Байесовский подход к построению таблиц «затраты-выпуск»

О.В. Луговой, А.В. Полбин, В.Ю. Поташников

С задачами обновления, балансировки, дезагрегации таблиц «затраты-выпуск» сталкиваются как исследователи в сфере экономики и статистики (например, для калибровки вычислимых моделей общего равновесия), так и статистические службы (при построении таблиц). В отличие от известных популярных методов (RAS, минимизации перекрестной энтропии и их аналогов) в предлагаемом подходе вместо точечных оценок коэффициентов прямых затрат оценивается их совместное вероятностное распределение. Для этого строится вероятностная модель совместного распределения ячеек, являющаяся функцией правдопо-добия новой наблюдаемой информации (например, выпуск, добавленная стоимость, промежуточный спрос), которая с помощью формулы Байеса объединяется с априорной информацией о ячейках (например, известных таблиц предыдущих лет).

Получаемая в итоге апостериорная совместная плотность вероятности оценивается методами сэмплирования Монте-Карло по схеме цепи Маркова. Характеристики апостериорного распределения определяются набором (выборкой) искомых таблиц из этого распределения. При этом каждая из полученных таблиц не противоречит имеющимся данным, ограничениям и не слишком далека от априорно заданной таблицы или любой другой информации о ячейках. В отличие от точечных оценок стохастические таблицы напрямую инкорпорируют информацию о неопределенности каждого оцененного коэффициента прямых затрат таблиц «затраты-выпуск», учитывая существующие между ними взаимосвязи. Предлагаемая методика может использоваться для экстраполяции, интерполяции, дезагрегации и балансировки таблиц «затраты-выпуск» и более широко - матриц социальных счетов. С целью апробации метода проводится экспериментальная оценка таблиц «затраты-выпуск» российской экономики за 2003 г. на основе таблиц 1998-2002 гг.

Экспериментальное применение метода Байеса на реальных данных продемонстрировало адекватность и вычислительную доступность предлагаемой методики.

Ключевые слова: балансировка, таблицы «затраты-выпуск», метод Байеса.
JEL: C11, C6, C8.

Введение

Задача построения таблиц «затраты-выпуск» широко используется как в прикладных исследованиях (например, для калибровки вычислимых моделей общего равновесия), так и статистическими службами для оценки таблиц между базовыми статистическими обследованиями предприятий. Для этих целей обычно используются технические методы построения матриц, которые на выходе дают сбалансированную оценку таблиц «затраты-выпуск» за рассматриваемый год, согласующуюся с доступными статистическими данными СНС. В довольно общей постановке задача актуализации соответствующих таблиц сводится к нахождению матрицы прямых затрат A, которая удовлетворяет следующим ограничениям:

$$Y = AX; \quad \sum_i a_{i,j} = \bar{a}_j, \; a_{i,j} \geq 0, \qquad (1)$$

где Y, X - известные векторы промежуточного спроса и выпуска соответственно; $a_{i,j}$ - коэффициенты пря-мых затрат, которые являются элементами матрицы A;

$\bar{a}_j$ - известная сумма по столбцам матрицы A, которая представляет собой долю промежуточного потребления в выпуске отрасли j.

В общем случае система уравнений (1) является неопределенной и имеет бесконечное число решений. При актуализации обычно предполагается, что известна матрица прямых затрат

Луговой Олег Валерьевич (olugovoy@gmail.com) - канд. экон. наук, вед. научн. сотрудник, Центр экономического моде-лирования энергетики и экологии Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (РАНХиГС).

Полбин Андрей Владимирович (apolbin@gmail.com) - ст. научн. сотрудник, Центр экономического моделирования энер-гетики и экологии Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (РАНХиГС); ст. научн. сотрудник, Институт экономической политики им. Е.Т. Гайдара (ИЭП).

Поташников Владимир Юрьевич (potashnikov.vu@gmail.com) - ст. научн. сотрудник, Центр экономического моделиро-вания энергетики и экологии Российской академии народного хозяйства и государственной службы при Президенте Рос-сийской Федерации (РАНХиГС).

A^0 некоторого предыдущего периода. Одним из наиболее популярных подходов построения таблиц «затраты-выпуск» по причине простоты имплементации является метод RAS [21], в котором проводится итеративная процедура балансировки неизвестной таблицы «затраты-выпуск», где для первого приближения используется матрица прямых затрат A^0.

Более сложные методы оптимизации сводятся к нахождению такой матрицы коэффициентов прямых затрат $\hat{A}$, которая минимизирует некоторую меру расстояния $\{\Psi (A, A^0)\}$ от известной матрицы A^0, то есть $\hat{A} = \mathrm{argmin}\{\Psi (A, A^0)\}$ при условии совместности оцениваемой матрицы с системой уравнений (1). Например, Алмон [9] минимизирует меру квадратов разностей между коэффициентами прямых затрат A и априорным распределением коэффициентов A^0 при ограничениях (1); Байрон [10] минимизирует меру взвешенных квадратов разностей; Голан и др. [13] для актуализации таблиц «затраты-выпуск» минимизируют меру перекрестной энтропии между матрицами (детальный обзор альтернативных подходов к построению таблиц «затраты-выпуск» см., например, в работе [22]).

