

Математическое моделирование и его применение в научных исследованиях

УДК 519.862.6

DOI:10.25729/ESI.2024.36.4.005

Оценивание методом наименьших квадратов параметров неэлементарных линейных регрессий с равномерно квантованными объясняющими переменными

Базилевский Михаил Павлович

Иркутский государственный университет путей сообщения,

Россия, Иркутск, *mik2178@yandex.ru*

Аннотация. Статья посвящена разработке новой структурной спецификации регрессионных моделей. Ранее автором были введены в рассмотрение неэлементарные линейные регрессии, в которых объясняющие переменные преобразуются с помощью таких неэлементарных операций, как минимум, максимум и модуль. В данной статье для преобразования объясняющих переменных в регрессионной модели предлагается использовать операции округления их значений до ближайшего целого числа в меньшую сторону (пол) или в большую сторону (потолок). В математике и цифровой обработке сигналов такой процесс преобразования называется квантованием. Рассмотрен известный равномерный квантователь с границей округления 0,5. Предложена неэлементарная линейная регрессия с квантованными объясняющими переменными. Определены области возможных значений шагов квантования для модели с одной объясняющей переменной. На основе этого разработан алгоритм приближенного оценивания с помощью метода наименьших квадратов параметров предложенной структурной спецификации. С использованием искусственно сгенерированных статистических данных в пакете Gretl проведены вычислительные эксперименты, подтвердившие корректность приведенных математических рассуждений. Все полученные в ходе экспериментов неэлементарные линейные регрессии с квантованными переменными оказались адекватнее классических линейных регрессий.

Ключевые слова: регрессионный анализ, неэлементарная линейная регрессия, округление, пол, потолок, квантование, квантователь, метод наименьших квадратов, мультиколлинеарность

Цитирование: Базилевский М.П. Оценивание методом наименьших квадратов параметров неэлементарных линейных регрессий с равномерно квантованными объясняющими переменными / М.П. Базилевский // Информационные и математические технологии в науке и управлении, 2024. – № 4(36). – С. 46-56. – DOI:10.25729/ESI.2024.36.4.005.

Введение. Для выявления скрытых знаний в статистических данных в настоящее время весьма эффективны методы регрессионного анализа [1]. Эта область машинного обучения активно развивается в последние годы и не ограничивается только лишь оцениванием параметров, известных широкому кругу специалистов линейных регрессионных моделей методом наименьших квадратов (МНК). Так, например, в [2] предложен подход к построению нечетких регрессионных моделей, в [3] разработан эффективный алгоритм динамического оценивания параметров регрессий с помощью метода наименьших модулей, в [4] предложен метод, устойчивый к наличию экстремальной стационарной координатной помехи, в [5] сформулирован алгоритм оценки модели полиномиальной логит-регрессии.

Реальные статистические зависимости между выходной и входными переменными редко носят линейный характер, поэтому актуальной задачей является поиск новых, вполне интерпретируемых нелинейных математических форм связи между исследуемыми факторами. Такие формы иначе называют структурными спецификациями регрессионных моделей. Среди таких новых спецификаций хотелось бы выделить неэлементарные линейные регрессии [6], содержащие в уравнении модели бинарные операции \min и \max . Позднее эти регрессии были обобщены [7] с использованием тернарных, кватернарных, ..., l -арных операций \min и \max .

Частным случаем такой спецификации можно считать производственную функцию Леонтьева. В [8] введены модульные линейные регрессии, в которых объясняющие переменные преобразуются с помощью операции модуль, и предложен алгоритм точного оценивания их параметров методом наименьших модулей, а в [9] – алгоритм приближенного МНК-оценивания.

Операции \min , \max и модуль носят дискретный характер. К ним же относится операция «антье» $[x]$ – целая часть вещественного числа x . Цель данной статьи состоит в создании новой спецификации регрессионных моделей, включающей операции «антье», а также алгоритма оценивания её параметров с помощью МНК.

1. Регрессии с квантованными объясняющими переменными. Впервые операцию $[x]$ ввёл К. Ф. Гаусс более 200 лет назад. Как отмечено в [10], эта операция округляет x до ближайшего целого числа в меньшую сторону. Например, $[6.2] = 6$, $[6.5] = 6$, $[6.9] = 6$. Операции округления x до ближайшего целого в большую сторону не существовало, пока в начале 60-х годов К. Э. Айверсон [11] не ввел названия «пол» (floor) и «потолок» (ceiling). Эти операции обозначаются $\lfloor x \rfloor$ и $\lceil x \rceil$ соответственно и строго определяются по следующим правилам:

$$\lfloor x \rfloor = \max \{n \mid n \leq x, n - \text{целое}\}, \lceil x \rceil = \min \{n \mid n \geq x, n - \text{целое}\}.$$

Таким образом, операция пол $\lfloor x \rfloor$ – округление x до ближайшего целого числа в меньшую сторону, т.е. $\lfloor x \rfloor = \lfloor x \rfloor$, а операция потолок $\lceil x \rceil$ – округление в большую сторону. Подробное описание свойств этих функций можно найти в работах [10, 12].

