УДК 519.862.6

DOI: http://doi.org/10.25728/pu.2021.3.3

ПОСТРОЕНИЕ СТЕПЕННО-ПОКАЗАТЕЛЬНЫХ И ЛИНЕЙНО-ЛОГАРИФМИЧЕСКИХ РЕГРЕССИОННЫХ МОДЕЛЕЙ

М.П. Базилевский

Аннотация. Часто при использовании нелинейных регрессионных моделей оценки полученной зависимости затруднительно или вовсе невозможно интерпретировать. С целью создания таких нелинейных спецификаций регрессий, в которых любому оцененному параметру, за исключением свободного члена, всегда может быть дана какая-либо содержательная интерпретация, разработаны мультипликативная степенно-показательная регрессия, представляющая собой обобщение производственной функции Кобба — Дугласа, и аддитивная линейно-логарифмическая регрессия. Для каждой из них сформулированы три стратегии построения и подробно рассмотрены вопросы интерпретации их оценок. Стратегии построения предложенных моделей с помощью метода наименьших модулей формализованы в виде задач линейного и частично-булевого линейного программирования. Для демонстрации разработанного в настоящей работе математического аппарата решена задача моделирования железнодорожных грузоперевозок в Иркутской области.

Ключевые слова: регрессионная модель, интерпретация, мультипликативная степенно-показательная регрессия, линейно-логарифмическая регрессия, отбор «информативных» регрессоров, метод наименьших модулей, железнодорожные грузоперевозки.

ВВЕДЕНИЕ

Регрессионный анализ представляет собой признанный во всем мире инструмент построения математических моделей статистического типа [1, 2]. Одним из первых и, возможно, самым важным этапом построения регрессионной модели является ее спецификация, т. е. выбор состава переменных и математической формы связи между ними. К настоящему времени разработано значительное количество таких спецификаций, большинство из которых можно найти в работах [3—6]. Самой простой спецификацией является модель множественной линейной регрессии:

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n}, \tag{1}$$

где y_i, $i = \overline{1, n}$, — наблюдаемые значения объясняемой (выходной) переменной y; x_{ij}, $i = \overline{1, n}$, $j = \overline{1, l}$, — наблюдаемые значения объясняющих (входных) переменных $x_1, x_2, ..., x_l$; ε_i, $i = \overline{1, n}$, — ошибки аппроксимации; $\alpha_0, \alpha_1, \alpha_2, ..., \alpha_l$ — неизвестные параметры.

Линейная регрессия (1) легко оценивается, например, с помощью метода наименьших квадратов (МНК). Пусть ее оцененное уравнение имеет вид:

$$\tilde{y} = \tilde{\alpha}_0 + \sum_{j=1}^{l} \tilde{\alpha}_j x_j, \tag{2}$$

где $\tilde{y}$ — рассчитываемое по модели значение объясняемой переменной; $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_l$ — оценки неизвестных параметров.

Коэффициент уравнения (2) $\tilde{\alpha}_s$ при объясняющей переменной x_s интерпретируется так: если значение объясняющей переменной x_s изменится на одну единицу, то значение объясняемой переменной y изменится в среднем на $\tilde{\alpha}_s$ единиц.

Отметим, что разработка новых спецификаций регрессионных моделей продолжается и по сей день. Так, в работе [7] предложена линейно-мультипликативная регрессия (ЛМР), в работе [8] — регрессия, противоположная по смыслу производственной функции Леонтьева, в работе [9] рассмотрен их симбиоз, а в работе [10] — индексная регрессия.

Для решения проблемы спецификации разработана технология организации «конкурса» регрессионных моделей, подробно описанная в монографии [6]. Суть конкурса состоит в формировании множества альтернативных вариантов регрессий и многокритериальном выборе лучшей из них. В работе [6] рассмотрен следующий алгоритм формирования альтернатив. Сначала набор исходных объясняющих переменных расширяется с помощью преобразований, в качестве которых могут выступать, например, элементарные функции $\ln x$, e^x, x^{-1}, x^2, x^3, $\sqrt{x}$ и т. д. Затем полным перебором комбинаций решается задача отбора m информативных регрессоров (ОИР) [11]. К сожалению, полученное таким образом уравнение регрессии может оказаться значительно нелинейным, что затруднит или вовсе сделает невозможной интерпретацию найденных оценок.