Оговоренные выше методы ставят целью получение единственной оценки неизвестной матрицы A, наиболее близкой к искомой и удовлетворяющей новым данным. Поскольку таких матриц может быть множество, а различие методов определяется лишь мерой расстояния, то вопрос неопределенности оцениваемых пара-метров является не менее информативным, чем сама оценка, и может существенно дополнить любые точечные оценки. Полученные вероятнос-тные оценки могут иметь широкое прикладное значение.

Идея описания коэффициентов прямых затрат в виде некоторых случайных величин не нова. Воп-рос учета неопределенности при оценке коэффи-циентов матриц «затраты-выпуск», вероятностном описании коэффициентов матриц развивается по нескольким направлениям, в зависимости от целей и методологии. В качестве примеров можно вы-делить задачу учета ошибок измерения, проблему нечеткой классификации отраслей и продуктов, проблему агрегации данных по отдельным от-раслям и фирмам с негомогенной технологией и др. (см., например, [8, 11, 12, 18]). И данная идея о стохастичности коэффициентов прямых затрат может быть расширена на задачу построения и балансировки таблиц «затраты-выпуск».

В работах [10, 16, 19, 23, 24] в достаточно общей постановке задачи при балансировке данных национальных счетов минимизиро-валась мера взвешенных квадратов разностей $(p - p^0)\Sigma^{-1}(p - p^0)$ при соответствующих балан-совых ограничениях, где p - вектор оцениваемых параметров; p^0 - вектор априорных значений оцениваемых параметров, в качестве которого при построении таблиц «затраты-выпуск» мо-жет выступать матрица предыдущего периода (в векторизованном виде); Σ^{-1} - взвешивающая матрица, которая отражает меру неопределенности оцениваемых параметров. Нахождение минимума рассматриваемой меры взвешен-ных квадратов разностей будет эквивалентно оценке моды апостериорного распределения, используемого также в данной работе, если в качестве априорного распределения использовать многомерное нормальное распределение $p \sim N(p^0,\Sigma)$. В работах [15, 20] предлагалось следовать Байесовскому подходу при построении и балансировке данных национальных счетов, но целью авторов было нахождение точечной оценки неизвестных параметров, в отличие от распределений.

Предлагаемый в данной работе метод позволяет, наряду с точечными оценками (моды, среднего), оценивать распределения коэффициентов прямых затрат, а также множественные ковариации между одновременно оцениваемыми показателями. Метод основан на вероятностном подходе Байеса с использованием широкого круга априорной информа-ции, как количественной статистической, так и качественной, включая экспертные оценки, и отслеживать влияние этой информации на конечный результат. Предлагаемый метод реализован в кластерных компьютерных системах и апробирован.

Общие сведения о методе Байеса в статистическом анализе

В контексте Байесовской эконометрики предполагается, что перед началом эксперимента исследователь обладает некоторой информацией относительно значения оцениваемого параметра θ. Данная информация может быть представлена в виде вероятностного распределения $p(\theta)$, называемого *априорным распределением*. В результате учета статистических данных распределение параметра уточняется, переходя от априорного распределения к апостериорному, в соответствии с законом Байеса:

$$p(\theta \mid Y) = \frac{L(Y \mid \theta) \cdot p(\theta)}{\int L(Y \mid \theta) \cdot p(\theta) d\theta} \propto L(Y \mid \theta) \cdot p(\theta), \quad (2)$$

где $p(\theta \mid Y)$ - функция плотности апостериорного распределения параметров; $L(Y \mid \theta)$ - функция правдоподобия.

Апостериорное распределение содержит всю необходимую информацию об оцениваемых параметрах. На ее основе можно получить точечную оценку искомых величин, доверительное множество их значений и провести соответствующий анализ корреляции этих параметров. В качестве точечной оценки обычно выбирается такое значение параметров, которое минимизирует математическое ожидание некоторой функции потерь, где оператор ожидания берется по апостериорному распределению. Наиболее популярной является оценка апостериорного среднего, которая минимизирует функцию апос-териорного риска.

Несмотря на всю привлекательность Байесовского метода, в прошлом он не был таким популярным из-за численного интегрирования, необходимого для вычисления знаменателя в выражении (2). В некоторых случаях, когда априорное распределение является сопряженным с функцией правдоподобия, апостериорное распределение можно получить в аналитическом виде. Однако в общем случае решение не имеет аналитического вида.

С развитием вычислительных мощностей современных компьютеров и разработкой методов сэмплирования, таких, как методы Монте-Карло по схеме марковской цепи (MCMC), позволяю-щих получать выборку необходимой размерности из апостериорного распределения, значительно расширились возможности применения Байесовского подхода в статистике.

Формализация метода Байеса для построения таблиц «затраты-выпуск»

В контексте Байесовского метода построения таблиц «затраты-выпуск» функцию апостериорного распределения запишем в виде[1]:

$$z = vec(A); \quad (3)$$

$$p(z \mid data) \propto p(z) \cdot I_{\Psi}(z) \cdot L(Y \mid z). \quad (4)$$

В уравнении (3) *vec* обозначает оператор векторизации матрицы, то есть после применения данного оператора матрица переходит в вектор, в котором последовательно записаны столбцы исходной матрицы. В уравнении (4) $p(z \mid data)$ - апостериорное распределение вектора коэффициентов прямых затрат; $p(z)$ - некоторое априорное распределение. В прикладных задачах предлагаемого метода часть линейных ограничений может рассматриваться с ошибкой измерения, а часть - без. Например, возможна ситуация, когда вектор промежуточного спроса (или некоторые его компоненты) не известен с достоверностью, но для него существуют некоторые оценки. Соответственно, в качестве функции правдоподобия мы рассматриваем функцию $I_{\Psi}(z) \cdot L_{Y|z}(\ )$, где $I_{\Psi}(z)$ - индикаторная функция, которая равна еди-нице, если все линейные ограничения системы без ошибок измерения выполнены, и равна нулю в противном случае; $L_{Y|z}(\ )$ - функция правдоподобия для линейных ограничений с ошибкой измерения.

