основе недавно построенной Росстатом базовой системы таблиц за 2011 г. предстоит построить перспективный и ретроспективный ряды аналогичных таблиц за последующие годы (до года, за который будет выпущена следующая базовая таблица) и за предшествующие годы. О некоторых проблемах такого построения см.: [Баранов, Ким, Старицына, 2011; Баранов, Ким, Пионтковский, Старицына, 2014].

Одно из последних сравнительных эмпирических исследований методов проекции таблиц «затраты – выпуск» предпринято в работе [Temurshoev, Webb, Yamano, 2011]. В ней результаты проекций по 10 методам сравнивались с реальными данными на таблицах различного типа по Испании и Нидерландам. По результатам этого исследования наиболее эффективным показал себя метод GRAS. Вместе с тем, именно на интересующих нас таблицах использования результаты не выглядят однозначными. При проекции таблиц использования не менее эффективными, чем методы RAS и GRAS, показали себя два квадратичных метода – метод INSD и метод Куроды. Так, на данных Нидерландов при построении таблицы использования в ценах покупателей 2000 г. по данным 1995 г. наиболее эффективным оказался метод Куроды, а при проекции таблицы 2000 г. на 2005 г. – метод RAS [Erumban et al., 2012, table 3]. Аналогично, при проекции таблиц использования Испании в основных ценах с 2000 г. на 2005 г. в первом квадранте лучше оказался метод Куроды, а во втором квадранте – метод RAS (см. два последних графика на Fig.1 и Table 4 в [Op. cit.]).

3. Источники данных для вычислительных экспериментов

При выборе таблиц из базы WIOD для вычислительных экспериментов нам приходится учитывать, что при построении таблиц использования на основе базовых таблиц национальные статистические службы также используют математические методы (см., например: [Eurostat, 2008, 8.6.1]). Тем самым, простое сравнение методов прогнозирования на таком эмпирическом материале может привести к искажениям: например, если официальная таблица за 2001 г. построена на основе таблицы за 2000 г. методом RAS, то, естественно, метод RAS покажется наиболее эффективным при проекции с 2000 г. на 2001 г., но никакой информации о его релевантности применительно к прогнозированию реального состояния экономики это не даст. Аналогичным образом, если за какие-то годы официальные таблицы «затраты – выпуск» отсутствовали, то недостающие таблицы строились в рамках проекта WIOD с помощью проекции [Erumban et al., 2012], в частности, посредством предложенного в работе [Timurshoev, Timmer, 2011] варианта метода RAS. Например, все таблицы по России построены этим методом на основе таблиц за 1995 г., поэтому, как и показали наши вычислительные эксперименты, на материалах WIOD метод RAS оказывается при проекции таблиц использования по России (с 1995 г. на 2000-е годы) более точным, чем на примере других стран.

Чтобы избежать подобных статистических артефактов, было принято решение использовать только те таблицы из базы WIOD, которые построены путем агрегации *базовых* национальных таблиц. Однако если при этом источники данных проекта WIOD исчерпывающе описаны авторами проекта [Erumban et al., 2012], то, к сожалению, достоверных сведений о том, какие именно из построенных национальными статистическими службами таблиц были базовыми, по некоторым странам найти не удалось. За исключением этих стран, а также тех, в которых, подобно России, за рассматриваемый период была построена только одна базовая таблица или даже ни одной базовой официальной таблицы, осталось 28 стран, данные по которым и составили эмпирическую базу нашего исследования.

Для большинства стран источниками информации о том, какие таблицы являются базовыми, являются сайты национальных статистических служб, где размещены и сами таблицы, и информация о них. Для некоторых стран соответствующей информации найти не удалось. Для Кипра нет таблиц за данный период. Для Мексики составлена лишь одна таблица за 2003 г., поэтому она не используется в данной работе, аналогичная ситуация наблюдается с Канадой, имеющей только одну базовую таблицу за 1997 г., и с

Россией с единственной базовой таблицей за 1995 г. С Нидерландами и Люксембургом ситуация противоположная: насколько нам известно (см. источники сведений в работе [Пионтковский, Соколов, Старчикова, 2015]), в этих странах базовые таблицы составляются ежегодно. Поскольку по другим странам мы рассматриваем, в основном, проекцию на срок около 5 лет, мы приняли решение ограничиться рассмотрением таблиц по Нидерландам и Люксембургу за 2000 и 2005 гг. Не удалось найти точных сведений по Австрии, Дании, Латвии, Литве, Словакии, Франции и Эстонии, поэтому таблицы по этим странам в наших расчетах не использовались.

