Базилевский М. П.
M. P. Bazilevskiy

АЛГОРИТМ ОЦЕНИВАНИЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ ОДНОГО ВИДА МНОГОСЛОЙНЫХ НЕЭЛЕМЕНТАРНЫХ ЛИНЕЙНЫХ РЕГРЕССИЙ

AN ALGORITHM FOR ESTIMATING UNKNOWN PARAMETERS OF ONE TYPE OF MULTILAYER NON-ELEMENTARY LINEAR REGRESSIONS

Базилевский Михаил Павлович – кандидат технических наук, доцент кафедры «Математика» Иркутского государственного университета путей сообщения (Россия, Иркутск). E-mail: mik2178@yandex.ru.
Mikhail P. Bazilevskiy – PhD in Engineering, Associate Professor, Department of Mathematics, Irkutsk State Transport University (Russia, Irkutsk). E-mail: mik2178@yandex.ru.

Аннотация. В статье рассмотрен математический приём, состоящий в представлении функции Леонтьева в виде эквивалентной многослойной конструкции. На основе этого представления разработан алгоритм приближённого оценивания неизвестных параметров функции Леонтьева. С использованием реальной выборки данных из 101 наблюдения проведены вычислительные эксперименты. В результате работы алгоритма с использованием метода наименьших модулей приближённые оценки функций Леонтьева получились мало отличающимися от точных оценок, найденных посредством решения задач частично-булевого линейного программирования. При этом приближённые оценки были получены в разы быстрее, чем точные оценки. А при использовании метода наименьших квадратов скорость работы алгоритма оказалась ещё выше. Введена новая спецификация многослойных неэлементарных линейных регрессий, для которой справедлив разработанный алгоритм. С помощью метода наименьших квадратов по той же выборке построена одна из возможных форм новых моделей, оказавшаяся адекватнее функции Леонтьева.

Summary. The article discusses a mathematical technique consisting in representing the Leontief function as an equivalent multilayer structure. Based on this representation, an algorithm for approximate estimation of the Leontief function unknown parameters has been developed. Using a real data sample of 101 observations, computational experiments have been conducted. As a result of the algorithm's operation using the least absolute deviations method, approximate estimates of the Leontief functions turned out to be little different from the exact estimates found by solving 0-1 mixed integer linear programming problem. At the same time, approximate estimates were obtained several times faster than exact estimates. And when using the ordinary least squares method, the speed of the algorithm turned out to be even higher. A new specification of multilayer non-elementary linear regressions has been introduced, for which the developed algorithm is valid. Using the ordinary least squares method for the same sample, one of the possible forms of new models has been constructed, which turned out to be more adequate than the Leontief function.

Ключевые слова: регрессионный анализ, метод наименьших квадратов, метод наименьших модулей, функция Леонтьева, многослойная неэлементарная линейная регрессия, алгоритм оценивания.

Key words: regression analysis, ordinary least squares, least absolute deviations, Leontief function, multilayer non-elementary linear regression, estimation algorithm.

УДК 519.862.6

Введение. Развитие инструментария регрессионного анализа, относящегося к актуальному на сегодняшний день направлению разработки искусственного интеллекта – машинному обучению [1; 2], является актуальной научной задачей. Одним из возможных путей такого развития можно считать конструирование новых нелинейных структурных спецификаций регрессионных моделей и эффективных алгоритмов их оценки. Под структурной спецификацией [3] понимается математическая форма связи между переменными в модели регрессии. Новые нелинейные спецификации, с одной стороны, нужны для улучшения точности прогнозов зависимой переменной при исследовании действительно сложных процессов или явлений, с другой стороны, такие модели должны довольно просто интерпретироваться.

Среди существующих нелинейных спецификаций регрессионных моделей можно выделить следующих хорошо известных представителей. Во-первых, квазилинейные регрессии [4; 5], которые линейны по параметрам, но нелинейны по факторам. В них в качестве преобразований объясняющих переменных используются элементарные функции x^n, $\ln x$, $\sin x$ и т. д. Такие модели с лёгкостью оцениваются с помощью метода наименьших квадратов (МНК). В работах [6; 7] исследуется алгоритм идентификации параметров квазилинейной авторегрессии с помощью обобщённого метода наименьших модулей (МНМ). Во-вторых, регрессии, применяемые в экономике для построения производственных функций [8]. К ним относится функция Кобба – Дугласа [9; 10], Леонтьева [11; 12], функция с постоянной эластичностью замещения (CES-функция) [13; 14] и т. д.

Именно функция Леонтьева послужила основой при создании неэлементарных линейных регрессий (НЛР). Сначала в [15] был предложен алгоритм МНК-оценивания функции Леонтьева с двумя объясняющими переменными, что позволило ввести НЛР с бинарными операциями min и max (см., например, [16]). Обобщение НЛР с использованием операций min и max разной степени арности рассмотрено в [17]. При этом параллельно развивались так называемые модульные регрессии [18], в которых для преобразования переменных используется операция модуль. Поэтому в [19] были исследованы двухфакторные модели с симбиозом операций min, max и модуль. Они были названы двухслойными НЛР. Как известно, свойство многослойности присуще многим моделям машинного обучения. Например, многослойность подразумевают полином Колмогорова – Габора при реализации метода группового учёта аргументов [20; 21] и современные нейронные сети [22; 23]. В [24] предложена вложенная кусочно-линейная регрессия с двумя слоями, а построению многослойных модульных регрессий с помощью МНМ посвящена работа [25]. Изначально целью данной статьи была разработка нового алгоритма оценивания только функции Леонтьева с несколькими объясняющими переменными. Однако, успешно достигнув поставленной цели, автору удалось сформулировать ещё и новые спецификации многослойных НЛР.