В этой связи функция округления x до ближайшего целого числа может быть записана в виде:

$$f(x) = \lfloor x + 0,5 \rfloor. \quad (1)$$

Например, $f(6,2) = \lfloor 6,2 + 0,5 \rfloor = \lfloor 6,7 \rfloor = 6$, $f(6,9) = \lfloor 6,9 + 0,5 \rfloor = \lfloor 7,4 \rfloor = 7$.

Обобщением функции (1) является так называемый равномерный квантователь (quantizer):

$$f(x) = k \left\lfloor \frac{x}{k} + 0,5 \right\rfloor, \quad (2)$$

где $k > 0$ – шаг квантования.

Тот же самый квантователь (2) можно записать через операцию потолок:

$$f(x) = k \left\lceil \frac{x}{k} - 0,5 \right\rceil.$$

Квантование в математике и цифровой обработке сигналов [13, 14] – процесс преобразования входных значений в выходные в меньшем объеме. Разница между входным значением и его квантованным значением называется ошибкой квантования.

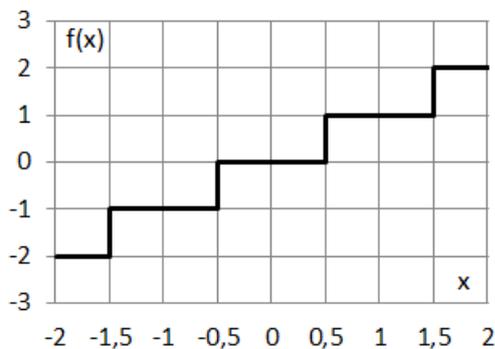
Графики квантователя (2) при $k = 1$ и при $k = 0,2$ на промежутке $x \in [-2, 2]$ представлены на рис. 1 (а), (б).

Как видно по рис. 1, при $k = 1$ расстояние между соседними квантованными значениями (уровнями) равно 1, а при $k = 0,2$ соответственно 0,2. При $k \rightarrow 0$ ступенчатые графики на рис. 1 трансформируются в обычные прямые линии.

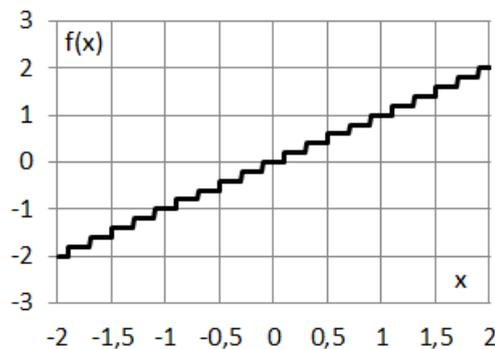
Функции (2) в настоящее время активно применяются при квантовании нейронных сетей (см., например, [15]). Интересная идея обсуждается в работе [16], в которой авторы предлагают следующее обобщение функции (2):

$$f(x) = k \left\lfloor \frac{x}{k} + b \right\rfloor, \quad (3)$$

где b – граница округления.



(a) $k=1$



(б) $k=0,2$

Рис. 1. Графики функции (2)

В [17] проведены теоретические исследования методов глубокого обучения нейронных сетей с использованием квантователя (2).

Используя квантователь (3), введем в рассмотрение неэлементарные линейные регрессии (НЛР) с квантованными объясняющими переменными:

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j k_j \left\lfloor \frac{x_{ij}}{k_j} + b_j \right\rfloor + \varepsilon_i, \quad i = \overline{1, n}, \quad (4)$$

где y_i – наблюдаемое значение выходной переменной y под номером i ; x_{ij} – наблюдаемое значение входной переменной x_j под номером i ; n – общее количество наблюдаемых значений; l – общее количество входных переменных; $\alpha_j, j = \overline{0, l}$ – неизвестные параметры; $k_j > 0$ – неизвестный шаг квантования входной переменной x_j ; b_j – неизвестная граница округления входной переменной x_j ; ε_i – ошибка аппроксимации в i -м наблюдении.