Таблица 1.

Сведения о наличии базовых таблиц «затраты – выпуск» для избранной группы стран (соответствующие годы отмечены знаком плюс)

Страны	Годы																
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Австралия		+							+	+				+			
Бельгия	+					+											
Бразилия	+	+				+					+						
Болгария	+					+				+							
Великобритания				+						+				+			
Венгрия						+					+					+	
Германия						+					+						+
Греция											+					+	
Индия				+					+					+			
Индонезия	+					+					+						
Ирландия	+					+					+						
Испания	+					+					+						
Италия	+					+					+						
Канада			+														
Китай			+					+					+				
Кипр																	
Корея	+					+					+			+			
Люксембург						+					+					+	
Мальта						+	+										
Мексика									+								
Нидерланды						+					+					+	
Польша						+					+						
Португалия							+					+					
Россия	+																
Румыния						+						+					
Словения						+								+			
США			+					+					+				
Тайвань							+					+					
Турция				+				+									
Финляндия	+					+											
Чехия	+					+					+					+	
Швеция	+					+					+						
Япония	+					+					+						+

Более подробные указания на источники сведений по каждой стране см. в работе [Пионтковский, Соколов, Старчикова, 2015].

4. Общая методология проекции таблиц и сравнение результатов прогнозов

Мы будем рассматривать проекцию следующих возможных частей таблицы использования: I квадранта таблицы (промежуточное потребление), II квадранта (конечный спрос) или I и II квадрантов вместе.

Предположим, что у нас имеется матрица A_g^s – соответствующая часть таблицы использования, построенной для страны s за год g исходя из статистических данных. Также предположим, что известны (из национальных счетов или из других источников) окаймляющие итоги строк и столбцов аналогичной матрицы A_{g+t}^s (которая полностью состоит из неизвестных) за год $(g+t)$, где $t=1,2,\ldots$ Пусть размерность прогнозируемой матрицы равна $n\times m$ (в формате таблиц WIOD, которые мы рассматриваем в этой статье, $n=59$, а m равно 36, 6 или 42 в зависимости от того, рассматриваем ли мы первый квадрант, второй квадрант или же оба вместе). Теперь построим матричное представление системы описанных выше линейных ограничений на матрицу неизвестных A_{g+t}^s.

$$a_{g+t}^s = Gc_{g+t}^s, \tag{SYS}$$

где G – матрица размера $(n+m)\times(n\cdot m)$, состоящая из нулей и единиц; a_{g+t}^s – векторизация матрицы неизвестных A_{g+t}^s (размера $(n\cdot m)\times 1$); c_{g+t}^s – вектор размеров $(n+m)\times 1$, составленный из векторов окаймляющих итогов строк и столбцов матрицы A_{g+t}^s.

Искомая матрица должна удовлетворять системе (SYS). Но для поиска решения нам необходимо начальное приближение, которое было бы сходно с искомым по структуре. Таким приближением выступит векторизация известной нам матрицы A_g^s. Обозначим ее $\boldsymbol{a_g^s}$.

Наконец, мы передаем посчитанные значения G, $\boldsymbol{a_g^s}$ и c_{g+t}^s в процедуру прогнозирования и на выходе получаем прогноз $\tilde{A}_{g+t}^s$ матрицы использования.

Теперь перейдем к сравнению результатов прогнозирования двух произвольных методов M_1 и M_2.

Нужно сразу оговориться, что для корректного сравнения нам понадобится не только знание таблицы использования для начального года g, но и знание такой таблицы для целевого года $(g+t)$, поскольку мы будем в конкретном вычислительном эксперименте рассматривать как более эффективный тот метод прогнозирования, результат применения которого ближе, в некотором смысле, к истинной матрице A_{g+t}^s.