Целью настоящей работы является разработка таких нелинейных спецификаций, для которых априори было бы известно, что в результате проведения «конкурса» моделей любому коэффициенту полученной регрессии, за исключением свободного члена, всегда можно будет дать какую-либо содержательную интерпретацию.

1. МУЛЬТИПЛИКАТИВНАЯ СТЕПЕННО-ПОКАЗАТЕЛЬНАЯ РЕГРЕССИЯ

Однофакторная показательная регрессия [12, 13] имеет вид:

$$y_i = \alpha_0 e^{\alpha_1 x_i} \varepsilon_i, \quad i = \overline{1, n}. \qquad (3)$$

Модель (3) является нелинейной по оцениваемым параметрам, но может быть линеаризована с помощью логарифмирования:

$$\ln y_i = c_0 + \alpha_1 x_i + u_i, \quad i = \overline{1, n}, \qquad (4)$$

где $c_0 = \ln\alpha_0$, $u_i = \ln\varepsilon_i$.

Линейную по параметрам модель (4) называют полулогарифмической (левой логарифмической, логарифмически-линейной) регрессией [13].

В работе [13] можно найти такую интерпретацию оцененного коэффициента $\tilde{\alpha}_1$ моделей (3) и (4): если значение объясняющей переменной x изменится на одну единицу, то значение объясняемой переменной y изменится в среднем на $100\tilde{\alpha}_1$ %.

Но, к сожалению, как отмечено в работе [13], такая интерпретация коэффициента $\tilde{\alpha}_1$ моделей (3) и (4) правомерна лишь для малых $\tilde{\alpha}_1$.

Рассмотрим обобщение модели (3) — аддитивную многофакторную показательную регрессию:

$$y_i = a_0 + \sum_{j=1}^{l} a_j e^{\beta_j x_{ij}} + \varepsilon_i, \quad i = \overline{1, n}, \qquad (5)$$

где β_j, $j = \overline{1, l}$ — неизвестные параметры.

Линеаризовать модель (5) не представляется возможным. Но даже если бы были найдены ее оценки, то затруднительно было бы дать им какую-либо содержательную интерпретацию. Поэтому рассмотрим мультипликативную многофакторную показательную регрессию:

$$y_i = \alpha_0 \prod_{j=1}^{l} e^{\alpha_j x_{ij}} \varepsilon_i, \quad i = \overline{1, n}. \qquad (6)$$

Модель (6) линеаризуется с помощью логарифмирования и все ее коэффициенты можно интерпретировать так, как показано выше.

Регрессия (6) по своим свойствам очень напоминает производственную функцию (ПФ) Кобба — Дугласа (степенную регрессию):

$$y_i = \alpha_0 \prod_{j=1}^{l} x_{ij}^{\alpha_j} \varepsilon_i, \quad i = \overline{1, n}. \qquad (7)$$

Модель (7) тоже линеаризуется с помощью логарифмирования, а оцененный коэффициент $\tilde{\alpha}_s$ при объясняющей переменной x_s интерпретируется так: если значение объясняющей переменной x_s изменится на 1 %, то значение объясняемой переменной y изменится в среднем на $\tilde{\alpha}_s$ %. Иными словами, $\tilde{\alpha}_s$ представляет собой коэффициент эластичности переменной y по x_s.

Тогда составим мультипликативную конструкцию моделей (6) и (7):

$$y_i = \alpha_0 \prod_{j=1}^{l} x_{ij}^{\alpha_j} \prod_{j=1}^{l} e^{\beta_j x_{ij}} \varepsilon_i, \quad i = \overline{1, n}. \qquad (8)$$

Будем называть выражение (8) мультипликативной степенно-показательной регрессией (МСПР).

Отметим, что симбиоз степенных и показательных регрессий вводится не впервые. Так, например, в статье [14] рассмотрена модификация ПФ Кобба — Дугласа, в которую труд и капитал входят в виде степенных функций, а научно-техническая информация — в виде показательной. Нельзя также не упомянуть о ПФ Тинбергена [6], представляющей собой произведение степенной регрессии (7) и множителя $e^{\gamma t}$, учитывающего влияние «нейтрального» технического прогресса. Однако МСПР обобщает все эти известные модификации.