В качестве априорного распределения могут выступать, например, усеченное на отрезке [0, 1] нормальное распределение, бета-распределение и равномерное распределение на отрезке [0, 1]. Если имеется априорная информация, что некоторый коэффициент равен определенному значению, то данная информация будет вводиться в виде линейного ограничения и учитываться в индикаторной функции. Так, довольно естественными ограничениями могут быть равенства нулю тех коэффициентов, которые были нулевыми в предыдущие годы. Использование в уравнении (4) индикаторной функции в качестве компоненты функции правдоподобия может показаться необычным. Представим, что все линейные ограничения измерены с некоторой ошибкой измерения, которая имеет нормальное распределение. Тогда функция правдоподобия принимает стандартное Гауссовское распределение. Если теперь устре-мить дисперсии ошибок к нулю, то функция правдоподобия будет стремиться к функции Дирака. И, таким образом, апостериорное распределение будет прямо пропорционально произведению

[1] В работе [15] формулировалась аналогичная задача, но авторы не проводили оценку апостериорной плотности распределения коэффициентов прямых затрат, а оценивали только моду апостериорного распределения с помощью численных методов.

априорного распределения на индикатор того, что все ограничения выполнены.

Итак, в рамках формулы (4) нам известна функция апостериорного распределения вектора коэффициентов прямых затрат с точностью до константы, то есть в аналитическом виде функция плотности нам не известна. Мы будем оценивать функцию апостериорного распределения с помощью методов Монте-Карло по схеме марковской цепи, которые позволяют произвести генерацию выборки из оцениваемого распределения. В связи с тем что в уравнении (4) присутствует индикаторная функция и, соответственно, оцениваемые параметры являются линейно зависимыми, нам необходимо провести модификацию базового алгоритма.

Проблема сэмплирования при линейных ограничениях

Для описания метода сэмплирования рассмотрим сначала все линейные ограничения задачи, которые представимы в следующем виде:

$$Bz = T, \tag{5}$$

где B и T - известная матрица и известный вектор соответственно; z - неизвестный вектор коэффициентов прямых затрат. Здесь мы просто переписали все линей-ные ограничения в стандартном представлении системы линейных алгебраических уравнений (СЛАУ). Если исходная оцениваемая матрица A имеет размерность $n\times n$, в модели также k ограничений на нули и не рассматриваются ошибки измерения, то размерность матрицы B равна $(2n + k)\times n^2$, а размерность вектора T - $(2n + k)\times 1$.

Первые n строк матрицы B будут просто представлять собой кронекерово произведение единичной матрицы размера $n\times n$ и строки выпуска рассматриваемых отраслей $1\times n$, а первые n строк вектора T будут совпадать с вектором промежуточного спроса, что соответствует первой части линейных ограничений системы (1). Следующие n строк матрицы B будут представлять собой кронекерово произведение строки выпуска отраслей $1\times n$ и единичной матрицы размера $n\times n$, а следующие n строк вектора T - вектор промежуточного потребления соответствующих отраслей. Каждая из последних k строк матрицы B будет содержать только нули за исключением одного единичного элемента на i-м месте, который будет соответствовать априорному ограничению на нуль i-го элемента вектора z. Соответственно, все послед-ние k строк вектора T будут нулевыми.

Так как в системе (5) в общем случае количество неизвестных превосходит количество уравнений, то данная система имеет бесконечное множество решений. Из линейной алгебры из-вестно, что любое решение вырожденной СЛАУ представимо в виде:

$$z = \tilde{z} + F^{(1)}\xi^{(1)}, \tag{6}$$

где $\tilde{z}$ - некоторое частное решение системы (5); $F^{(1)}$ - фун-даментальная матрица решений однородной системы уравнений $Bz = 0$.

Частное решение и матрица фундаменталь-ных решений могут быть найдены с помощью метода Гаусса последовательного исключения переменных. Столбцы фундаментальной матрицы решений $F^{(1)} = f_1^{(1)},\dots, f_{n^2-p}^{(1)}$, являются линейно независимыми и представляют собой базис Евклидова подпространства, где p - число линейно независимых ограничений. В данном контексте функция апостериорного распределения принимает достаточно наглядный вид. Действительно, если найти базис ортогонального дополнения $F^{(2)} = f_1^{(2)},\dots, f_p^{(2)}$, то можно перейти в новую систему координат:

$$\begin{pmatrix}\xi^{(1)}\\ \xi^{(2)}\end{pmatrix} = \left(F^{(1)}\; F^{(2)}\right)^{-1}(z - \tilde{z}). \tag{7}$$