Упомянутая близость двух матриц может быть формализована различными способами. Здесь мы, следуя работе [Temurshoev, Webb, Yamano, 2011], рассматриваем пять следующих квазиметрик. Все эти квазиметрики хорошо известны, см., например: [Miller, Blair, 2009, 7.4.8]. Мы следуем обозначениям из [Temurshoev, Webb, Yamano, 2011, section 3]: в каждом из пяти случаев вычисляется своя «метрика» между матрицами (x_{ij}) и (x_{ij}^{true}). В дальнейшем мы вычисляем их для случая, когда $\left(x_{ij}^{true}\right) = A_{g+t}$ – истинная матрица прогнозируемого года, а $\left(x_{ij}\right) = A_{(g+t),m}$ – матрица прогноза на нужный нам год, где m – номер метода.

1) **Средняя абсолютная ошибка в процентах** (MAPE – Mean absolute percentage error):

$$MAPE = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\left|x_{ij} - x_{ij}^{true}\right|}{\left|x_{ij}^{true}\right|}\cdot 100.$$

2) **Взвешенная абсолютная процентная ошибка** (WAPE – Weighted absolute percentage):

$$WAPE = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{\left|x_{ij}^{true}\right|}{\sum_k\sum_l x_{kl}^{true}}\right)\frac{\left|x_{ij} - x_{ij}^{true}\right|}{\left|x_{ij}^{true}\right|}\cdot 100.$$

3) **Стандартизированная взвешенная абсолютная разница** (SWAD – Standardized weighted absolute difference):

$$SWAD = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left|x_{ij}^{true}\right|\cdot\left|x_{ij} - x_{ij}^{true}\right|}{\sum_k\sum_l \left(x_{kl}^{true}\right)^2}.$$

4) **Пси-статистика** (PsiStat):

$$\hat{\psi} = \frac{1}{\sum_k\sum_l x_{kl}^{true}}\sum_i\sum_j\left[\left|x_{ij}^{true}\right|\cdot\left|\ln\left(\frac{x_{ij}^{true}}{s_{ij}}\right)\right| + \left|x_{ij}\right|\cdot\left|\ln\left(\frac{x_{ij}}{s_{ij}}\right)\right|\right],$$

где $s_{ij} = \left(\left|x_{ij}^{true}\right| + \left|x_{ij}\right|\right)/2.$

5) **RSQ** – квадрат коэффициента корреляции между построенной матрицей и целевой матрицей.

В результате мы получаем для каждой пары матриц A_g и A_{g+t} по три значения каждой из метрик (по одной на метод). Теперь проранжируем эти три значения для всех метрик, кроме: для метрик 1)–4) – по убыванию (чем меньше – тем лучше), а для метрики 5) – по возрастанию (так как корреляция увеличивается при увеличении сходства). В итоге мы получаем три пятимерных вектора рангов R_m для каждого из трех методов ($m = 1,2,3$): i-я компонента вектора R_m равна рангу метода m среди трех методов относительно метрики i, где ранг 1 соответствует наиболее эффективному методу, а ранг 3 – наименее эффективному.

Наконец, вычислим теперь суммы элементов векторов R_1, R_2, R_3 и обозначим их через r_1, r_2 и r_3 соответственно. Сравнив эти три числа, ранжируем три метода и определяем наиболее эффективный из них применительно к проекции матрицы A_g на год $g+t$: более эффективному методу соответствует наименьшее значение r_i, а если r_i и r_j оказались равны, то считаем, что оба метода одинаково эффективны.

5. Результаты расчетов

На основе полученных данных о базовых годах были проведены следующие расчеты: для каждой страны, имеющей больше одного базового года в рассматриваемом промежутке, для таблиц в основных ценах и ценах покупателей были спрогнозированы первый, второй квадранты по отдельности и первый и второй квадранты вместе с базового года на следующий соседний базовый год для всех таких возможных пар. Например, у Германии базовые годы следующие: 2000, 2005, 2011, поэтому имеется две пары соседних базовых годов 2000 -> 2005 и 2005 -> 2011, так что было построено двенадцать прогнозов (по шесть на каждую пару: по три на пару для таблицы в основных ценах и по три на пару для таблицы в ценах покупателей).

В итоге описанные прогнозы были построены для 52 пар базовых годов. В результате каждого из прогнозов получалась таблица следующего вида (пример для данных по Нидерландам для объединенных I и II квадрантов в ценах покупателей для прогнозов 2000 -> 2005.

Таблица 2.

Значение метрик для различных методов проекции таблиц в ценах покупателей (объединенные I и II квадранты) для Нидерландов на 2005 г. от базового 2000 г.