Прологарифмированная МСПР (8) имеет вид:

$$\ln y_i = c_0 + \sum_{j=1}^{l} \alpha_j \ln x_{ij} + \sum_{j=1}^{l} \beta_j x_{ij} + u_i,$$
$$i = \overline{1, n}. \qquad (9)$$

Как видно, с оцениванием МСПР проблем нет. Но возникает проблема с интерпретацией ее коэффициентов, поскольку в модель (9) каждая объясняющая переменная входит и линейно, и логарифмически. Поэтому для того, чтобы иметь возможность интерпретировать любой коэффициент МСПР, при ее построении обязательно нужно проводить ОИР.

Для дальнейшего изложения введем булевы переменные:

$$\sigma_j^{\text{ст}} = \begin{cases} 1, \text{если } x_j \text{ входит в МСПР} \\ \quad \text{в степенном виде,} \\ 0, \text{если не входит,} \end{cases}$$

$$\sigma_j^{\text{пок}} = \begin{cases} 1, \text{если } x_j \text{ входит в МСПР} \\ \quad \text{в показательном виде,} \\ 0, \text{если не входит.} \end{cases}$$

Тогда можно ввести линейные ограничения на коэффициенты моделей (8) и (9):

$$-M\sigma_j^{\text{ст}} \le \alpha_j \le M\sigma_j^{\text{ст}}, \quad j = \overline{1, l}, \qquad (10)$$

$$-M\sigma_j^{\text{пок}} \le \beta_j \le M\sigma_j^{\text{пок}}, \quad j = \overline{1, l}, \qquad (11)$$

где M — большое положительное число.

Понятно, что если $\sigma_j^{\text{ст}} = 1$, $\sigma_j^{\text{пок}} = 0, j = \overline{1, l}$, то МСПР (8) трансформируется в степенную регрессию (7), а если $\sigma_j^{\text{ст}} = 0$, $\sigma_j^{\text{пок}} = 1, j = \overline{1, l}$, — в показательную (6).

Сформулируем три стратегии построения МСПР:

- Стратегия 1. Нет никаких ограничений на то, в каком виде переменные входят в модель. В этом случае нужно просто оценить линейную регрессию (9) с $(2l + 1)$ параметром и перейти к МСПР (8). Оцененное уравнение можно использовать для прогнозирования, но интерпретировать коэффициенты невозможно.
- Стратегия 2. Каждая объясняющая переменная входит в модель либо в степенном виде, либо в показательном. На формальном языке эта стратегия имеет вид:

$$\sigma_j^{\text{ст}} + \sigma_j^{\text{пок}} = 1, \quad j = \overline{1, l}. \qquad (12)$$

В этом случае нужно оценить 2^l линейных регрессий (9) с $(l + 1)$ параметром, выбрать из них наилучшую и перейти к МСПР (8). В полученном уравнении гарантированно можно дать интерпретацию любому коэффициенту (если его знак соответствует содержательному смыслу задачи), за исключением, быть может, свободного члена. Также полученное уравнение можно использовать для прогнозирования. Но если число переменных l велико, то возникает задача отбора заданного числа наиболее информативных из них.

- Стратегия 3. Каждая объясняющая переменная входит в модель либо в степенном виде, либо в показательном, а общее количество регрессоров равно m. На формальном языке эта стратегия имеет вид:

$$\sigma_j^{\text{ст}} + \sigma_j^{\text{пок}} \le 1, j = \overline{1, l}, \qquad (13)$$

$$\sum_{j=1}^{l} (\sigma_j^{\text{ст}} + \sigma_j^{\text{пок}}) = m. \qquad (14)$$

В этом случае нужно оценить $C_l^m \cdot 2^m$ линейных регрессий (9) с $(m + 1)$ параметром, выбрать из них наилучшую и перейти к МСПР (8). Полученное уравнение можно использовать как для прогнозирования, так и для интерпретации.