В данной системе координат функция плотности распределения для компоненты $p(z)L(\tilde{Y}|z)$ принимает вид:

$$p_{\xi^{(1)}\xi^{(2)}}(\xi^{(1)}\xi^{(2)}) \propto \det[F^{(1)} F^{(2)}]\times$$
$$\times p_z(\tilde{z} + F^{(1)}\xi^{(1)} + F^{(2)}\xi^{(2)})\times$$
$$\times L(Y \mid \tilde{z} + F^{(1)}\xi^{(1)} + F^{(2)}\xi^{(2)}). \tag{8}$$

Тогда функция плотности апостериорно-го распределения просто является условной функцией плотности распределения вектора $\xi^{(1)}$ при условии, что координаты вектора $\xi^{(2)}$ равны нулю. И в случае непрерывного распределения принимает вид:

$$p_\xi(\xi \mid \text{data}) = p_{\xi^{(1)}|\xi^{(2)}}(\xi^{(1)} \mid \xi^{(2)} = 0) = \frac{p_{\xi^{(1)}\xi^{(2)}}(\xi^{(1)}, \xi^{(2)} = 0)}{p_{\xi^{(2)}}(\xi^{(2)} = 0)}. \tag{9}$$

Если бы априорное распределение было нормальным (в совокупности с нормальной ошибкой измерения), то мы могли бы получить аналитическое выражение для функции апостериорного распределения. Действительно, условное распределение нормального распределения также является нормальным. Но из-за того что рассматриваемые параметры, по определению, ограничены на отрезке [0, 1], мы не можем использовать нормальное распределение в явном виде, так как ничто не будет гарантировать неотрицательность коэффициентов прямых затрат.

Таким образом, в окончательной форме апос-териорное распределение в новом базисе запи-сывается в виде:

$$p_{\xi}(\xi \mid \text{data}) \propto p_z(z + F^{(1)}\xi^{(1)}) \cdot L(Y \mid z + F^{(1)}\xi^{(1)}). \quad (10)$$

Соответственно, мы можем применить базовые методы сэмплирования Монте-Карло по схеме марковской цепи из апостериорного распределения для вектора $\xi^{(1)}$, после чего перейти в исходную систему координат и получить выборку из апостериорного распределения для вектора коэффициентов прямых затрат.

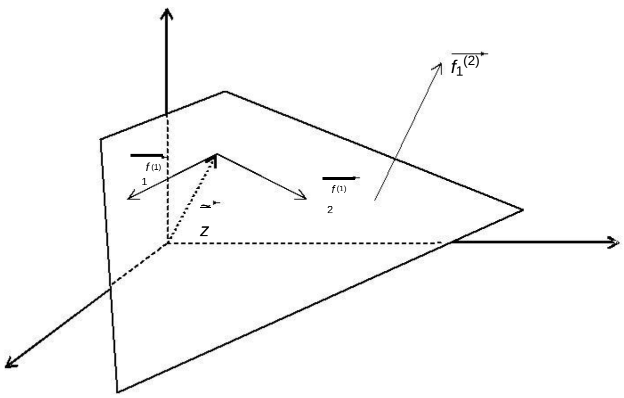

Рис. 1. **Переход в новую систему координат**

Основная идея представленного выше изложения отображена на рис. 1 на иллюстративном примере в трехмерном пространстве. Представленный рисунок иллюстрирует ситуацию, когда в модели три параметра и одно линейное ограничение. Данное линейное ограничение описывает плоскость в трехмерном пространстве, которая является подпространством размерностью 2 исходного пространства. Соответственно, три рассматриваемых параметра могут меняться только вдоль этой плоскости. Если перейти в новую систему координат, то функция апостери-орного распределения будет зависеть лишь от двух параметров - значений координат по векторам [$f_1^{(1)}$ $f_2^{(1)}$]. Таким образом, мы можем провести сэмплирование из функции апостериорного распределения по данным двум параметрам и получить множество точек на плоскости, которые будут, по определению, удовлетворять линейным ограничениям. После этого можно перейти обрат-но в исходное пространство, получив тем самым требуемую выборку из апостериорного распределения в трехмерном пространстве.

В численной имплементации генерации выборки из апостериорного распределения коэффициентов прямых затрат мы будем использовать схему Метрополиса-Гастингса [14, 17] с поэлементным обновлением параметров. Метод был реализован в программных пакетах Matlab[2] и R[3].

Изначально свойства алгоритма были протестированы на задачах, которые имеют аналитическое решение. Как было отмечено ранее, если в качестве априорного распределения коэффициентов прямых затрат выбрать нормальное распределение без каких-либо ограничений на диапазон изменения оцениваемых коэффициентов, то можно получить аналитическую формулу для апостериорного распределения, которое также будет нормальным распределением.

Рассмотрим другой упрощенный пример, имеющий аналитическое решение. Предполо-жим, что сумма коэффициентов прямых затрат по столбцам равна единице [например, доля валовой добавленной стоимости (ВДС) в выпуске также является оцениваемой величиной], и дополнительных линейных ограничений в модель не вводится. Тогда, если в качестве априорного распределения коэффициентов прямых затрат выбрать равномерное распределение на отрезке [0, 1], то апостериорное распределение для отдельного коэффициента имеет бета-распределение $B(1, N\text{-}1)$, где N - количество оцениваемых параметров в отдельном столбце неизвестной матрицы. Проведенные эксперименты показали, что алгоритм качественно воспроизводит оцениваемое апостериорное распределение, имеющее аналитическое решение.