Алгоритм оценивания функции Леонтьева. Пусть по выборке объёма n строится зависимость объясняемой переменной y от объясняющих переменных x_1, x_2, …, x_l. Если в качестве структурной спецификации выбрана производственная функция Леонтьева, то регрессионная модель с неизвестными параметрами α_1, α_2, …, α_l и ошибками ε_i, $i=\overline{1,n}$, имеет вид

$$y_i = \min\left\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, ..., \alpha_l x_{il}\right\} + \varepsilon_i,\ i=\overline{1,n}. \tag{1}$$

При использовании МНМ могут быть найдены точные оценки модели (1). Для этого требуется решить задачу частично-булевого линейного программирования (ЧБЛП), сформулированную, например, в [4; 12]. При использовании МНК можно получить лишь приближённые оценки модели (1). Один из возможных алгоритмов такого оценивания предложен в [4]. Он предполагает формирование l-мерного многогранника, из которого простым перебором выбираются наилучшие с точки зрения некоторого критерия МНК-оценки. Иной алгоритм предложен в [17]. Он также предполагает формирование многогранника, но меньшего объёма. Сформулируем далее ещё один алгоритм приближённого оценивания модели (1).

Нетрудно показать, что модель (1) равносильна следующей многослойной конструкции:

$$y_i = \min\{z_{i,l-2}, \alpha_l x_{il}\} + \varepsilon_i,\ i=\overline{1,n}, \tag{2}$$

где

$$z_{i,l-2} = \min\{z_{i,l-3}, \alpha_{l-1} x_{i,l-1}\},\ \ i=\overline{1,n},$$

$$z_{i,l-3} = \min\{z_{i,l-4}, \alpha_{l-2} x_{i,l-2}\},\ \ i=\overline{1,n},$$

$$\dots$$

$$z_{i2} = \min\{z_{i1}, \alpha_3 x_{i3}\}, \quad i = \overline{1,n},$$

$$z_{i1} = \min\{\alpha_1 x_{i1}, \alpha_2 x_{i2}\}, \quad i = \overline{1,n}.$$

Например, если $l = 4$, то получим многослойную конструкцию вида

$$y_i = \min\{z_{i2}, \alpha_4 x_{i4}\} + \varepsilon_i, \quad i = \overline{1,n}, \tag{3}$$

где

$$z_{i2} = \min\{z_{i1}, \alpha_3 x_{i3}\}, \quad i = \overline{1,n}, \tag{4}$$

$$z_{i1} = \min\{\alpha_1 x_{i1}, \alpha_2 x_{i2}\}, \quad i = \overline{1,n}. \tag{5}$$

Подставляя (5) в (4), а (4) в (3), получим модель

$$y_i = \min\left\{\min\left\{\min\left\{\alpha_1 x_{i1}, \alpha_2 x_{i2}\right\}, \alpha_3 x_{i3}\right\}, \alpha_4 x_{i4}\right\} + \varepsilon_i, \quad i = \overline{1,n},$$

которая равносильна производственной функции Леонтьева при $l = 4$:

$$y_i = \min\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, \alpha_3 x_{i3}, \alpha_4 x_{i4}\} + \varepsilon_i, \quad i = \overline{1,n}.$$

Действительно, выбор минимума из нескольких чисел может быть организован путём последовательного выбора на каждом шаге минимума из двух соседних чисел. Таким же образом определяется максимум из нескольких чисел. Поскольку, как отмечено в [17], модель (1) является частным случаем НЛР, то конструкцию (2) будем называть многослойной НЛР. Число слоёв в ней равно $(l-1)$.

Так, как это сделано в работе [15] для двухфакторных функций Леонтьева, вынесем из-под знака min в модели (1) параметр $\alpha_1 > 0$ и добавим в неё свободный член α_0:

$$y_i = \alpha_0 + \alpha_1 \min\left\{x_{i1}, \frac{\alpha_2}{\alpha_1} x_{i2}, \ldots, \frac{\alpha_l}{\alpha_1} x_{il}\right\} + \varepsilon_i, \quad i = \overline{1,n}. \tag{6}$$

При $\alpha_0 = 0$ модель (6) равносильна регрессии (1). Свободный член α_0 введён в модель (6) для того, чтобы при её МНК-оценивании была справедлива формула для коэффициента детерминации R^2.