2. ЛИНЕЙНО-ЛОГАРИФМИЧЕСКАЯ РЕГРЕССИЯ

Однофакторная логарифмическая [12] (правая логарифмическая, линейно-логарифмическая) регрессия имеет вид:

$$y_i = \alpha_0 + \alpha_1 \ln x_i + \varepsilon_i, \quad i = \overline{1, n}. \qquad (15)$$

На сайте [15] можно найти такую интерпретацию оцененного коэффициента $\tilde{\alpha}_1$ модели (15): если значение объясняющей переменной x изменится на 1 %, то значение объясняемой переменной y изменится в среднем на $\tilde{\alpha}_1/100$ единиц.

На наш взгляд, оценку $\tilde{\alpha}_1$ логарифмической модели (15) можно интерпретировать и так: если значение объясняющей переменной x изменится в e раз, то значение объясняемой переменной y изменится в среднем на $\tilde{\alpha}_1$ единиц.

Рассмотрим обобщение модели (15) — аддитивную многофакторную логарифмическую регрессию:

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j \ln x_{ij} + \varepsilon_i, \quad i = \overline{1, n}. \qquad (16)$$

Модель (16) является линейной по параметрам и любой ее оцененный коэффициент при логарифме объясняющей переменной можно интерпретировать вышеуказанными способами.

Стоит отметить, что нет смысла использовать в выражении (16) логарифмы с различными основаниями. Рассмотрим, например, двухфакторную модель

$$y_i = \alpha_0 + \alpha_1 \log_2 x_1 + \alpha_2 \log_3 x_2 + \varepsilon_i, \quad i = \overline{1, n}.$$

Используя известное свойство логарифма $\log_a x = \dfrac{\log_c x}{\log_c a}$, эту модель можно представить в виде:

$$y_i = \alpha_0 + \alpha_1 \frac{\ln x_1}{\ln 2} + \alpha_2 \frac{\ln x_2}{\ln 3} + \varepsilon_i, \quad i = \overline{1, n}.$$

Если в полученном выражении переобозначить параметры $\dfrac{\alpha_1}{\ln 2}$ и $\dfrac{\alpha_2}{\ln 3}$, то получим частный случай регрессии (16).

Составим аддитивную конструкцию моделей (1) и (16):

$$y_i = \gamma_0 + \sum_{j=1}^{l} \gamma_j x_{ij} + \sum_{j=1}^{l} \delta_j \ln x_{ij} + \varepsilon_i, \quad i = \overline{1, n}. \qquad (17)$$

Будем называть выражение (17) линейно-логарифмической регрессией (ЛЛР).

При попытке интерпретировать ЛЛР возникает та же проблема, что и при интерпретации МСПР, связанная с тем, что в уравнение (17) каждая объясняющая переменная входит и линейно, и логарифмически.

Введем булевы переменные:

$$\sigma_j^{\text{лин}} = \\ = \begin{cases} 1, \text{ если } x_j \text{ входит в ЛЛР в линейном виде,} \\ 0, \text{ если не входит,} \end{cases}$$

$$\sigma_j^{\text{лог}} = \begin{cases} 1, \text{ если } x_j \text{ входит в ЛЛР} \\ \quad \text{в логарифмическом виде,} \\ 0, \text{ если не входит.} \end{cases}$$

Тогда можно ввести линейные ограничения на коэффициенты модели (17):

$$-M\sigma_j^{\text{лин}} \le \gamma_j \le M\sigma_j^{\text{лин}}, \quad j = \overline{1, l}, \qquad (18)$$

$$-M\sigma_j^{\text{лог}} \le \delta_j \le M\sigma_j^{\text{лог}}, \quad j = \overline{1, l}. \qquad (19)$$

Очевидно, что если $\sigma_j^{\text{лин}} = 1$, $\sigma_j^{\text{лог}} = 0$, $j = \overline{1, l}$, то ЛЛР (17) трансформируется в линейную регрессию (1), а если $\sigma_j^{\text{лин}} = 0$, $\sigma_j^{\text{лог}} = 1$, $j = \overline{1, l}$ — в логарифмическую (16).

По аналогии с МСПР, сформулируем три стратегии построения ЛЛР.