Методы	Метрики													
	MAPE	R	SWAD	R	WAPE	R	PsiStat	R	RSQ	R	Inac	R	N0	R_all	CmR
GRAS	38,347	1	0,049	1	13,698	1	0,136	1	0,9944	1	6,1E+02	3	52	10	1
Kuroda1	38,999	3	0,059	3	14,110	3	0,140	3	0,9931	3	6,3E-09	1	52	2	3
INSD	38,613	2	0,050	2	13,871	2	0,138	2	0,9943	2	3,7E-08	2	52	6	2

Строки таблицы соответствуют методам прогнозирования, а по столбцам приведены значения метрик на полученных прогнозах и ранг (R) методов относительно этих значений. Кроме описанных выше метрик, в таблицы заносилась также величина невязки $(Inac)$, под которой подразумевался максимальный модуль разности между суммой строки или столбца построенной таблицы и аналогичной суммой в таблице из WIOD, а также $N0$ – количество ненулевых элементов в таблице из WIOD, которым соответствуют нулевые значения в построенной таблице. Значение в столбце R_all рассчитывается как сумма пяти слагаемых, по одному для каждой из первых пяти метрик, равное 2, 1 или 0 в зависимости от того, был соответствующий метод первым, вторым или последним по рангу относительно рассматриваемой метрики. Другими словами, значение R_all равно разности между максимально возможной суммой рангов по первым пяти метрикам (т.е. 15) и наблюдаемой суммой рангов по этим метрикам. Столбец CmR содержит кумулятивный ранг на основе R_all, который мы принимаем в качестве основного критерия при сравнении методов.

В таблицах 3 и 4 приведены количества прогнозов, в которых каждый из методов был первым, вторым и третьим по кумулятивному рангу. Эти таблицы приводятся отдельно по прогнозам в основных ценах и в ценах покупателей, а также отдельно по первому квадранту, второму квадранту и по объединенной матрице первого и второго квадрантов.

Таблица 3.

Количества прогнозов по кумулятивным рангам для таблиц в основных ценах

	I квадрант			II квадрант			I и II квадранты		
	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD
Кумулятивный ранг									
1	51	0	2	49	1	5	50	0	3
2	1	4	46	1	13	36	1	3	47
3	0	48	4	2	38	11	1	49	2

Таблица 4.

Количества прогнозов по кумулятивным рангам для таблиц в ценах покупателей

	I квадрант			II квадрант			I и II квадранты		
	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD
Кумулятивный ранг									
1	50	0	2	48	3	2	50	0	3
2	1	3	48	1	11	40	2	5	44
3	1	49	2	3	38	10	0	47	5

Некоторые таблицы не могут быть сбалансированы точно методом RAS или GRAS (например, в случае разреженных матриц см. [Miller, Blair, 2009, 7.4.9]). Иногда таблица вообще не может быть сбалансирована точно при соблюдении условия постоянства знаков, поскольку, например, иногда в СНС начинают учитывать новые категории, которым в таблице за предыдущий (базовый) год отвечают нулевые элементы. Препятствием к точной балансировке может также послужить изменение знака в какой-либо категории конечного спроса на противоположный. Поэтому может представлять интерес построение приближенной таблицы, в которой наблюдаются относительно небольшие невязки.

В ряде случаев (перечисленных в работе [Пионтковский, Соколов, Старчикова, 2015]) матрица прогноза первого квадранта не получилась достаточно сбалансированной, т.е. значение $Inac$ (максимальное абсолютное отклонение итога по столбцу или строке от экзогенно заданного) не менее 10 (т.е. не менее 10 млн долл.). Из-за наличия отрицательных элементов во втором квадранте и из-за того, что матрица второго квадранта обычно разрежена, при балансировке отдельно второго и одновременно первого и второго квадрантов такие случаи более распространены. В целом, оценки сравнительной эффективности методов, отраженные в табл. 3 и 4, не меняются, если ограничить статистическую базу только достаточно сбалансированными таблицами (с $Inac < 10$).

Итак, мы условно считаем сбалансированными прогнозные таблицы, в которых величина $Inac$ не больше 10. В зависимости от выбранного варианта таблицы (I квадрант, II квадрант или объединенные I и II квадранты) и от того, посчитаны они в основных ценах или в ценах покупателей, получаем списки сбалансированных прогнозов по разным странам и годам, указанные в табл. 5. Для перечисленных в таблице стран в указанные годы сбалансированы прогнозы, построенные по каждому из трех методов.