С учётом замен $\lambda_2 = \dfrac{\alpha_2}{\alpha_1}$, …, $\lambda_l = \dfrac{\alpha_l}{\alpha_1}$, модель (6) принимает вид

$$y_i = \alpha_0 + \alpha_1 \min\left\{x_{i1}, \lambda_2 x_{i2}, \ldots, \lambda_l x_{il}\right\} + \varepsilon_i, \; i = \overline{1,n}. \tag{7}$$

Представим модель (7) в виде следующей многослойной конструкции:

$$y_i = \alpha_0 + \alpha_1 \min\{z^*_{i,l-2}, \lambda_l x_{il}\} + \varepsilon_i, \quad i = \overline{1,n}, \tag{8}$$

где

$$z^*_{ij} = \min\left\{z^*_{i,j-1}, \lambda_{j+1} x_{i,j+1}\right\}, \quad j = \overline{2,l-2}, \quad i = \overline{1,n},$$

$$z^*_{i1} = \min\{x_{i1}, \lambda_2 x_{i2}\}, \quad i = \overline{1,n}.$$

Как следует из [15], область возможных значений параметра λ_2 на первом слое НЛР (8)

$$D_1: \; \min\left\{\frac{x_{11}}{x_{12}}, \frac{x_{21}}{x_{22}}, \ldots, \frac{x_{n1}}{x_{n2}}\right\} \le \lambda_2 \le \max\left\{\frac{x_{11}}{x_{12}}, \frac{x_{21}}{x_{22}}, \ldots, \frac{x_{n1}}{x_{n2}}\right\}; \tag{9}$$

области возможных значений параметров λ_{j+1}, $j=\overline{2,l-1}$, на остальных слоях

$$D_j:\ \min\left\{\frac{z^*_{1,j-1}}{x_{1,j+1}},\frac{z^*_{2,j-1}}{x_{2,j+1}},...,\frac{z^*_{n,j-1}}{x_{n,j+1}}\right\}\le\lambda_{j+1}\le\max\left\{\frac{z^*_{1,j-1}}{x_{1,j+1}},\frac{z^*_{2,j-1}}{x_{2,j+1}},...,\frac{z^*_{n,j-1}}{x_{n,j+1}}\right\},\quad j=\overline{2,l-1}. \qquad (10)$$

Тогда для приближённого оценивания многослойной НЛР (8) нужно действовать последовательно. Сначала на первом слое найти область D_1, разбить её точками и вычислить в каждой из них значения переменной z_1^*. Затем на втором слое для каждой переменной z_1^* найти область D_2, разбить её точками и вычислить в каждой из них значения переменной z_2^*. Строгое математическое описание этого алгоритма можно сформулировать следующим образом:

Шаг 1 (первый слой). Идентифицировать область D_1 возможных значений параметра λ_2 по формуле (9). Равномерно разбить полученный отрезок r точками $\lambda_2^{(1)}$, $\lambda_2^{(2)}$, …, $\lambda_2^{(r)}$. В каждой такой точке вычислить значения переменных $z_1^*\left(\lambda_2^{(i)}\right)=\min\{x_1,\lambda_2^{(i)}x_2\}$, $i=\overline{1,r}$.

Шаг 2 (второй слой). Для каждой переменной $z_1^*\left(\lambda_2^{(i)}\right)$, $i=\overline{1,r}$, идентифицировать область D_2 возможных значений параметра λ_3 по формуле (10). Первый полученный отрезок равномерно разбить r точками $\lambda_3^{(1,1)}$, $\lambda_3^{(1,2)}$, …, $\lambda_3^{(1,r)}$, второй – $\lambda_3^{(2,1)}$, $\lambda_3^{(2,2)}$, …, $\lambda_3^{(2,r)}$ и т. д. В каждой такой точке вычислить значения переменных $z_2^*\left(\lambda_3^{(i,j)}\right)=\min\{z_1^*,\lambda_3^{(i,j)}x_3\}$, $i=\overline{1,r}$, $j=\overline{1,r}$.

Шаг 3 (третий слой). Для каждой переменной $z_2^*\left(\lambda_3^{(i,j)}\right)$, $i=\overline{1,r}$, $j=\overline{1,r}$, идентифицировать область D_3 возможных значений параметра λ_4 по формуле (10). Первый полученный отрезок равномерно разбить r точками $\lambda_4^{(1,1,1)}$, $\lambda_4^{(1,1,2)}$, …, $\lambda_4^{(1,1,r)}$, второй – $\lambda_4^{(1,2,1)}$, $\lambda_4^{(1,2,2)}$, …, $\lambda_4^{(1,2,r)}$ и т. д. В каждой такой точке вычислить значения переменных $z_3^*\left(\lambda_4^{(i,j,k)}\right)=\min\{z_2^*,\lambda_4^{(i,j,k)}x_4\}$, $i=\overline{1,r}$, $j=\overline{1,r}$, $k=\overline{1,r}$.

…

Шаг (l-1) (последний слой). Для каждой переменной $z_{l-2}^*\left(\lambda_{l-1}^{(i_1,i_2,...,i_{l-2})}\right)$, $i_1,i_2,...,i_{l-2}=\overline{1,r}$, идентифицировать область D_{l-1} возможных значений параметра λ_l по формуле (10). Каждый из полученных отрезков равномерно разбить r точками. В каждой такой точке $\lambda_l^{(i_1,i_2,...,i_{l-1})}$, $i_1,i_2,...,i_{l-1}=\overline{1,r}$, оценить с помощью МНК или МНМ модель парной линейной регрессии (8). Из полученного набора r^{l-1} моделей выбрать ту, у которой сумма квадратов или модулей ошибок минимальна.