- Стратегия 1. Нет никаких ограничений на то, в каком виде переменные входят в модель. В этом случае нужно просто оценить линейную регрессию (17) с $(2l + 1)$ параметром. Оцененное уравнение нельзя использовать для интерпретации, но можно использовать для прогнозирования.
- Стратегия 2. Каждая объясняющая переменная входит в модель либо в линейном виде, либо в логарифмическом. На формальном языке эта стратегия имеет вид:

$$\sigma_j^{\text{лин}} + \sigma_j^{\text{лог}} = 1, \quad j = \overline{1, l}. \qquad (20)$$

В этом случае нужно оценить 2^l линейных регрессий (17) с $(l + 1)$ параметром и выбрать из них наилучшую. Ее можно использовать как для интерпретации, так и для прогнозирования.

- Стратегия 3. Каждая объясняющая переменная входит в модель либо в линейном виде, либо в логарифмическом, а общее количество регрессоров равно m. На формальном языке эта стратегия имеет вид:

$$\sigma_j^{\text{лин}} + \sigma_j^{\text{лог}} \le 1, \quad j = \overline{1, l}, \qquad (21)$$

$$\sum_{j=1}^{l} (\sigma_j^{\text{лин}} + \sigma_j^{\text{лог}}) = m. \qquad (22)$$

В этом случае нужно оценить $C_l^m \cdot 2^m$ линейных регрессий (17) с $(m + 1)$ параметром и выбрать из них наилучшую.

3. ФОРМАЛИЗАЦИЯ ЗАДАЧИ ПОСТРОЕНИЯ МСПР И ЛЛР В ВИДЕ ЗАДАЧИ ЧАСТИЧНО-БУЛЕВОГО ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Аппарат математического программирования находит широкое применение в регрессионном анализе (см., например, работы [16—18]).

Пусть прологарифмированная МСПР (9) оценивается с помощью метода наименьших модулей (МНМ). Тогда, как показано в монографии [6], МНМ-оценки этой регрессии могут быть найдены в результате решения задачи линейного программирования (ЛП):

$$\lambda_i^+ + \lambda_i^- \to \min, \qquad (23)$$

$$v_i = c_0 + \sum_{j=1}^{l} \alpha_j z_{ij} + \sum_{j=1}^{l} \beta_j x_{ij} + \lambda_i^+ - \lambda_i^-, \quad i = \overline{1, n}, \qquad (24)$$

$$\lambda_i^+, \lambda_i^- \ge 0, \qquad (25)$$

где $v_i = \ln y_i$, $z_{ij} = \ln x_{ij}$,

$$\lambda_i^+ = \begin{cases} v_i - c_0 - \sum_{j=1}^{l} \alpha_j z_{ij} - \sum_{j=1}^{l} \beta_j x_{ij}, \\ \text{если } v_i - c_0 - \sum_{j=1}^{l} \alpha_j z_{ij} - \sum_{j=1}^{l} \beta_j x_{ij} > 0, \\ 0, \text{ в противном случае,} \end{cases}$$

$$\lambda_i^- = \begin{cases} c_0 + \sum_{j=1}^{l} \alpha_j z_{ij} + \sum_{j=1}^{l} \beta_j x_{ij} - v_i, \\ \text{если } c_0 + \sum_{j=1}^{l} \alpha_j z_{ij} + \sum_{j=1}^{l} \beta_j x_{ij} - v_i > 0, \\ 0, \text{ в противном случае.} \end{cases}$$

Тогда в зависимости от выбранной стратегии построения МСПР необходимо решить такие задачи:

- для первой стратегии — задачу ЛП с целевой функцией (23) и с линейными ограничениями (24), (25);
- для второй стратегии — задачу частично-булевого линейного программирования (ЧБЛП) с целевой функцией (23) и с линейными ограничениями (24), (25), (10)—(12);
- для третьей стратегии — задачу ЧБЛП с целевой функцией (23) и с линейными ограничениями (24), (25), (10), (11), (13), (14).