[2] http://www.mathworks.com/
[3] http://www.r-project.org/

Экспериментальное построение симметричной таблицы «затраты-выпуск» 2003 г.

В качестве иллюстрации предложенного метода приводится актуализация таблиц «затраты-выпуск» для российской экономики в номенклатуре ОКОНХ[4] за 1998-2003 гг. [1-5]. В эксперименте оценивается «Симметричная таблица “затраты-выпуск” в основных ценах за 2003 г.» в разрезе 22-х отраслей. Таким образом, данная таблица трактуется как неизвестная, и проводится оценка коэффициентов прямых затрат в предположении, что на момент построения известны векторы выпуска и промежуточного спроса за 2003 г. и все опубликованные таблицы за предыдущие годы. В методе Байеса матрица коэффициентов прямых затрат 2002 г. использовалась как априорное среднее усеченного нормального распределения. Стандартные отклонения коэффициентов прямых затрат априорного распределения вычислялись на основе матриц с 1998 по 2002 г. Для оценки апостериорного распределения строились две марковские цепи длиной в 5 млн симуляций и отбрасывались первые 20% наблюдений. Далее для снижения автокорреляции выборка была разрежена, для чего на заключительном этапе построения выборки апостериорного распределения использовалась каждая 1000-я матрица исходной выборки.

На рис. 2 приведены диаграммы апостериорного распределения коэффициентов прямых затрат. Для краткости мы приводим только первые 8×6 коэффициентов исходной матрицы (в исходной матрице 484 коэффициента). Закрашенная область представляет собой функцию плотности апостериорного распределения; сплошная линия - истинное значение оцениваемого коэффициента; пунктирная линия - априорное среднее; штрихпунктирная линия - мода апостериорного распределения (вычислялась методами нелинейной оптимизации в программном пакете GAMS[5]). Для коэффициентов (6,3) и (6,4) графики отсутствуют, так как для них было наложено линейное ограничение о равенстве нулю.

Как показано на рис. 2, в целом апостериорное распределение достаточно хорошо накрывает истинные значения оцениваемых коэффициентов. В ряде случаев, например для ячеек (5,5) и (6,6), апостериорное распределение сильно сдвинулось от априорного в сторону истинного значения, то есть здесь доступная информация в линейных ограничениях значительно улучшает оценку данных параметров. Для некоторых коэффициентов, таких, как (2,6) и (8,3), апостериорное распределение сконцентрировано около априорного, и плотность вероятности в окрестности истинного коэффициента прямых затрат невелика. Указанная проблема указывает на недостаточность информации в учтенных данных, чтобы предсказать наблюдаемый сдвиг. Поэтому апостериорное значение близко к априорному. Это не является свойством метода, а скорее, информативностью данных и выбранной мерой расстояния. Дополнительная информация, включая статистические данные или экспертные оценки об отраслях, может быть учтена при оценивании в виде дополнительных ограничений и может улучшить оценки.

Отметим, что распределения отдельных коэффициентов на рис. 2 не являются независимыми между собой. Метод дает именно выборку матриц коэффициентов прямых затрат из апостериорного распределения, которые удовлетворяют всем ограничениям, и значения коэффициентов не могут браться из разных сгенерированных матриц. Однако среднее значение и мода удовлетворяют тем же ограничениям, что и сами матрицы.

С целью определения качества полученных точечных оценок было проведено сравнение результатов построения рассмотренной выше таблицы с альтернативными методами. Для этого использовался широкий набор мер качества приближения, в том числе принимающих во внимание различный вес в потерях от неправильного приближения больших и малых по величине коэффициентов прямых затрат (см., например, [22])[6]. Согласно результатам проведенного эксперимента, точечные оценки метода Байеса дают сопоставимые оценки с другими методами. Однако для данной конкретной таблицы рассматриваемого года полученные Байесовским методом оценки несколько уступают методу RAS, но при этом дают лучшие оценки по сравнению с рядом

[4] Проблематика построения таблиц «затраты-выпуск» в классификации экономической деятельности ОКВЭД обсуждается, например, в работах [6, 7].

[5] http://www.gams.com/.

[6] Для краткости изложения мы не приводим детальное описание данных экспериментов. Более подробное описание соответс-твующих результатов может быть предоставлено по запросу к авторам.