Стоит отметить, что описанный выше алгоритм предполагает использование только точек разбиения, расположенных внутри идентифицированных отрезков. При оценивании многослойной НЛР надёжнее будет использовать и концы этих отрезков, но при этом количество парных регрессий на последнем шаге алгоритма увеличится с r^{l-1} до $(r+2)^{l-1}$.

Разработанный в [17] алгоритм оценивания функции Леонтьева (7) предполагает формирование для оценок параметров λ_2, λ_3, …, λ_l системы линейных неравенств, решением которой является выпуклый многогранник в многомерном пространстве. Проблема состояла в том, как выбрать внутри этого многогранника точки. Предложенный выше алгоритм не требует решения этой проблемы, поскольку все точки разбиения лежат внутри этого многогранника.

Заметим, что в многослойной конструкции (8) на первом слое преобразуются переменные x_1 и x_2, а далее на каждом слое последовательно добавляются переменные x_3, x_4 и т. д. Изменится ли результат работы алгоритма, если поменять порядок переменных в слоях? Ответ – да. С изменением порядка переменных в слоях выпуклый многогранник не меняется, но меняются точки его разбиения. Поэтому при малых r разница при изменении порядка факторов может быть существенной, но с ростом r эта разница уменьшается.

Вычислительные эксперименты. Главной целью проведения вычислительных экспериментов было сравнить на реальных данных скорость и качество точного МНМ-оценивания функции Леонтьева, пользуясь аппаратом ЧБЛП, и приближённого МНМ- и МНК-оценивания с помощью предложенного алгоритма.

Для проведения вычислительных экспериментов использовался обычный персональный компьютер с процессором AMD Ryzen 3 4300U (тактовая частота 2.7 ГГц) и 16 ГБ оперативной памяти.

Для точного МНМ-оценивания функции Леонтьева был использован математический аппарат из работ [4; 12]. При формировании задач ЧБЛП большое число M было выбрано равным 100 000. Для решения задач ЧБЛП был использован оптимизационный решатель LPSolve. Лимит времени в нём для получения оптимального решения был выбран 1800 с. (30 мин.). При превышении лимита времени автоматически фиксировалось полуоптимальное решение задачи.

Для приближённого оценивания функции Леонтьева с помощью МНК и МНМ был использован предложенный выше алгоритм. Предварительно он был реализован в виде программы на скриптовом языке программирования hansl эконометрического пакета Gretl. Для МНМ-оценивания парных регрессий на последнем шаге алгоритма была использована встроенная в пакет Gretl функция «lad», а для МНК-оценивания – «ols».

Для обработки были выбраны встроенные в пакет Gretl данные по кабельным системам (файл data7-22.gdt). Объём выборки n составляет 101. Объясняемой переменной выступает sub – число подписчиков кабельной системы. К объясняющим переменным относятся:

homes – количество домов с установленной кабельной системой;

inst – плата за установку (в долларах);

svc – ежемесячная плата за обслуживание;

cblchanl – количество телевизионных сигналов, передаваемых кабельной системой;

tvchanl – количество полученных телевизионных сигналов;

princome – доход на душу населения.

Для удобства обозначим все эти переменные, как y, x_1, x_2, x_3, x_4, x_5 и x_6 соответственно.

В ходе экспериментов оценивались функции Леонтьева с пятью объясняющими переменными, причём как без свободного члена, так и с ним. Сначала эти 12 моделей были оценены точно с помощью МНМ. Затем эти 12 регрессий оценивались приближённо с помощью МНМ и МНК. При этом для каждой модели число точек r разбиения отрезков в алгоритме устанавливалось равным 6, 9 и 12. Также использовались точки на концах идентифицированных отрезков. Результаты проведённых экспериментов приведены в табл. 1.

В табл. 1 в первом столбце приведены составы объясняющих переменных в функциях Леонтьева, во втором и третьем столбце – результаты точного МНМ-оценивания, а именно суммы модулей остатков моделей $\sum|e|$ и время t_1 решения задач ЧБЛП в пакете LPSolve. Если в третьем столбце зафиксировано значение 1800, то это означает, что оптимальность решения не доказана в установленный лимит времени. Такое решение принято называть полуоптимальным. В четвёртом столбце указаны значения параметра r, а в пятом, шестом, седьмом и восьмом столбцах – результаты приближённого МНМ-оценивания, а конкретно суммы модулей остатков моделей $\sum|e|$, относительные отклонения δ в процентах, время t_2 работы алгоритма в пакете Gretl и отношения t_1/t_2. Относительное отклонение δ находится по формуле

$$\delta = \frac{J_{\text{алг}} - J_{\text{ЧБЛП}}}{J_{\text{алг}}} \cdot 100\%,$$

где $J_{\text{алг}}$ – сумма модулей остатков, полученная в результате приближённого оценивания в Gretl; $J_{\text{ЧБЛП}}$ – сумма модулей остатков, полученная в результате решения задачи ЧБЛП в LPSolve. Заметим, что $J_{\text{ЧБЛП}}$ не обязательно принимает оптимальное значение, поскольку в LPSolve установлен лимит времени на поиск решения. Близкое к нулю положительное значение δ говорит о несущественном превосходстве использования аппарата ЧБЛП по сравнению с предложенным алгоритмом.