Модель (4), как и все известные НЛР, относится к нелинейным по оцениваемым параметрам. Несколько упростим её. Будем рассматривать НЛР (4) с границами округления $b_j = 0,5, j = \overline{1, l}$, т.е. НЛР, в которой квантование переменных подразумевает округление до ближайшего целого значения:

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j k_j \left\lfloor \frac{x_{ij}}{k_j} + 0,5 \right\rfloor + \varepsilon_i, \quad i = \overline{1, n}. \quad (5)$$

Шаги квантования $k_j, j = \overline{1, l}$ в НЛР (5) по определению положительны. Определимся с областью возможных значений шагов квантования. Сделаем это сначала для модели с одной объясняющей переменной вида:

$$y_i = \alpha_0 + \alpha_1 k_1 \left\lfloor \frac{x_{i1}}{k_1} + 0,5 \right\rfloor + \varepsilon_i, \quad i = \overline{1, n}. \quad (6)$$

Если в НЛР (6) операция $\left\lfloor \frac{x_{i1}}{k_1} + 0,5 \right\rfloor$ будет принимать одно и то же значение $\theta \in \mathbb{Z}$ для любого наблюдения i , то из-за вхождения в модель свободного члена α_0 возникнет совершенная мультиколлинеарность, из-за которой становится невозможно единственным образом идентифицировать МНК-оценки. Определим возможные значения шага квантования k_1 , при которых в регрессии (6) возникает совершенная мультиколлинеарность. Известно [10], что $\lfloor x \rfloor = \theta$ тогда и только тогда, когда

$$\theta \leq x < \theta + 1.$$

Тогда совершенная мультиколлинеарность в НЛР (6) возникает тогда, когда шаг квантования k_1 удовлетворяет системе линейных неравенств

$$\theta \leq \frac{x_{i1}}{k_1} + 0,5 < \theta + 1, \quad i = \overline{1, n}, \quad k_1 > 0, \quad \theta \in \mathbb{Z}. \quad (7)$$

Пусть переменная x_1 принимает как положительные, так и отрицательные значения.

Очевидно, что при бесконечно больших значениях шага квантования k_1 , операция $\left\lfloor \frac{x_{i1}}{k_1} + 0,5 \right\rfloor$ будет равна нулю для любого i . Определим, начиная с какого значения k_1 это будет происходить. Для этого подставим в систему (7) $\theta = 0$. Получим

$$-0,5k_1 \leq x_{i1} < 0,5k_1, \quad i = \overline{1, n}, \quad k_1 > 0. \quad (8)$$

Очевидно, что всегда найдется число k_1 такое, что множество решений системы (8) не пусто. Из (8) следует, что операция $\left\lfloor \frac{x_{i1}}{k_1} + 0,5 \right\rfloor$ в любом наблюдении обращается в нуль, если

$$k_1 > \max \{ |2x_{11}|, |2x_{21}|, \dots, |2x_{n1}| \}. \quad (9)$$

Заметим, что неравенство (9) справедливо тогда, когда в операции $\max \{ |2x_{11}|, |2x_{21}|, \dots, |2x_{n1}| \}$ сработала компонента с положительным наблюдением переменной x_1 , иначе строгий знак неравенства в (9) нужно заменить на нестрогий.

Таким образом, верхнее ограничение на шаг квантования k_1 для НЛР (6), как следует из вышесказанного, имеет вид

$$k_1 \leq \max \{ |2x_{11}|, |2x_{21}|, \dots, |2x_{n1}| \} \quad \text{или} \quad k_1 < \max \{ |2x_{11}|, |2x_{21}|, \dots, |2x_{n1}| \}.$$

Далее будем использовать только первое неравенство с нестрогим знаком.

К сожалению, внутри промежутка $0 < k_1 \leq \max \{ |2x_{11}|, |2x_{21}|, \dots, |2x_{n1}| \}$ могут быть точки, удовлетворяющие системе (7), т.е. точки, в которых в модели (6) возникает совершенная мультиколлинеарность. Рассмотрим возможные случаи.

Случай № 1. Среди значений переменной x_1 есть как положительные, так и отрицательные. В таком случае множество решений системы (7) пусто для любого $\theta \neq 0$, поэтому внутри промежутка $0 < k_1 \leq \max \{ |2x_{11}|, |2x_{21}|, \dots, |2x_{n1}| \}$ отсутствуют точки с совершенной мультиколлинеарностью. Сюда же можно отнести случай, когда среди значений переменной x_1 есть хотя бы 1 нуль.

Случай № 2. Все значения переменной x_1 положительны. В этом случае система (7) может иметь решения при $\theta > 0$, которые можно записать в виде:

$$\max \left\{ \frac{x_{11}}{\theta + 0,5}, \dots, \frac{x_{n1}}{\theta + 0,5} \right\} < k_1 \leq \min \left\{ \frac{x_{11}}{\theta - 0,5}, \dots, \frac{x_{n1}}{\theta - 0,5} \right\}, \quad \theta > 0, \quad \theta \in \square. \quad (10)$$

Поэтому из промежутка $0 < k_1 \leq \max \{|2x_{11}|, |2x_{21}|, \dots, |2x_{n1}|\}$ нужно исключить точки, удовлетворяющие ограничениям (10).