Аналогично формализуется задача построения ЛЛР. При этом МНМ-оценки ЛЛР (17) могут быть найдены на основе решения задачи ЛП:

$$\theta_i^+ + \theta_i^- \to \min, \qquad (26)$$

$$y_i = \gamma_0 + \sum_{j=1}^{l} \gamma_j x_{ij} + \sum_{j=1}^{l} \delta_j z_{ij} + \theta_i^+ - \theta_i^-, \quad i = \overline{1, n}, \qquad (27)$$

$$\theta_i^+, \theta_i^- \ge 0, \qquad (28)$$

где $z_{ij} = \ln x_{ij}$,

$$\theta_i^+ = \begin{cases} y_i - \gamma_0 - \sum_{j=1}^{l} \gamma_j x_{ij} - \sum_{j=1}^{l} \delta_j z_{ij}, \\ \text{если } y_i - \gamma_0 - \sum_{j=1}^{l} \gamma_j x_{ij} - \sum_{j=1}^{l} \delta_j z_{ij} > 0, \\ 0, \text{ в противном случае,} \end{cases}$$

$$\theta_i^- = \begin{cases} \gamma_0 + \sum_{j=1}^{l} \gamma_j x_{ij} + \sum_{j=1}^{l} \delta_j z_{ij} - y_i, \\ \text{если } \gamma_0 + \sum_{j=1}^{l} \gamma_j x_{ij} + \sum_{j=1}^{l} \delta_j z_{ij} - y_i > 0, \\ 0, \text{ в противном случае.} \end{cases}$$

Тогда в зависимости от выбранной стратегии построения ЛЛР необходимо решить такие задачи:

- для первой стратегии — задачу ЛП с целевой функцией (26) и с линейными ограничениями (27), (28);
- для второй стратегии — задачу ЧБЛП с целевой функцией (26) и с линейными ограничениями (27), (28), (18)—(20);
- для третьей стратегии — задачу ЧБЛП с целевой функцией (26) и с линейными ограничениями (27), (28), (18), (19), (21), (22).

4. МОДЕЛИРОВАНИЕ ЖЕЛЕЗНОДОРОЖНЫХ ГРУЗОПЕРЕВОЗОК ИРКУТСКОЙ ОБЛАСТИ

Актуальной на сегодняшний день является задача моделирования объемов железнодорожных перевозок (см., например, работы [19, 20]). Для демонстрации предложенного в настоящей работе математического аппарата решалась задача моделирования железнодорожных грузоперевозок в Иркутской области. Для построения моделей на официальном сайте Федеральной службы государственной статистики были собраны годовые данные за период 2000—2018 гг. по показателям:

— отправление грузов железнодорожным транспортом общего пользования (млн т) y;

— численность рабочей силы (тыс. чел.) x_3;

— валовой региональный продукт (млн руб.) x_{14};

— число предприятий и организаций x_{18};

— объем промышленной продукции (млн руб.);

— производство электроэнергии (млрд кВт/ч) x_{22};

— среднегодовая номинальная начисленная заработная плата работников в области добычи полезных ископаемых (руб.) x_{23};

— среднегодовая номинальная начисленная заработная плата работников в области обрабатывающих производств (руб.) x_{24};

— продукция сельского хозяйства (млн руб.) x_{25};

— среднегодовая номинальная начисленная заработная плата работников сельского хозяйства, охоты и лесного хозяйства (руб.);

— число действующих строительных организаций;

— оборот розничной торговли (млн руб.) x_{31}.

Для построения МСПР и ЛЛР на языке hansl эконометрического пакета Gretl был разработан специальный скрипт.

Сначала по исходным данным осуществлялось построение МСПР по третьей стратегии. Задача решалась методом перебора, оценивание производилось с помощью МНК, число отбираемых регрессоров m задавалось равным 3. В результате перебора $C_{11}^3 \cdot 2^3 = 1320$ альтернатив была выбрана наилучшая по величине коэффициента детерминации R^2. В прологарифмированном виде эта регрессия имеет вид

$$\ln\tilde{y} = -1{,}2502 + \underset{(2{,}972)}{8{,}431 \cdot 10^{-6}}x_{23} - \underset{(-6{,}098)}{3{,}388 \cdot 10^{-5}}x_{25} + \underset{(11{,}13)}{0{,}5176}\ln x_{31}, \qquad (29)$$

а ее коэффициент детерминации $R^2 = 0{,}9334$. В уравнении (29) под коэффициентами при объясняющих переменных указаны значения t-критерия Стьюдента. С их помощью можно сделать вывод, что для уровня значимости $\alpha = 0{,}05$ все коэффициенты значимы.