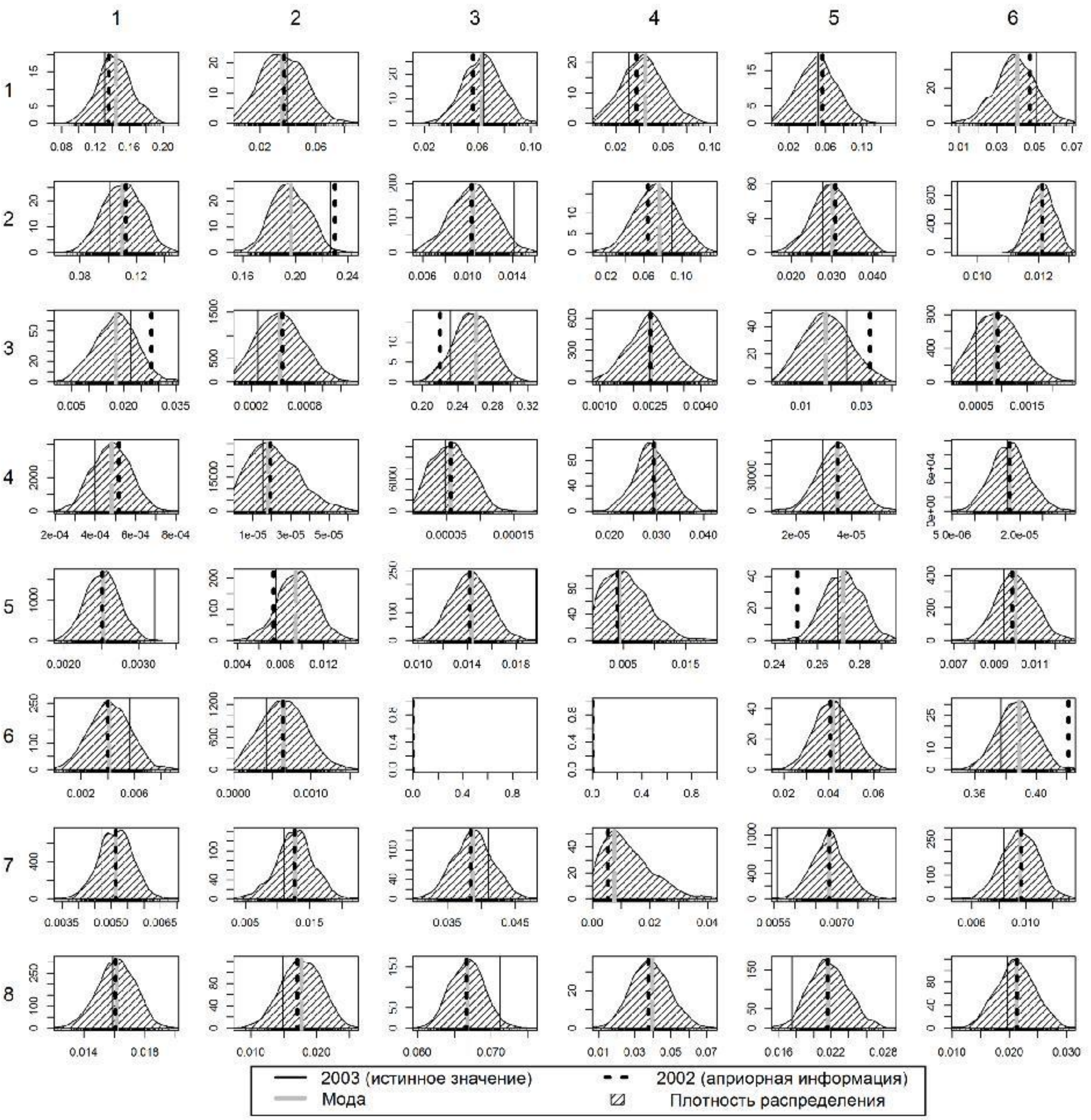

Рис. 2. **Апостериорное распределение первых 8 × 6 ячеек оцениваемой матрицы коэффициентов прямых затрат**

других методов, таких, как методы минимизации квадратов разностей, взвешенных квадратов разностей между оцениваемыми коэффициентами прямых затрат и соответствующими коэффициентами таблицы предыдущего года. Вероятно, это может объясняться тем, что публикуемые таблицы «затраты-выпуск» также были продлены с помощью технических методов и экспертных оценок, близких по логике продления к методу RAS (статистические обследования предпри-ятий для построения таблиц «затраты-выпуск» в 2003 г. не проводились). Тем не менее основным преимуществом предлагаемого в настоящей статье метода является не наилучшая точечная оценка, а выборка таблиц «затраты-выпуск» из апостериорного вероятностного распределения, несущая в себе дополнительную информацию, не предлагаемую другими методами.

* *
*

В работе рассмотрен Байесовский подход для построения таблиц «затраты-выпуск». Основной целью рассмотренной методики является вероятностная оценка неизвестных таблиц «затраты-выпуск». Предложен алгоритм сэмплирования матриц коэффициентов прямых затрат из апостериорного распределения, на выходе которого получается сколь угодно большая выборка таблиц «затраты-выпуск», не противоречащих имеющейся априорной информации и агрегированной статистике. Данная методика может использоваться для экстраполяции, интерполяции, дезагрегации и балансировки таблиц «затраты-выпуск», и более широко - матриц социальных счетов. На основе полученной выборки матриц, не противоречащих данным, могут быть сделаны оценки апостериорного среднего, моды, ковариации ячеек, а также выводы относительно наиболее неопределенных ячеек таблиц, вероятных структурных сдвигов, влияния ошибок наблюдения на неопределенность ячеек. Также полученные стохастические таблицы могут иметь широкое прикладное значение, например для анализа чувствительности результатов при калибровке экономико-математических моделей. Экспериментальное применение предложенного метода на реальных данных продемонстрировало адекватность и вычислительную доступность предлагаемой методики.