Таблица 1

Результаты вычислительных экспериментов

Состав пере-менных	Задача ЧБЛП		Предложенный алгоритм							
				МНМ				МНК		
	$\sum\|e\|$	t_1, с	r	$\sum\|e\|$	δ, %	t_2, с	t_1/t_2	$\sum e^2$	t_3, с	t_1/t_3
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Без свободного члена										
x_1, x_2, x_3, x_4, x_5	2414,7	598,8	6	2423,4	0,36	46,7	12,82	190 231,9	19,8	30,24
			9	2422,7	0,33	160,7	3,73	189 799,0	67,9	8,82
			12	2440,1	1,04	426,4	1,40	187 820,1	173,9	3,44
x_1, x_2, x_3, x_4, x_6	2414,7	1499,9	6	2423,4	0,36	47,8	31,38	190 231,9	20,2	74,25
			9	2422,7	0,33	159,6	9,40	189 799,0	67,8	22,12
			12	2440,1	1,04	429,8	3,49	187 820,1	182,6	8,21
x_1, x_2, x_3, x_5, x_6	2414,7	339,3	6	2423,4	0,36	46,9	7,23	191 165,8	19,9	17,05
			9	2422,7	0,33	166,7	2,04	190 758,4	72,4	4,69
			12	2440,1	1,04	440,3	0,77	189 652,5	177,0	1,92
x_1, x_2, x_4, x_5, x_6	2414,7	839,1	6	2423,4	0,36	45,1	18,61	190 231,9	20,0	41,96
			9	2422,7	0,33	158,1	5,31	189 799,0	67,7	12,39
			12	2440,1	1,04	425,7	1,97	187 820,1	184,4	4,55
x_1, x_3, x_4, x_5, x_6	2480,1	790,7	6	2490,9	0,43	46,7	16,93	192 009,4	20,2	39,14
			9	2488,0	0,32	165,3	4,78	190 307,3	68,9	11,48
			12	2485,2	0,21	423,1	1,87	190 244,6	185,1	4,27
x_2, x_3, x_4, x_5, x_6	4472,6	1800	6	3914,3	-14,26	46,7	38,54	428 241,9	20,0	90,00
			9	3920,4	-14,09	160,9	11,19	430 038,0	68,0	26,47
			12	3920,1	-14,09	429,6	4,19	427 806,7	184,3	9,77
Со свободным членом										
x_1, x_2, x_3, x_4, x_5	2413,6	1558,4	6	2423,1	0,39	53,9	28,91	188 300,2	20,5	76,02
			9	2420,2	0,27	184,6	8,44	188 426,3	72,5	21,50
			12	2437,8	0,99	480,7	3,24	186 235,9	174,0	8,96
x_1, x_2, x_3, x_4, x_6	2413,6	1800	6	2423,1	0,39	54,4	33,09	188 300,2	19,8	90,91
			9	2420,2	0,27	192,2	9,37	188 246,6	68,2	26,39
			12	2437,8	0,99	488,2	3,69	186 235,9	183,2	9,83
x_1, x_2, x_3, x_5, x_6	2413,6	905,1	6	2423,1	0,39	55,4	16,34	189 977,0	20,1	45,03
			9	2420,2	0,27	189,6	4,77	189 348,2	69,0	13,12
			12	2437,8	0,99	505,3	1,79	187 981,8	176,3	5,13
x_1, x_2, x_4, x_5, x_6	2413,6	1800	6	2423,1	0,39	54,9	32,79	188 300,2	20,3	88,67
			9	2420,2	0,27	192,7	9,34	188 246,6	74,2	24,26
			12	2437,8	0,99	503,4	3,58	186 235,9	177,6	10,14
x_1, x_3, x_4, x_5, x_6	2479,3	1800	6	2490,8	0,46	55,2	32,61	189 982,7	20,0	90,00
			9	2487,9	0,35	193,3	9,31	188 421,2	73,1	24,62
			12	2485,1	0,23	522,7	3,44	188 303,2	176,9	10,18
x_2, x_3, x_4, x_5, x_6	4339,3	1800	6	3899,6	-11,28	56,8	31,69	427 996,2	20,6	87,38
			9	3892,6	-11,48	195,3	9,22	428 762,3	69,7	25,82
			12	3895,7	-11,39	509,1	3,54	427 797,1	185,7	9,69

В девятом, десятом и одиннадцатом столбцах табл. 1 приведены результаты приближённого МНК-оценивания функций Леонтьева, а именно суммы квадратов остатков моделей $\sum e^2$, время t_3 работы алгоритма в пакете Gretl и отношения t_1/t_3.

По табл. 1 можно сделать следующие выводы.