К сожалению, из-за эффекта мультиколлинеарности исказился знак коэффициента при переменной x_{25}, поэтому попытка интерпретации уравнения (29) приводит к абсурдному выводу: чтобы увеличить объемы железнодорожных перевозок, нужно снижать объемы продукции сельского хозяйства. Этот противоречивый результат указывает на то, что в процессе перебора моделей нужно организовывать проверку соответствия знаков коэффициентов уравнения регрессии содержательному смыслу факторов. Если хотя бы один коэффициент не соответствует смыслу, то такая модель исключается из дальнейшего рассмотрения. Эту известную рекомендацию можно найти в монографии [6]. Поэтому было принято решение перестроить МСПР. Группой экспертов было установлено, что все вышеперечисленные объясняющие переменные должны влиять на y со знаком «+». После корректировки разработанного скрипта и запуска его с теми же настройками оказалось, что из 1320 альтернатив только 64 удовлетворяют содержательному смыслу задачи. Лучшей из них в прологарифмированном виде является модель

$$\ln\tilde{y} = -6{,}4889 + \underset{(2{,}812)}{0{,}00127}x_3 + \underset{(3{,}555)}{0{,}533}\ln x_{18} + \underset{(3{,}447)}{0{,}754}\ln x_{22}, \qquad (30)$$

в которой значимы все коэффициенты при объясняющих переменных, а критерий $R^2 = 0{,}7437$.

Уравнению (30) соответствует МСПР вида

$$\tilde{y} = 0{,}00152 \cdot e^{0{,}00127x_3} \cdot x_{18}^{0{,}533} \cdot x_{22}^{0{,}754}. \qquad (31)$$

Сумма квадратов остатков для модели (31) составляет 229,598.

Интерпретация модели (31): с увеличением численности рабочей силы x_3 на 1 тыс. чел. перевозки грузов y возрастают в среднем на 0,127 %; с увеличением числа предприятий и организаций x_{18} на 1 % перевозки грузов y возрастают в среднем на 0,533 %; с увеличением производства электроэнергии x_{22} на 1 % перевозки грузов y возрастают в среднем на 0,754 %.

Затем по исходным данным осуществлялось построение ЛЛР по третьей стратегии. Настройки задавались те же, что и для МСПР. В результате перебора из 1320 альтернатив была выбрана модель

$$\tilde{y} = -267{,}173 + \underset{(2{,}55)}{0{,}000639}x_{24} - \underset{(-6{,}185)}{0{,}00193}x_{25} + \underset{(12{,}87)}{29{,}124}\ln x_{14}, \qquad (32)$$

в которой все коэффициенты при регрессорах значимы, а критерий $R^2 = 0{,}9328$.

В модели (32) коэффициент при переменной x_{25} снова не удовлетворяет содержательному смыслу задачи. Поэтому эта модель была перестроена с учетом контроля соответствия знаков коэффициентов. Оказалось, что из 1320 альтернатив только 64 удовлетворяют смыслу задачи. Лучшей из них оказалась регрессия

$$\tilde{y} = -552{,}38 + \underset{(2{,}764)}{0{,}0746}x_3 + \underset{(3{,}489)}{31{,}1352}\ln x_{18} + \underset{(3{,}282)}{42{,}7013}\ln x_{22}, \qquad (33)$$

в которой все коэффициенты при объясняющих переменных значимы, критерий $R^2 = 0{,}7312$, а сумма квадратов остатков составляет 233,236.

Как видно, по величине суммы квадратов остатков ЛЛР (33) оказалась несколько хуже, чем МСПР (31). При этом ЛЛР (33) имеет в своем составе те же регрессоры, что и МСПР (31).