Литература

1. Система таблиц «затраты-выпуск» России за 1998-1999 годы. Стат. сб. / Госкомстат России. М., 2002. - 224 с.
2. Система таблиц «затраты-выпуск» России за 2000 год. Стат. сб. / Госкомстат России. М., 2003. - 116 с.
3. Система таблиц «затраты-выпуск» России за 2001 год. Стат. сб. / Федеральная служба государствен-ной статистики. М., 2004. - 116 с.
4. Система таблиц «затраты-выпуск» России за 2002 год. Стат. сб. / Росстат. М., 2005. - 116 с.
5. Система таблиц «затраты-выпуск» России за 2003 год. Стат. сб./ Росстат. М., 2006. - 116 с.
6. **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.
7. **Баранов Э.Ф., Ким И.А., Старицына Е.А.** Методологические вопросы реконструкции системы таблиц «затраты-выпуск» России за 2003 и последующие годы в структуре ОКВЭД - ОКПД // Вопросы статистики. 2011. №. 12. С. 3-8.
8. **Ершов Э.Б.** Неопределенность информации и устойчивость решения статической модели планового межотраслевого баланса / Проблемы народнохозяйственного оптимума. М.: Экономика, 1969.
9. **Almon C.** Recent methodological advances in input-output in the United States and Canada // Fourth International Conference on Input-Output Techniques, Geneva, 1968.
10. **Byron R.P.** The estimation of large social account matrices // Journal of the Royal statistical Society. Series A (General). 1978. Vol. 141. No. 3. P. 359-367.
11. **Dìaz B., Morillas A.** Incorporating uncertainty in the coefficients and multipliers of an IO table: A case study // Papers in Regional Science. 2011. Vol. 90. No. 4. P. 845-861.
12. **Goicoechea A., Hansen D.R.** An input-output model with stochastic parameters for economic analysis // AIIE Transactions. 1978. Vol. 10. No. 3. P. 285-291.
13. **Golan A., Judge G., Robinson S.** Recovering information from incomplete or partial multisectoral economic data // The Review of Economics and Statistics. 1994. Vol. 76. No. 3. P. 541-549.
14. **Hastings W.K.** Monte Carlo sampling methods using Markov chains and their applications // Biometrika. 1970. Vol. 57. No. 1. P. 97-109.
15. **Heckelei T., Mittelhammer R., Jansson T.** A Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models // Food and Resource Economics Discussion Paper. 2008. Vol. 2.
16. **Lenzen M**. et al. A cycling method for constructing input-output table time series from incomplete data // Economic Systems Research. 2012. Vol. 24. No. 4. P. 413-432.
17. **Metropolis N.** et al. Equation of state calculations by fast computing machines // The journal of chemical physics. 1953. Vol. 21. No. 6. P. 1087-1092.
18. **Quandt R.E.** Probabilistic errors in the Leontief system // Naval Research Logistics Quarterly. 1958. Vol. 5. No. 2. P. 155-170.
19. **Rampa G.** Using weighted least squares to deflate input-output tables // Economic Systems Research. 2008. Vol. 20. No. 3. P. 259-276.
20. **Rodrigues J.F.D.** A Bayesian approach to the balancing of statistical economic data // Entropy. 2014. Vol. 16. No. 3. P. 1243-1271.
21. **Stone R.** Input-output and national accounts. Paris, Organisation for European Economic Co-operation, 1961.
22. **Temurshoev U., Webb C., Yamano N.** Projection of supply and use tables: methods and their empirical assessment // Economic Systems Research. 2011. Vol. 23. No. 1. P. 91-123.
23. **Van der Ploeg F.** Reliability and the adjustment of sequences of large economic accounting matrices // Journal of the Royal Statistical Society. Series A (General). 1982. Vol. 145. No. 2. P. 169-194.
24. **Weale M.** The reconciliation of values, volumes and prices in the national accounts // Journal of the Royal Statistical Society. Series A (Statistics in Society). 1988. Vol. 151. No. 1. P. 211-221.

A Bayesian method for constructing input-output tables

Oleg Lugovoy
Author affiliation: The Russian Presidential Academy of National Economy and Public Administration (Moscow, Russia). E-mail: olugovoy@gmail.com.

Andrey Polbin
Author affiliation: The Russian Presidential Academy of National Economy and Public Administration (Moscow, Russia). E-mail: apolbin@gmail.com.

Vladimir Potashnikov
Author affiliation: The Russian Presidential Academy of National Economy and Public Administration (Moscow, Russia). E-mail: potashnikov.vu@gmail.com..

Tasks of updating, balancing, disaggregation of Input-Output Tables (IOT) face the researchers in the field of economics and statistics (for example, to calibrate computable general equilibrium models), as well as statistical services (to compile the IOTs). As compared to the well-known popular methods (RAS, cross-entropy minimization, and their analogs), which provide point estimates of unknown tables, the proposed approach targets estimation of joint probability distribution of input-output (IO) coefficients. With this goal the authors of this article develop a probabilistic model of joint distribution of the IO coefficients as a likelihood function of observed information (for example, output, value added, intermediate demand).This information from newly arrived data is being mixed with prior information of IO parameters (for example, known IOTs from former years) by Bayes rule.

The resultant posterior joint distribution can be estimated using Markov chain Monte Carlo (MCMC) sampling methods. Properties of the posterior distribution are defined by a set (sample) the required tables from this distribution. Whereas, each of the obtained tables is consistent with the observed data, constrains, and is also near to the prior given table or any other information on the cells. In contradiction to the point estimates, the stochastic tables directly incorporate information on the uncertainty of each estimated coefficient of direct costs of the IOTs, taking into account all the multivariate correlation between them. The proposed technique can be applied to extrapolate, interpolate, disaggregate, and balance the IOTs, and more widely - national accounts. In order to test the abovementioned technique the authors performed an experimental valuation of the Input-Output Tables for the Russian economy in 2003, based on tables for 1998-2002.