Случай № 3. Все значения переменной x_1 отрицательны. В этом случае система (7) при $\theta < 0$ может иметь следующие решения:

$$\max \left\{ \frac{x_{11}}{\theta + 0,5}, \dots, \frac{x_{n1}}{\theta + 0,5} \right\} \leq k_1 < \min \left\{ \frac{x_{11}}{\theta - 0,5}, \dots, \frac{x_{n1}}{\theta - 0,5} \right\}, \quad \theta < 0, \quad \theta \in \square. \quad (11)$$

Следовательно, из промежутка $0 < k_1 \leq \max \{|2x_{11}|, |2x_{21}|, \dots, |2x_{n1}|\}$ нужно исключить точки, удовлетворяющие ограничениям (11).

В зависимости от конкретной ситуации, исходя из вышеуказанного, можно определить область возможных значений шага квантования k_1 в модели (6). Тогда для любой точки из этой области легко вычисляются МНК-оценки параметров α_0 и α_1 регрессии. Поэтому, взяв некоторое количество точек из области возможных значений параметра k_1 , оценив в каждой из них линейную относительно параметров α_0 и α_1 модель, и выбрав регрессию с минимальной величиной суммы квадратов ошибок, получим близкие к оптимальным оценки НЛР (6). Мера такой близости зависит от величины шага разбиения области значений параметра k_1 : чем меньше шаг, тем ближе оценки к оптимальным, но тем выше вычислительная сложность задачи. Таким образом, на основании вышеизложенного можно сформулировать следующий алгоритм приближенного МНК-оценивания параметров НЛР (5).

Шаг 1. Сформировать области возможных значений шагов квантования входных переменных. При этом:

- если среди значений j -й переменной есть как положительные, так и отрицательные, либо хотя бы одно нулевое значение, то область имеет вид

$$0 < k_j \leq \max \{|2x_{1j}|, |2x_{2j}|, \dots, |2x_{nj}|\}; \quad (12)$$

- если все значения j -й переменной положительны, то ограничение (12) требуется дополнить условиями

$$\max \left\{ \frac{x_{1j}}{\theta + 0,5}, \dots, \frac{x_{nj}}{\theta + 0,5} \right\} < k_j \leq \min \left\{ \frac{x_{1j}}{\theta - 0,5}, \dots, \frac{x_{nj}}{\theta - 0,5} \right\}, \quad \theta > 0, \quad \theta \in \square; \quad (13)$$

- если все значения j -й переменной отрицательны, то ограничение (12) следует дополнить условиями

$$\max \left\{ \frac{x_{1j}}{\theta + 0,5}, \dots, \frac{x_{nj}}{\theta + 0,5} \right\} \leq k_j < \min \left\{ \frac{x_{1j}}{\theta - 0,5}, \dots, \frac{x_{nj}}{\theta - 0,5} \right\}, \quad \theta < 0, \quad \theta \in \square.$$

Шаг 2. Выбрать в каждой области некоторое количество точек.

Шаг 3. Используя выбранные точки вместо шагов квантования переменных в (5), оценить с помощью МНК параметры всех возможных линейных регрессий (5).

Шаг 4. Выбрать лучшую модель, например, по величине коэффициента детерминации R^2 .

2. Вычислительные эксперименты. Вычислительные эксперименты проводились с использованием искусственно сгенерированных статистических данных, представленных в первых трех столбцах таблицы 1. Для оценивания параметров НЛР по предложенному алгоритму в эконометрическом пакете Gretl был написан специальный скрипт на языке hansl. При этом для округления значений использована встроенная функция floor.

Эксперимент № 1. Сначала по этим данным с помощью МНК была построена парная линейная регрессия вида:

$$\tilde{y} = -1,3139 + 1,4491x_1, \quad (14)$$

для которой $R^2 = 0,3994$.

Таблица 1. Статистические данные

y	x_1	x_2	$2,6 \left\lfloor \frac{x_1}{2,6} + \frac{1}{2} \right\rfloor$	$4,05 \left\lfloor \frac{x_2}{4,05} + \frac{1}{2} \right\rfloor$	$3,6 \left\lfloor \frac{x_1}{3,6} + \frac{1}{2} \right\rfloor$	$2,1 \left\lfloor \frac{x_2}{2,1} + \frac{1}{2} \right\rfloor$
2,61	7,1	6,33	7,8	8,1	7,2	6,3
12,96	7,3	15,145	7,8	16,2	7,2	14,7
4,15	4,8	9,462	5,2	8,1	3,6	10,5
13,79	7,1	11,902	7,8	12,15	7,2	12,6
8,55	5,9	9,523	5,2	8,1	7,2	10,5
5,11	5,3	12,636	5,2	12,15	3,6	12,6
14,26	11,3	10,375	10,4	12,15	10,8	10,5
1,27	2,9	6,016	2,6	4,05	3,6	6,3
6,34	8,3	6,117	7,8	8,1	7,2	6,3
14,33	6,6	14,366	7,8	16,2	7,2	14,7

Затем с помощью разработанного алгоритма оценивались параметры НЛР (6) с квантованной переменной x_1 . Поскольку все значения этой переменной положительны, то на первом шаге алгоритма область возможных значений шага квантования k_1 определялась из неравенств (12), (13). Перебирая различные комбинации переменной θ в (13), оказалось, что эта система неравенств решений не имеет. Тогда область возможных значений шага квантования

$$0 < k_1 \leq 22,6. \quad (15)$$

На втором шаге алгоритма на промежутке (15) равномерно были выбраны точки (начальное значение – 0,01, шаг – 0,01). На третьем шаге с помощью МНК оценивались параметры уже линейных регрессий (6). На рис. 2 представлен график зависимости полученных коэффициентов детерминации R^2 оцененных регрессий от шага квантования k_1 . Заметим, что промежуток (15) был взят несколько шире.