Таблица 5.

Список сбалансированных прогнозов по странам (отмечены знаком плюс)

Страна	Годы	Основные цены			Цены покупателя		
		I кв.	II кв.	I и II кв.	I кв.	II кв.	I и II кв.
Австралия	2003 -> 2004	+	+	+	+	+	+
Болгария	1995 -> 2000	+	+	+	+	+	+
Бразилия	1995 -> 1996	+	+	+	+	+	+
Бразилия	1996 -> 2000	+		+	+		+
Бразилия	2000 -> 2005		+			+	
Великобритания	2004 -> 2008	+			+		
Венгрия	2005 -> 2010	+	+	+	+		
Германия	2000 -> 2005	+			+		
Германия	2005 -> 2011	+		+	+		+
Греция	2005 -> 2011	+			+		
Индия	2003 -> 2008	+					

Окончание табл. 5.

Страна	Годы	Основные цены			Цены покупателя		
		I кв.	II кв.	I и II кв.	I кв.	II кв.	I и II кв.
Ирландия	1995 -> 2000	+	+	+	+	+	+
Италия	1995 -> 2000	+			+		
Италия	2000 -> 2005	+		+			
Корея	2005 -> 2008	+	+	+	+	+	+
Люксембург	2000 -> 2005	+			+		
Люксембург	2005 -> 2010	+		+	+		+
Португалия	2001 -> 2006	+	+	+	+	+	+
Словения	2000 -> 2008	+			+		
Япония	2000 -> 2005	+	+	+	+	+	+
Япония	2002 -> 2007	+			+		
Япония	2005 -> 2011	+		+	+		+

Целесообразно рассмотреть результаты расчетов отдельно по сбалансированным матрицам. Они приведены в табл. 6 и 7, где даны значения кумулятивного рейтинга CmR. Результаты расчетов по отдельным метрикам см. в работе [Пионтковский, Соколов, Старчикова, 2015].

Таблицы 6 и 7 показывают, что по кумулятивному рангу (т.е. по критерию максимизации суммы рангов) метод GRAS наиболее эффективен при проекции I квадранта и объединенных I и II квадрантов.

Таблица 6.

Количества прогнозов по кумулятивным рангам для сбалансированных таблиц в основных ценах

	I квадрант			II квадрант			I и II квадранты		
	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD
Кумулятивный ранг									
1	18	0	3	5	2	2	11	0	2
2	3	1	17	3	4	3	2	3	10
3	0	20	1	1	3	4	0	10	1

Таблица 7.

Количества прогнозов по кумулятивным рангам для сбалансированных таблиц в ценах покупателей

	I квадрант			II квадрант			I и II квадранты		
	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD	GRAS	Kuroda1	INSD
Кумулятивный ранг									
1	17	0	2	2	4	2	9	0	2
2	2	2	15	5	1	3	2	2	8
3	0	17	2	1	3	3	0	9	1

Для каждого из шести вариантов таблицы задачу выбора наиболее эффективного метода по нашим данным можно рассматривать как задачу многокритериальной оптимизации с пятью критериями: в качестве критериев выступают ранги рассматриваемых методов по результатам расчетов каждой из пяти рассматриваемых квазиметрик. Тогда кумулятивный ранг является одним из возможных агрегированных критериев. Анализ детальных результатов расчетов показывает, что при проекции таблицы промежуточного потребления (I квадрант) и объединенной таблицы промежуточного спроса и конечного потребления (I и II квадранты) как в базовых ценах, так и в ценах покупателей метод GRAS превосходит остальные по каждому из отдельных пяти критериев, т.е. является оптимальным по Парето (это видно из таблиц, приведенных в пп. 4.5 и 4.6 препринта [Пионтковский, Соколов, Старчикова, 2015]).

6. Выводы и заключительные замечания

Результаты расчетов показали, что из трех рассматриваемых методов проекции наиболее эффективным показал себя метод GRAS. Преимущество этого метода абсолютно при проекции первого квадранта таблицы использования (матрица промежуточного использования), на котором метод сводится к методу RAS, а также по результатам проекции первого и второго квадрантов при известных суммарных окаймляющих итогах. Это преимущество наблюдается при проекции таблиц как в ценах покупателей, так и в основных ценах. Тем не менее при проекции таблицы конечного спроса (II квадрант) все три метода дают сопоставимые результаты: хотя метод GRAS все же несколько более эффективен в среднем, здесь оснований для однозначной рекомендации не имеется.