1. Время МНМ-оценивания функции Леонтьева, предполагающего решение задачи ЧБЛП в пакете LPSolve, непредсказуемо. Так, без использования свободного члена в моделях для одного и того же числа ограничений в задачах ЧБЛП это время варьировалось в диапазоне от 339,3 до 1800 с., а со свободным членом – от 905,1 до 1800 с. Также видно, что без использования свободного члена только в одном случае из шести не было получено оптимальное решение, тогда как включение свободного члена увеличило число таких случаев до четырёх из шести.

2. При МНМ-оценивании функции Леонтьева без свободного члена с использованием предложенного алгоритма относительные отклонения лежат в диапазоне: от –14,26 до 0,43 % при $r = 6$; от –14,09 до 0,33 % при $r = 9$; от –14,09 до 1,04 % при $r = 12$. Положительные значения в указанных диапазонах довольно малы, т. е. нет никакой серьёзной разницы между суммами модулей остатков, полученными точным и приближённым методом. А отрицательные значения относительных отклонений и вовсе говорят о том, что предложенный алгоритм нашёл более качественную по величине суммы модулей остатков модель, чем метод ЧБЛП в установленный лимит времени. Время МНМ-оценивания предложенным алгоритмом лежит в диапазоне: от 45,1 до 47,8 с. при $r = 6$; от 158,1 до 166,7 с. при $r = 9$; от 423,1 до 440,3 с. при $r = 12$. Отношение времени t_1 МНМ-оценивания с помощью аппарата ЧБЛП ко времени t_2 МНМ-оценивания с помощью нашего алгоритма попадает в диапазон: от 7,23 до 38,54 при $r = 6$; от 2,04 до 11,19 при $r = 9$; от 0,77 до 4,19 при $r = 12$. Практически все эти отношения превосходят единицу, т. е. скорость работы предложенного алгоритма в подавляющем большинстве случаев оказалась выше, чем скорость решения задач ЧБЛП.

3. Включение свободного члена в функцию Леонтьева повышает вычислительную трудность задачи её МНМ-оценивания. Так, время оценивания предложенным алгоритмом лежит в диапазоне: от 53,9 до 56,8 с. при $r = 6$; от 184,6 до 195,3 с. при $r = 9$; от 480,7 до 522,7 с. при $r = 12$. Отношение времени t_1 МНМ-оценивания с помощью аппарата ЧБЛП ко времени t_2 МНМ-оценивания с помощью нашего алгоритма попадает в диапазон: от 16,34 до 33,09 при $r = 6$; от 4,77 до 9,37 при $r = 9$; от 1,79 до 3,69 при $r = 12$. Из этого следует, что скорость работы предложенного алгоритма во всех случаях оказалась выше, чем скорость решения задач ЧБЛП. Относительные отклонения лежат в диапазоне: от –11,28 до 0,46 % при $r = 6$; от –11,48 до 0,35 % при $r = 9$; от –11,39 до 0,99 % при $r = 12$. Следовательно, точный метод МНМ-оценивания по величине сумм модулей остатков либо незначительно выигрывает у приближённого алгоритма, либо существенно ему уступает.

4. Время МНК-оценивания функции Леонтьева без свободного члена с помощью предложенного алгоритма лежит в диапазоне: от 19,8 до 20,2 с. при $r = 6$; от 67,7 до 72,4 с. при $r = 9$; от 173,9 до 185,1 с. при $r = 12$. А при использовании свободного члена: от 19,8 до 20,6 с. при $r = 6$; от 68,2 до 74,2 с. при $r = 9$; от 174 до 185,7 с. при $r = 12$. Из этого следует, во-первых, что при использовании МНК включение в модель свободного члена абсолютно не повлияло на скорость работы алгоритма. Во-вторых, применение МНК в алгоритме гораздо эффективнее, чем применение МНМ. Так, без использования свободного члена отношение времени t_1 МНМ-оценивания с помощью аппарата ЧБЛП ко времени t_3 МНК-оценивания с помощью нашего алгоритма попадает в диапазон: от 17,05 до 90 при $r = 6$; от 4,69 до 26,47 при $r = 9$; от 1,92 до 9,77 при $r = 12$. А при включении свободного члена это отношение лежит в диапазоне: от 45,03 до 90,91 при $r = 6$; от 13,12 до 26,39 при $r = 9$; от 5,13 до 10,18 при $r = 12$.

5. Увеличение числа точек r при использовании любого метода оценивания модели приводит к увеличению времени работы алгоритма. При этом нет гарантии, что качество модели при увеличении r будет улучшено.

По табл. 1 наилучшим качеством аппроксимации обладает следующая оценённая с помощью МНК функция Леонтьева:

$$\tilde{y} = 5,9792 + 0,4302\min\left\{x_1, 23.178x_2, 18.343x_4, 90.575x_5, 44.035x_6\right\}, \tag{11}$$

для которой $\sum e^2 = 186235,9$, а коэффициент детерминации $R^2 = 0,7056$. Заметим, что модель (11) адекватнее линейной регрессии со всеми шестью объясняющими переменными, для которой $\sum e^2 = 218447,4$, а $R^2 = 0,6547$.