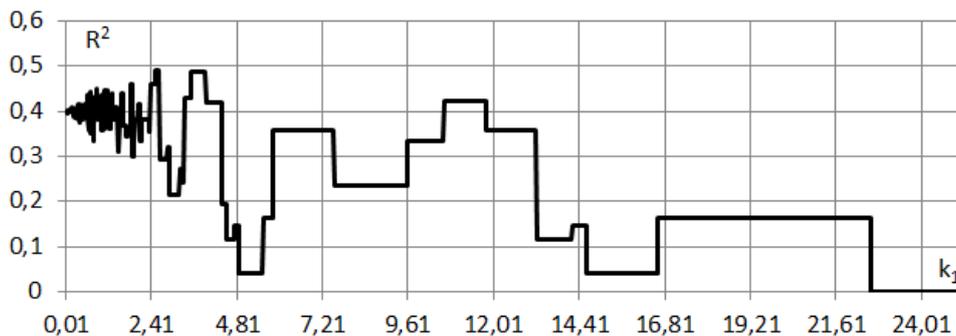


Рис. 2. Зависимость R^2 от k_1

По рис. 2 видно, что когда шаг квантования k_1 мал, то коэффициент детерминации НЛР (6) близок к коэффициенту линейной регрессии (14). На промежутке $0 < k_1 \leq 2,4$ частота изменения R^2 очень высока, а, начиная со значения 2,4, она постепенно снижается. Амплитуда же, наоборот, сначала низкая, а потом постепенно разрастается. При $k_1 > 22,6$ коэффициент детерминации равен 0, поскольку возникает совершенная мультиколлинеарность. Ни в каких других точках он в нуль не обращается. Полученные результаты подтверждают корректность вышеописанных математических рассуждений. Своё наибольшее значение 0,49021 коэффициент R^2 достигает на промежутке $2,52 \leq k_1 \leq 2,63$. При $k_1 = 2,6$ НЛР (6) имеет вид:

$$\tilde{y} = -2,7366 + 1,6381 \left[\frac{x_1}{2,6} + 0,5 \right].$$

Значения квантователя $2,6 \left[\frac{x_1}{2,6} + 0,5 \right]$ представлены в четвертом столбце табл. 1. Тем самым, при квантовании переменной x_1 удалось повысить качество линейной регрессии (14) по коэффициенту детерминации с 0,3994 до 0,49021.

Эксперимент № 2. Аналогичным образом строились зависимости y от x_2 . Оцененная с помощью МНК парная линейная регрессия имеет вид

$$\tilde{y} = -3,092 + 1,1219x_2, \tag{16}$$

а её коэффициент $R^2 = 0,5358$.

Область возможных значений шага квантования k_2 также определялась из неравенств (12), (13). Но в этот раз оказалось, что система (13) имеет решения

$$\theta = 1, \quad 10,097 < k_2 \leq 12,032.$$

Поэтому область возможных значений шага квантования

$$k_2 \in (0, 10,097] \cup (12,032, 30,29). \tag{17}$$

На промежутке (17) равномерно были выбраны точки (начальное значение – 0,01, шаг – 0,01). График зависимости полученных коэффициентов детерминации R^2 оцененных линейных регрессий от шага квантования k_2 представлен на рис. 3.

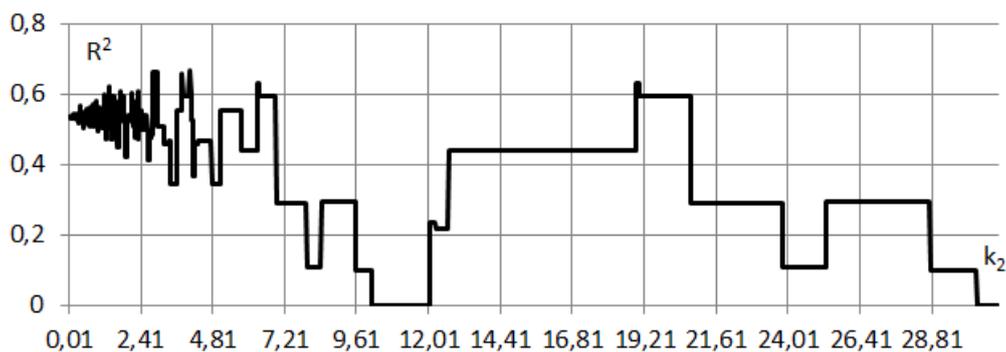


Рис. 3. Зависимость R^2 от k_2

По рис. 3 видно, что коэффициент детерминации оказался равен 0 в точках, не принадлежащих промежутку (17). Полученные результаты вновь согласуются с описанной выше математической теорией. Своё наибольшее значение 0,66779 коэффициент R^2 достигает на промежутке $4,02 \leq k_2 \leq 4,07$. При $k_2 = 4,05$ НЛР имеет вид:

$$\tilde{y} = -2,9445 + 1,0714 \left\lfloor \frac{x_2}{4,05} + 0,5 \right\rfloor.$$

Значения квантователя $4,05 \left\lfloor \frac{x_2}{4,05} + 0,5 \right\rfloor$ представлены в пятом столбце табл. 1. Тем самым, при квантовании переменной x_2 удалось повысить качество линейной регрессии (16) по коэффициенту детерминации с 0,5358 до 0,6678.

Эксперимент № 3. Строились зависимости y от x_1 и x_2 . Сначала была получена модель множественной линейной регрессии вида:

$$\tilde{y} = -9,9341 + 1,2194x_1 + 0,9963x_2, \quad (18)$$

для которой $R^2 = 0,811975$.

Для построения НЛР (5) с квантованными переменными x_1 и x_2 были использованы те же области возможных шагов квантования (15) и (17). На этих промежутках равномерно были выбраны точки (начальное значение – 0,1, шаг – 0,1). Решив переборную задачу, была получена следующая НЛР:

$$\tilde{y} = -11,679 + 1,406 \left\lfloor \frac{x_1}{3,6} + 0,5 \right\rfloor + 1,038 \left\lfloor \frac{x_2}{2,1} + 0,5 \right\rfloor,$$

для которой $R^2 = 0,924549$.

Значения квантователей $3,6 \left\lfloor \frac{x_1}{3,6} + 0,5 \right\rfloor$ и $2,1 \left\lfloor \frac{x_2}{2,1} + 0,5 \right\rfloor$ представлены в шестом и седьмом столбцах табл. 1. Как видно, при квантовании переменных x_1 и x_2 удалось повысить качество линейной регрессии (18) по коэффициенту детерминации с 0,811975 до 0,924549.

Заключение. В статье предложены и исследованы неэлементарные линейные регрессии с квантованными объясняющими переменными. Результаты исследований позволили разработать алгоритм приближенного МНК-оценивания параметров предложенных моделей. Применение разработанного алгоритма при обработке искусственно сгенерированных статистических данных показало, что при квантовании объясняющих переменных качество регрессионных моделей может ощутимо улучшаться.

К сожалению, разработанный и описанный в статье алгоритм не гарантирует нахождения оптимальных МНК-оценок и шагов квантования переменных в неэлементарной линейной регрессии. Однако области возможных значений неизвестных параметров удалось идентифицировать однозначно, поэтому при разбиении этих областей как можно большим количеством точек будет получаться решение, мало отличающееся от оптимального. Таким образом, разработанный в статье математический и алгоритмический аппарат можно успешно применять для решения реальных прикладных задач анализа данных в различных предметных областях. Для этого, в первую очередь, возникает необходимость в разработке вместо скрипта пакета Gretl специализированного программного обеспечения, имеющего интерфейс. Научный интерес вызывает проведение сравнительного анализа качества решения конкретных задач моделирования с использованием разработанных ранее неэлементарных линейных регрессий (с операциями минимум, максимум и модуль) и новых спецификаций (с операциями пол и потолок).