Отметим, что наши результаты по Испании и Нидерландам аналогичны упомянутым во введении результатам из работы [Temurshoev, Webb, Yamano, 2011], полученным на основе более подробных национальных таблиц использования, см., например, табл. 3. Однако наше исследование показывает, что обнаруженные там случаи относительной эффективности квадратичных методов по сравнению с GRAS достаточно редки и могут считаться статистической флуктуацией.

Полученное очередное подтверждение эффективности пропорциональных методов при проекции таблиц «затраты – выпуск» служит аргументом в пользу того, чтобы при построении таблиц использования по России за 2000-е годы использовать именно такие методы.

При проекции таблицы промежуточного использования (I квадрант), а также всей таблицы использования (объединенных I и II квадрантов) как в основных ценах, так и в ценах покупателей на втором месте по сравнительной эффективности оказался метод INSD. Возможно, причина достаточно высокой сравнительной эффективности этого метода чисто математическая. Как отмечено в статье [Huang, Kobayashi, Tanji, 2008], целевая функция (5) метода INSD представляет собой аппроксимацию Тейлора второго порядка аналогичной функции (3) метода GRAS при $z_{ij} = 1$, поэтому естественно, что построенные этими методами матрицы близки (особенно при проекции на небольшие временные промежутки в 2–5 лет, когда изменения коэффициентов a_{ij} малы, т.е. значения z_{ij} близки к единице).

Сравнительно низкая эффективность метода Куроды оказалась для нас сюрпризом. Вероятно, это означает, что наблюдаемая раньше относительно высокая эффективность этого метода на таблицах использования основана на случайных флуктуациях. Поскольку результаты наших расчетов по Испании и Нидерландам близки к результатам из статьи [Temurshoev, Webb, Yamano, 2011], маловероятно, что тут играет роль какая-то особенность агрегации национальных таблиц в таблицы WIOD. Мы видим, тем не менее, что этот метод относительно эффективен при измерении результатов метрикой MAPE, что позволяет рекомендовать построенные на его основе прогнозы в задачах, в которых важны лишь относительные, а не абсолютные значения изменений отдельных показателей.

При этом обнаружилась еще одна область возможного применения квадратичных методов проекции Kuroda 1 и INSD. Мы видим, что в ряде случаев таблица не может быть сбалансирована соответствующим вариантом пропорционального метода из-за значительного изменения расположения в ней нулевых элементов между базовым годом и прогнозным (особенно в случае разреженных матриц): в нашем вычислительном эксперименте даже по первому квадранту (в котором метод RAS наиболее эффективен) это происходило в 31 случае из 52 для расчетов в основных ценах и в 33 случаях из 52 – в ценах покупателей (см. п. 4.3). В то же время в 25 случаях из этих 31 (соответственно в 27 из 33), т.е. в 80% случаев, методы Kuroda 1 и INSD давали достаточно удовлетворительный результат. Таким образом, в случае, когда метод RAS не дает сбалансированную матрицу, можно рекомендовать для проекции таблиц использования квадратичные методы: INSD и метод Куроды.

Наша методика сравнительного исследования может быть улучшена при дальнейших исследованиях для получения более точных результатов. Одна возможная причина неточности в наших расчетах – заметное количество несбалансированных таблиц. Более тонкая работа с данными должна исключить большинство таких случаев и уточнить результаты расчетов. Исключение импорта (известного из официальных статистических данных) из таблицы конечного спроса может несколько изменить значения метрик на прогнозах по II квадранту. Еще одна причина возможной неточности (хотя и не главная) – зависимость результата применения квадратичных методов от выбранного алгоритма условной квадратичной оптимизации: мы обнаружили, что результаты несколько различаются при применении стандартных средств такой оптимизации, реализованных в системе MATLAB, и при применении более эффективного на этой задаче барьерного метода, реализованного в библиотеке GUROBI.

* *
*

СПИСОК ЛИТЕРАТУРЫ

Баранов Э.Ф., Ким И.А., Старицына Е.А. Методологические вопросы реконструкции системы таблиц «затраты – выпуск» России за 2003 год и последующие годы в структуре ОКВЭД-ОКПД // Вопросы статистики. 2011. № 12. С. 3–8.