Многослойная НЛР. Достоинство разработанного алгоритма приближённого оценивания функций Леонтьева в том, что он будет справедлив и для некоторых других математических форм многослойных НЛР.

Представим многослойную конструкцию (8) в следующем виде:

$$y_i = \alpha_0 + \alpha_1 g_{k_{l-1}}(\lambda_l, z^*_{i,l-2}, x_{i,s_l}) + \varepsilon_i, \ \ i = \overline{1,n}, \tag{12}$$

где

$$z^*_{ij} = g_{k_j}(\lambda_{j+1}, z^*_{i,j-1}, x_{i,s_{j+1}}), \ \ j = \overline{2, l-2}, \ \ i = \overline{1,n},$$

$$z^*_{i1} = g_{k_1}\left(\lambda_2, x_{i,s_1}, x_{i,s_2}\right), \ \ i = \overline{1,n};$$

здесь s_1, s_2, …, s_l – элементы вектора S, содержащего последовательность номеров, включаемых в слои объясняющих переменных; k_j, $j = \overline{1, l-1}$ – номер выбранного в каждом слое неэлементарного преобразования g_{k_j} двух переменных из набора $G = \left(\min\{u_1, \lambda u_2\}, \max\{u_1, \lambda u_2\}, \min\{u_1, \lambda + u_2\}, \max\{u_1, \lambda + u_2\}\right)$.

Например, если $S = \left(1, 2, 3, ..., l\right)$, а $k_j = 1$, $j = \overline{1, l-1}$, то многослойная НЛР (8) трансформируется в функцию Леонтьева.

Комбинируя в (12) элементы вектора S и состав неэлементарных преобразований, можно получить множество различных форм многослойных НЛР. При этом для них будет справедлив предложенный нами алгоритм. При использовании в нём преобразования $\min\{u_1, \lambda + u_2\}$ или $\max\{u_1, \lambda + u_2\}$, как следует из работ [15–17], область возможных значений параметра λ_2 на первом слое НЛР (12)

$$D_1: \ \min\left\{x_{1,s_1} - x_{1,s_2}, ..., x_{n,s_1} - x_{n,s_2}\right\} \le \lambda_2 \le \max\left\{x_{1,s_1} - x_{1,s_2}, ..., x_{n,s_1} - x_{n,s_2}\right\};$$

области возможных значений параметров λ_{j+1}, $j = \overline{2, l-1}$, на остальных слоях

$$D_j: \ \min\left\{z^*_{1,j-1} - x_{1,s_{j+1}}, ..., z^*_{n,j-1} - x_{n,s_{j+1}}\right\} \le \lambda_{j+1} \le \max\left\{z^*_{1,j-1} - x_{1,s_{j+1}}, ..., z^*_{n,j-1} - x_{n,s_{j+1}}\right\},$$

$$j = \overline{2, l-1}.$$

Для демонстрации работоспособности алгоритма по тем же данным с помощью МНК оценивалась многослойная НЛР с параметрами $S = (5, 2, 4, 1, 6)$, $k_1 = 3$, $k_2 = 1$, $k_3 = 3$, $k_4 = 1$. В результате при $r = 12$ была построена следующая четырёхслойная модель:

$$\tilde{y} = 503,0913 + 0,4639\min\left\{z^*_3, -30,387x_6\right\}, \tag{13}$$

где

$$z^*_3 = \min\{z^*_2, -1079,1 + x_1\},$$

$$z_2^* = \min\{z_1^*, -7,6481x_4\},$$

$$z_1^* = \min\{x_5, -55,838 + x_2\},$$

для которой $\sum e^2 = 182905,4$, а коэффициент детерминации $R^2 = 0,7109$. Как видно, по обоим показателям многослойная НЛР (13) адекватнее функции Леонтьева (11).

Заключение. С использованием математического приёма представления функции Леонтьева в виде многослойной конструкции сформулирован новый алгоритм приближённого оценивания её неизвестных параметров. Этот алгоритм может быть использован при оценивании функции Леонтьева как с помощью МНМ, так и с помощью МНК. В ходе вычислительных экспериментов установлено, что при использовании МНМ качество полученных по нашему алгоритму моделей оказалось либо несущественно ниже, либо существенно выше (в установленный лимит времени), чем качество построенных с использованием аппарата ЧБЛП регрессий. При этом скорость работы предложенного алгоритма в подавляющем большинстве случаев оказалась в разы выше, чем скорость решения задач ЧБЛП. А при использовании МНК скорость работы алгоритма ещё больше увеличилась. Полученные результаты подтверждают целесообразность применения разработанного алгоритма на практике. К тому же вызывает научный интерес дальнейшее исследование предложенной в статье многослойной НЛР, которая является обобщением функции Леонтьева и для которой справедлив разработанный алгоритм.