Список источников

1. Arkes J. Regression analysis: a practical introduction. London, Routledge, 2023, 412 p.
2. Исмагилов И.И. Классификация регрессионных моделей и метод построения линейных нечётких регрессий / И.И. Исмагилов, Г. Алсаид // Инновационная экономика: информация, аналитика, прогнозы, 2023. – № 2. – С. 130-138.
3. Голованов А.А. Регрессионный анализ данных на основе метода наименьших модулей в динамических задачах оценивания / А.А. Голованов, А.Н. Тырсин // Заводская лаборатория. Диагностика материалов, 2023. – Т. 89. – № 5. – С. 71-80.
4. Бочаров Д.А. Метод линейной регрессии, устойчивый к экстремальным стационарным помехам / Д.А. Бочаров // Сенсорные системы, 2020. – Т. 34. – № 1. – С. 44-56.
5. Farghali R.A., Qasim M., Kibria B.G., Abonazel M.R. Generalized two-parameter estimators in the multinomial logit regression model: methods, simulation and application. Communications in statistics-simulation and computation, 2023, vol. 52, no. 7, pp. 3327-3342, DOI:10.1080/03610918.2021.1934023.
6. Базилевский М.П. Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования / М.П. Базилевский // Проблемы управления, 2022. – № 4. – С. 3-14.
7. Базилевский М.П. Обобщение неэлементарных линейных регрессий / М.П. Базилевский // Моделирование и анализ данных, 2023. – Т. 13. – № 2. – С. 85-98.
8. Базилевский М.П. Оценивание модульных линейных регрессионных моделей с помощью метода наименьших модулей / М.П. Базилевский, А.Б. Ойдопова // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления, 2023. – № 45. – С. 130-146.
9. 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.
10. Грэхем Р. Конкретная математика. Основание информатики: пер. с англ. / Р. Грэхем, Д. Кнут, О. Паташник. – М.: Мир, 1998. – 703 с.
11. Iverson K.E. A programming language. Proceedings of the May 1-3: Spring joint computer conference, 1962, pp. 345-351.
12. Трапезников С.С. Целая и дробная часть числа / С.С. Трапезников, Г.С. Изергина // Краевая научно-практическая конференция «Прикладные и фундаментальные вопросы математики». – Пермь, 2014. – 30 с.
13. Марьев А.А. Методы и устройства цифровой обработки сигналов. Дискретизация. Квантование. Цифровой анализ сигналов / А.А. Марьев. – Ростов-на-Дону – Таганрог: Изд-во ЮФУ, 2020. – 132 с.
14. Малыгин И.Г. Высокоскоростной алгоритм передачи видеoinформации о чрезвычайных ситуациях на объектах транспорта / И.Г. Малыгин, О.А. Королев // Морские интеллектуальные технологии, 2021. – Т. 1. – № 1. – С. 64-70.
15. Louizos C., Reisser M., Blankevoort T. et al. Relaxed quantization for discretized neural networks. arXiv preprint, 2018, available at: <https://arxiv.org/pdf/1810.01875>(accessed: 05/12/2024).
16. Li Z., Guo C., Zhu Z. et al. Efficient adaptive activation rounding for post-training quantization. arXiv preprint, 2022, available at: <https://arxiv.org/pdf/2208.11945> (accessed: 05/12/2024).
17. Li H., De S., Xu Z. et al. Training quantized nets: a deeper understanding. Advances in neural information processing systems, 2017, vol. 30, available at: <https://proceedings.neurips.cc/paper/2017/file/1c303b0eed3133200cf715285011b4e4-Paper.pdf> (accessed: 05/12/2024).

Базилевский Михаил Павлович. *Доцент, к.т.н., Иркутский государственный университет путей сообщения, AuthorID:679277, SPIN:4347-5028, ORCID:0000-0002-3253-5697, mik2178@yandex.ru, Россия, Иркутск, Чернышевского д. 15.*

UDC 519.862.6

DOI:10.25729/ESI.2024.36.4.005

Ordinary least squares estimation of non-elementary linear regressions parameters with uniformly quantized explanatory variables

Mikhail P. Bazilevskiy

Irkutsk state transport university,

Russia, Irkutsk, *mik2178@yandex.ru*

Abstract. This article is devoted to the development of a new structural specification of regression models. Previously, the author introduced non-elementary linear regressions, in which explanatory variables are transformed using such non-elementary operations as minimum, maximum and modulus. In this article, to transform explanatory variables in a regression model, it is proposed to use the operations of rounding their values to the nearest integer downwards (floor) or up (ceiling). In mathematics and digital signal processing, this conversion process is called quantization. The well-known uniform quantizer with a rounding boundary of 0.5 is considered. A non-elementary linear regression with quantized explanatory variables is proposed. The ranges of possible values of quantization steps size for a model with one explanatory variable are determined. Based on this, an algorithm has been developed for approximate estimation using the ordinary least squares method of the proposed structural specification parameters. Using artificially generated statistical data in the Gretl package, computational experiments were carried out that confirmed the correctness of the above mathematical reasoning. All non-elementary linear regressions with quantized variables obtained during the experiments turned out to be more adequate than classical linear regressions.

Keywords: regression analysis, non-elementary linear regression, rounding, floor, ceiling, quantization, quantizer, ordinary least squares method, multicollinearity.