Баранов Э.Ф., Ким И.А., Пионтковский Д.И., Старицына Е.А. Вопросы построения таблиц «затраты – выпуск» России в международных классификаторах // Экономический журнал Высшей школы экономики. 2014. Т. 18. № 1. С. 7–42.

Пионтковский Д.И., Соколов Д.Д., Старчикова О.С. Сравнение математических методов прогнозирования таблиц «затраты – выпуск» на основе базы данных WIOD: Препринт WP2/2015/07. М.: Изд. дом ВШЭ, 2015.

Erumban A.E., Gouma R., Timmer M., de Vries G., de Vries K. Sources for National Supply and Use Table Input Files. World Input-Output Database (WIOD). April, 2012. (http://www.wiod.org/publications/source_docs/SUT_Input_Sources.pdf)

Eurostat Manual of Supply, Use and Input-Output Tables. Eurostat, 2008.

Friedlander D. A Technique for Estimating a Contingency Table, Given the Marginal Totals and Some Supplementary Data // Journal of the Royal Statistical Society. Series A (General). 1961. Vol. 124. P. 412–420.

Günlük-Şenesen G., Bates J.M. Some Experiments with Methods of Adjusting Unbalanced Data Matrices // Journal of the Royal Statistical Society. Series A (Statistics in Society). 1988. Vol. 151. № 3. P. 473–490.

Huang W., Kobayashi S., Tanji H. Updating an Input–output Matrix with Sign-preservation: Some Improved Objective Functions and Their Solutions // Economic Systems Research. 2008. Vol. 20. № 1. P. 111–123.

Kuroda M. A Method of Estimation for the Updating Transaction Matrix in the Input-Output Relationships / K. Uno, S. Shishido (eds.) // Statistical Data Bank Systems. Amsterdam: North Holland, 1988.

Lenzen M., Gallego B., Wood R. Matrix Balancing under Conflicting Information // Economic Systems Research. 2009. Vol. 21. № 1. P. 23–44.

Miller R.E., Blair P.D. Input-output Analysis: Foundations and Extensions. Cambridge University Press, 2009.

Stone R.A. Input-output Accounts and National Accounts // Organization for European Economic Cooperation. Paris, 1961.

Temurshoev U., Miller R.E., Bouwmeester M.C. A Note on the GRAS Method // Economic Systems Research. 2013. Vol. 25. № 3. P. 361–367.

Temurshoev U., Timmer M.P. Joint Estimation of Supply and Use Tables // Papers in Regional Science. 2011. Vol. 90. № 4. P. 863–882.

Temurshoev U., Webb C., Yamano N. Projection of Supply and Use Tables: Methods and their Empirical Assessment // Economic Systems Research. 2011. Vol. 23. № 1. P. 91–123.

Wilcoxen P.J. Kuroda's Method for Constructing Consistent Input-output Data Sets. Impact Research Centre, University of Melbourne, 1989.

An Empirical Comparison of the Mathematical Methods for the Time Series of the Supply and Use Tables Construction

Sergey Kuznetsov[1], **Dmitri Piontkovski**[2], **Denis Sokolov**[3], **Olga Starchikova**[4]

[1] National Research University Higher School of Economics,
20, Myasnitskaya str., Moscow, 101000, Russian Federation.
E-mail: sergeysmith1995@ya.ru

[2] National Research University Higher School of Economics,
20, Myasnitskaya str., Moscow, 101000, Russian Federation.
E-mail: dpiontkovski@hse.ru

[3] Universitat Autònoma de Barcelona,
Departament d'Economia i d'Història Econòmica – Office B3-112G,
Facultat d'Economia i Empresa, Edifici B, Barcelona, 08193 Bellaterra, Spain.
E-mail: denisdsokolov@gmail.com

[4] Universitat Autònoma de Barcelona,
Departament d'Economia i d'Història Econòmica – Office B3-112G,
Facultat d'Economia i Empresa, Edifici B, Barcelona, 08193 Bellaterra, Spain.
E-mail: starchikovaos@gmail.com