ЛИТЕРАТУРА

1. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2 (3), 160.
2. Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. Nature reviews Molecular cell biology, 23 (1), 40-55.
3. Айвазян, С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – 1005 с.
4. Носков, С. И. Технология моделирования объектов с нестабильным функционированием и неопределённостью в данных / С. И. Носков. – Иркутск: РИЦ ГП Облинформпечать, 1996. – 320 с.
5. Никонорова, Ю. В. Применение метода наименьших квадратов для нахождения параболических функций, описывающих динамику цены биткоина и цены акций МММ / Ю. В. Никонорова // Вестник НИЯУ МИФИ. – 2021. – Т. 10. – № 4. – С. 357-362.
6. Аботалеб, М. С. А. Исследование возможностей параллелизма для прогнозирования с использованием квазилинейного рекуррентного уравнения / М. С. А. Аботалеб, Т. А. Макаровских, А. В. Панюков // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. – 2023. – Т. 12. – № 4. – С. 94-109.
7. Аботалеб, М. С. А. Установление значимости коэффициентов квазилинейного уравнения n-факторной авторегрессии / М. С. А. Аботалеб // Проблемы информатики. – 2024. – № 3 (64). – С. 5-28.
8. Клейнер, Г. Б. Производственные функции: теория, методы, применение / Г. Б. Клейнер. – М.: Финансы и статистика, 1986. – 239 с.
9. Mohajan, H. K. (2021). Estimation of cost minimization of garments sector by Cobb-Douglas production function: Bangladesh perspective. Annals of Spiru Haret University. Economic Series, 21 (2), 267-299.
10. Vasyl’yeva, O. (2021). Assessment of factors of sustainable development of the agricultural sector using the Cobb-Douglas production function. Baltic Journal of Economic Studies, 7 (2), 37-49.
11. Михеев, А. В. Вероятностный подход к определению производственных функций / А. В. Михеев // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2021. – № 4. – С. 82-94.
12. Носков, С. И. Программный комплекс построения некоторых типов кусочно-линейных регрессий / С. И. Носков, А. А. Хоняков // Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – № 3. – С. 47-55.
13. Усова, А. А. Анализ модели роста с производственной CES-функцией / А. А. Усова, А. М. Тарасьев // Математическая теория игр и её приложения. – 2022. – Т. 14. – № 4. – С. 96-114.
14. Черемухин, А. Д. Исследование применимости CES-функций для описания процессов сельскохозяйственного производства / А. Д. Черемухин, Г. В. Груздев // Вестник НГИЭИ. – 2022. – № 9 (136). – С. 65-83.

15. Базилевский, М. П. МНК-оценивание параметров специфицированных на основе функций Леонтьева двухфакторных моделей регрессии / М. П. Базилевский // Южно-Сибирский научный вестник. – 2019. – № 2 (26). – С. 66-70.
16. Базилевский, М. П. Отбор информативных операций при построении линейно-неэлементарных регрессионных моделей / М. П. Базилевский // International Journal of Open Information Technologies. – 2021. – Т. 9. – № 5. – С. 30-35.
17. Базилевский, М. П. Обобщение неэлементарных линейных регрессий / М. П. Базилевский // Моделирование и анализ данных. – 2023. – Т. 13. – № 2. – С. 85-98.
18. Базилевский, М. П. Программное обеспечение для оценивания модульных линейных регрессий / М. П. Базилевский // Информационные и математические технологии в науке и управлении. – 2023. – № 3 (31). – С. 136-146.
19. Базилевский, М. П. Алгоритм приближённого оценивания с помощью метода наименьших квадратов двухслойных неэлементарных линейных регрессий с двумя объясняющими переменными / М. П. Базилевский // Современные наукоёмкие технологии. – 2024. – № 4. – С. 10-14.
20. Установление технологических зависимостей работы концентратора KC-CVD6 с помощью метода группового учёта аргументов / В. В. Пелих, В. М. Салов, А. Е. Бурдонов, Н. Д. Лукьянов // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2020. – Т. 331. – № 2. – С. 139-150.
21. Муравина, О. М. Применение метода группового учёта аргументов для анализа петрофизических данных / О. М. Муравина, И. А. Пономаренко, М. В. Минц // Вестник Камчатской региональной организации Учебно-научный центр. Серия: Науки о Земле. – 2021. – № 3. – С. 5-15.
22. Kumar, S., Haq, M. A., Jain, A., Jason, C. A., Moparthi, N. R., Mittal, N., & Alzamil, Z. S. (2023). Multilayer Neural Network Based Speech Emotion Recognition for Smart Assistance. Computers, Materials & Continua, 75 (1).
23. Nguyen, P. M., & Pham, H. T. (2023). A rigorous framework for the mean field limit of multilayer neural networks. Mathematical Statistics and Learning, 6 (3), 201-357.
24. Носков, С. И. Идентификация параметров простой формы вложенной кусочно-линейной регрессии / С. И. Носков // Учёные записки Комсомольского-на-Амуре государственного технического университета. Науки о природе и технике. – 2023. – № III (67). – С. 57-61.
25. Базилевский, М. П. Оценивание неизвестных параметров многослойной модульной регрессии методом наименьших модулей / М. П. Базилевский // Моделирование, оптимизация и информационные технологии. – 2024. – Т. 12. – № 2 (45). – URL: https://moitvivt.ru/ru/journal/pdf?id=1581 (дата обращения: 20.12.2024). – Текст: электронный.