References

1. Arkes J. Regression analysis: a practical introduction. London: Routledge, 2023, 412 p.
2. Ismagilov I.I., Alsaied Gh. Klassifikatsiya regressionnykh modeley i metod postroeniya lineynykh nechetkikh regressiy [Classification of regression models and a method for constructing linear fuzzy regressions]. Innovatsionnaya ekonomika: informatsiya, analitika, prognozy [Innovative economy: information, analysis, prognoses], 2023, no. 2, pp. 130-138.
3. Golovanov A.A., Tyrsin A.N. Regressiionnyy analiz dannykh na osnove metoda naimen'shikh moduley v dinamicheskikh zadachakh otsenivaniya [Regression analysis of data based on the method of least absolute deviations in dynamic estimation problems]. Zavodskaya laboratoriya. Diagnostika materialov [Industrial laboratory. Diagnostics of materials], 2023, vol. 89, no. 5, pp. 71-80.
4. Bocharov D.A. Metod lineynoy regressii, ustoychivyy k ekstremal'nym stacionarnym pomekham [A linear regression method robust to extreme stationary clutter]. Sensornye sistemy [Sensory Systems], 2020, vol. 34, no. 1, pp. 44-56.
5. Farghali R.A., Qasim M., Kibria B.G., Abonazel M.R. Generalized two-parameter estimators in the multinomial logit regression model: methods, simulation and application. Communications in statistics-simulation and computation, 2023, vol. 52, no. 7, pp. 3327-3342, DOI:10.1080/03610918.2021.1934023.
6. Bazilevskiy M.P. Metod postroeniya neelementarnykh lineynykh regressiy na osnove apparata matematicheskogo programmirovaniya [A method for constructing nonelementary linear regressions based on mathematical programming]. Problemy upravleniya [Control sciences], 2022, no. 4, pp. 3-14.
7. Bazilevskiy M.P. Obobshchenie neelementarnykh lineynykh regressiy [Generalization of non-elementary linear regressions]. Modelirovanie i analiz dannykh [Modeling and Data Analysis], 2023, vol. 13, no. 2, pp. 85-98.
8. Bazilevskiy M.P., Oydopova A.B. Otsenivanie modul'nykh lineynykh regressiionnykh modeley s pomoshch'yu metoda naimen'shikh moduley [Estimation of modular linear regression models using the least absolute deviations]. Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniya [Bulletin of Perm national research polytechnic university. Electrotechnics, informational technologies, control systems], 2023, no. 45, pp. 130-146.
9. Bazilevskiy M.P. Programmnoe obespechenie dlya otsenivaniya modul'nykh lineynykh regressiy [Software for estimating modular linear regressions]. Informatsionnye i matematicheskie tekhnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2023, no. 3 (31), pp. 136-146.
10. Graham R.L., Knuth D.E., Patashnik O. Konkretnaya matematika. Osnovanie informatiki [Concrete mathematics. A foundation for computer science]. Moscow, Mir [Worls], 1998, 703 p.

11. Iverson K.E. A programming language. Proceedings of the May 1-3: Spring joint computer conference, 1962, pp. 345-351.
12. Trapeznikov S.S., Izergina G.S. Tselaya i drobnaya chast' chisla [Integer and fractional part of a number]. Kraevaya nauchno-prakticheskaya konferentsiya «Prikladnye i fundamental'nye voprosy matematiki» [Regional scientific and practical conference «Applied and fundamental issues of mathematics»], Perm, 2014, 30 p.
13. Mar'ev A.A. Metody i ustroystva tsifrovoy obrabotki signalov. Diskretizatsiya. Kvantovanie. Tsifrovoy analiz signalov [Methods and devices of digital signal processing. Sampling. Quantization. Digital signal analysis]. Rostov-na-Donu – Taganrog, YuFU Publ., 2020, 132 p.
14. Malygin I.G., Korolev O.A. Vysokoskorostnoy algoritm peredachi videoinformatsii o chrezvychaynykh situatsiyakh na ob"ektakh transporta [High-speed algorithm for transmitting video information about emergency situations on transport objects]. Morskie intellektual'nye tekhnologii [Marine intellectual technologies], 2021, vol. 1, no. 1, pp. 64-70.
15. Louizos C., Reisser M., Blankevoort T. et al. Relaxed quantization for discretized neural networks. arXiv preprint, 2018, available at: <https://arxiv.org/pdf/1810.01875> (accessed: 05/12/2024).
16. Li Z., Guo C., Zhu Z. et al. Efficient adaptive activation rounding for post-training quantization. arXiv preprint, 2022, available at: <https://arxiv.org/pdf/2208.11945> (accessed: 05/12/2024).
17. Li H., De S., Xu Z. et al. Training quantized nets: a deeper understanding. Advances in neural information processing systems, 2017, vol. 30, available at: <https://proceedings.neurips.cc/paper/2017/file/1c303b0eed3133200cf715285011b4e4-Paper.pdf> (accessed: 05/12/2024).

Bazilevskiy Mikhail Pavlovich. Associate professor, candidate of technical sciences, Irkutsk state transport university, AuthorID:679277, SPIN:4347-5028, ORCID:0000-0002-3253-5697, mik2178@yandex.ru, Russia, Irkutsk, Chernyshevskogo St., 15.

Статья поступила в редакцию 13.05.2024; одобрена после рецензирования 27.11.2024; принята к публикации 17.12.2024.

The article was submitted 05/13/2024; approved after reviewing 11/27/2024; accepted for publication 12/17/2024.