The results of the experimental application of the Bayesian method on the real data suggest adequacy and computational accessibility of the proposed technique.

Keywords: balancing, Input-Output Tables, Bayesian method.
JEL: C11, C6, C8.

References

1. The Input-Output accounts in Russia for the years of 1998-1999. Stat. publ. Moscow, Goskomstat Rossii, 2002. 224 p. (In Russ.).

2. The Input-Output accounts in Russia for the year of 2000. Stat. publ. Moscow, Goskomstat Rossii, 2003. 116 p. (In Russ.).

3. The Input-Output accounts in Russia for the year of 2001. Stat. publ. Moscow, Federal State Statistics Service, 2004. 116 p. (In Russ.).

4. The Input-Output accounts in Russia for the year of 2002. Stat. publ. Moscow, Rosstat, 2005. 116 p. (In Russ.).

5. The Input-Output accounts in Russia for the year of 2003. Stat. publ. Moscow, Rosstat, 2006. 116 p. (In Rus.).

6. **Baranov E.F., Kim I.A., Piontkovskiy D.I., Staritsyna Ye.A.** Voprosy postroyeniya tablits «zatraty-vypusk» Rossii v mezhdunarodnykh klassifikatorakh [Problems of constructing Russian Input-Output Tables into the International Classifications]. *Economic Journal of Higher School of Economics*, 2014, vol. 18, no 1, pp. 7-42. (In Russ.).

7. **Baranov E.F., Kim I.A., Staritsyna Ye.A.** Metodologicheskiye voprosy rekonstruktsii sistemy tablits «zatraty-vypusk» Rossii za 2003 i posleduyushchiye gody v strukture OKVED-OKPD [Methodological issues of reconstruction of the Input-Output accounts in Russia for the year of 2003 and subsequent years in the structure of OKVED-OKPD]. *Voprosy statistiki*, 2011, no.12, pp. 3-8. (In Russ.).

8. **Yershov E.B.** Neopredelonnost’ informatsii i ustoychivost’ resheniya staticheskoy modeli planovogo mezhotraslevogo balansa. In: Problemy narodnokhozyaystvennogo optimuma [Uncertainty of information and stability of the solution of the static model of the planned Input-Output account. *Problems of economic optimum*]. Moscow, Ekonomika Publ., 1969. (In Russ.).

9. **Almon C.** Recent methodological advances in input-output in the United States and Canada. Unpublished paper presented at the Fourth International Conference on Input-Output Techniques, Geneva, 1968.

10. **Byron R.P.** The estimation of large social account matrices. *Journal of the Royal Statistical Society*. Series A (General). 1978, vol. 141, no. 3, pp. 359-367.

11. **Dìaz B., Morillas A.** Incorporating uncertainty in the coefficients and multipliers of an IO table: A case study. *Papers in Regional Science*, 2011, vol. 90, no. 4, pp. 845-861.

12. **Goicoechea A., Hansen D.R.** An input-output model with stochastic parameters for economic analysis. *AIIE Transaction,* 1978, vol. 10, no. 3, pp. 285-291.

13. **Golan A., Judge G., Robinson S.** Recovering information from incomplete or partial multisectoral economic data. *The Review of Economics and Statistics*. 1994, vol. 76, no. 3, pp. 541-549.

14. **Hastings W.K.** Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, vol. 57, no. 1, pp. 97-109.

15. **Heckelei T., Mittelhammer R., Jansson T.** A Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models. *Food and Resource Economics Discussion Paper*, 2008, vol. 2.

16. **Lenzen M.** et al. A cycling method for constructing input-output table time series from incomplete data. *Economic Systems Research,* 2012, vol. 24, no. 4, pp. 413-432.

17. **Metropolis N.** et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 1953, vol. 21, no. 6, pp. 1087-1092.

18. **Quandt R.E.** Probabilistic errors in the Leontief system. *Naval Research Logistics Quarterly*, 1958, vol. 5, no. 2, pp. 155-170.

19. **Rampa G.** Using weighted least squares to deflate input-output tables. *Economic Systems Research*, 2008, vol. 20, no. 3, pp. 259-276.

20. **Rodrigues J.F.D.** A Bayesian approach to the balancing of statistical economic data. *Entropy*, 2014, vol. 16, no. 3, pp. 1243-1271.

21. **Stone R.** *Input-output and national accounts*. Paris, Organisation for European Economic Co-operation Publ., 1961.

22. **Temurshoev U., Webb C., Yamano N.** Projection of supply and use tables: methods and their empirical assessment. *Economic Systems Research*, 2011, vol. 23, no. 1, pp. 91-123.

23. **Van der Ploeg F.** Reliability and the adjustment of sequences of large economic accounting matrices. *Journal of the Royal Statistical Society*. Series A (General), 1982, vol. 145, no. 2, pp. 169-194.

24. **Weale M.** The reconciliation of values, volumes and prices in the national accounts. *Journal of the Royal Statistical Society*. Series A (Statistics in Society), 1988, vol. 151, no. 1, pp. 211-221.