We investigate the relative effectiveness of the projection methods of Supply and Use tables in relation to Use tables. The empirical bases of the study are the Use tables of 28 countries for the period from 1995 to 2010 from WIOD project. We conduct a comparative study of three mathematical methods that have proven the most effective in constructing projection of Use tables for Spain and the Netherlands from the empirical study by Temurshoev, Webb, and Yamano (2011). In these methods, a Use table is constructed based on the benchmark table and the sums of the columns and the rows of the table under construction. The most effective of these methods is GRAS, a version of the classical RAS algorithm. The results of applying this method under the number of criteria are closer to the published tables than the results of the INSD method and Kuroda method, which are based on quadratic programming. We conclude that GRAS method is a priority in the extrapolation of Use tables for Russia. At the same time we have shown that in some cases the table cannot be balanced by GRAS method because of significant changes in the structure of the table. In 80% of these cases the tables were successfully balanced by the two quadratic methods. In these cases the Kuroda method is the most effective.

***Key words*:** WIOD project; RAS; Kuroda method; supply and use tables.

JEL Classification: D57, C53, C67.

* *
*

References

Baranov E.F., Kim I.A., Starytsina E.A. (2011) Metodologitcheskie voprosy rekonstruktsii sistemy tablits «zatraty – vypusk» Rossii za 2003 i posleduyshie gody v structure OKVJeD – OKPD [Constructing Retrospective Time Series of Russian Input-Output Accounts Based on the NACE/CPA Classifications]. *Voprosy Statistiki*, 12, pp. 3–8.

Baranov E.F., Kim I., Piontkovski D., Staritsyna E.A. (2014) Voprosy postroeniya tablits «zatraty – vypusk» Rossii v mezhdunarodnyh klassifilkatorah [Problems of Constructing Russian Input-Output Tables into the International Classifications]. *HSE Economic Journal*, 18, 1, pp. 7–42.

Piontkovski D.I., Sokolov D.D., Starchikova O.S. (2015) *Sravnenie matematicheskikh metodov prognozirovamiya tablic «zatraty – vypusk» na osnove bazy dannykh WIOD*. [A Comparison of the Mathematical Projection Methods of IO Accounts on the Database of the WIOD Project.] Working Paper WP2/2015/07. Moscow: Higher School of Economics Publ. House, Series WP2 «Quantitative Analysis of Russian Economy».

Erumban A.E., Gouma R., Timmer M., de Vries G., de Vries K. (2012) *Sources for National Supply and Use Table Input Files*. World Input-Output Database (WIOD). April. Available at: http://www.wiod.org/publications/source_docs/SUT_Input_Sources.pdf

Eurostat (2008) *Eurostat Manual of Supply, Use and Input-Output Tables.*

Friedlander D. (1961) A Technique for Estimating a Contingency Table, Given the Marginal Totals and Some Supplementary Data. *Journal of the Royal Statistical Society*. *Series A (General)*, 124, pp. 412–420.

Günlük-Şenesen G., Bates J.M. (1988) Some Experiments with Methods of Adjusting Unbalanced Data Matrices. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 151, 3, pp. 473–490.

Huang W., Kobayashi S., Tanji H. (2008) Updating an Input–output Matrix with Sign-preservation: Some Improved Objective Functions and Their Solutions. *Economic Systems Research*, 20, 1, pp. 111–123.

Kuroda M. (1988) A Method of Estimation for the Updating Transaction Matrix in the Input-Output Relationships (eds. K. Uno, S. Shishido). *Statistical Data Bank Systems*. Amsterdam: North Holland.

Lenzen M., Gallego B., Wood R. (2009) Matrix Balancing under Conflicting Information. *Economic Systems Research*, 21, 1, pp. 23–44.

Miller R.E., Blair P.D. (2009) *Input-output Analysis: Foundations and Extensions*. Cambridge University Press.

Stone R.A. (1961) Input-output Accounts and National Accounts. *Organization for European Economic Cooperation*. Paris.

Temurshoev U., Miller R.E., Bouwmeester M.C. (2013) A Note on the GRAS Method. *Economic Systems Research*, 25, 3, pp. 361–367.

Temurshoev U., Timmer M.P. (2011) Joint Estimation of Supply and Use Tables. *Papers in Regional Science*, 90, 4, pp. 863–882.

Temurshoev U., Webb C., Yamano N. (2011) Projection of Supply and Use Tables: Methods and their Empirical Assessment. *Economic Systems Research*, 23, 1, pp. 91–123.

Wilcoxen P.J. (1989) *Kuroda's Method for Constructing Consistent Input-output Data Sets*. Impact Research Centre, University of Melbourne.