Интерпретация модели (33): с увеличением численности рабочей силы x_3 на 1 тыс. чел. перевозки грузов y возрастают в среднем на 0,0746 млн т; с увеличением числа предприятий и организаций x_{18} на 1 % перевозки грузов y возрастают в среднем на 0,3113 млн т; с увеличением производства электроэнергии x_{22} на 1 % перевозки грузов y возрастают в среднем на 0,427 млн т. Также можно сделать вывод: если число предприятий и организаций x_{18} увеличится в e раз, то перевозки грузов y вырастут в среднем на 31,1352 млн т; если производство электроэнергии x_{22} увеличится в e раз, то перевозки грузов y вырастут в среднем на 42,7013 млн т.

Таким образом, если целью исследователя является получение только прогнозных значений объемов перевозок грузов y, то для этого лучше воспользоваться моделями (29) и (32). А если исследователя интересует еще и интерпретация влияния факторов на y, то нужно использовать примерно одинаковые по качеству, но разные по смыслу МСПР (31) и ЛЛР (33).

ЗАКЛЮЧЕНИЕ

В настоящей работе введены две новые спецификации регрессионных моделей — мультипликативная степенно-показательная регрессия (МСПР) и линейно-логарифмическая регрессия (ЛЛР). Рассмотрены вопросы их оценивания и интерпретации. Главным достоинством предложенных спецификаций является то, что в результате их построения каждому коэффициенту регрессии, за исключением свободного члена, гарантированно можно дать некоторую содержательную интерпретацию. Спецификации МСПР и ЛЛР позволяют выявлять и анализировать новые нелинейные закономерности функционирования исследуемых процессов или объектов, поэтому в целом повышают ценность регрессионного анализа.

ЛИТЕРАТУРА

1. *Pardoe, I.* Applied Regression Modeling. — Wiley, 2020. — 336 p.
2. *Westfall, P.H., Arias, A.L.* Understanding Regression Analysis: a Conditional Distribution Approach. — Chapman and Hall/CRC, 2020. — 514 p.
3. *Клейнер Г.Б.* Производственные функции: теория, методы, применение. — М.: Финансы и статистика, 1986. — 239 с. [*Kleiner, G.B.* Production Functions: Theory, Methods, Application. — Moscow: Finance and Statistics, 1986. — 239 s. (In Russian)]
4. *Клейнер Г.Б.* Экономика. Моделирование. Математика. Избранные труды. — М.: ЦЭМИ РАН, 2016. — 856 с. [*Kleiner, G.B.* Economy. Modeling. Maths. Selected Works. — Moscow: TsEMI RAN, 2016. — 856 s. (In Russian)]
5. *Хацкевич Г.А., Проневич А.Ф., Чайковский М.В.* Двухфакторные производственные функции с заданной предельной нормой замещения // Экономическая наука сегодня. — 2019. — № 10. — С. 169—181. [*Khatskevich, G.A., Pronevich, A.F., Chaikovskii, M.V.* Dvukhfaktornye proizvodstvennye funktsii s zadannoi predel'noi normoi zameshcheniya // Ekonomicheskaya nauka segodnya. — 2019. — No. 10. — P. 169—181. (In Russian)]
6. *Носков С.И.* Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. — Иркутск: Облинформпечать, 1996. — 320 с. [*Noskov, S.I.* Tekhnologiya modelirovaniya ob'ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh. — Irkutsk: Oblinformpechat', 1996. — 320 s. (In Russian)]
7. *Базилевский М.П., Носков С.И.* Формализация задачи построения линейно-мультипликативной регрессии в виде задачи частично-булевого линейного программирования // Современные технологии. Системный анализ. Моделирование. — 2017. — № 3 (55). — С. 101—105. [*Bazilevskii, M.P., Noskov, S.I.* Formalizatsiya zadachi postroeniya lineino-mul'tiplikativnoi regressii v vide zadachi chastichno-bulevogo lineinogo programmirovaniya // Sovremennye tekhnologii. Sistemnyi analiz. Modelirovanie. — 2017. — Vol. 55, No. 3. — P. 101—105. (In Russian)]
8. *Иванова Н.К., Лебедева С.А., Носков С.И.* Идентификация параметров некоторых негладких регрессий // Информационные технологии и проблемы математического моделирования сложных систем. — 2016. — № 17. — С. 107—110. [*Ivanova, N.K., Lebedeva, S.A., Noskov, S.I.* Identifikatsiya parametrov nekotorykh negladkikh regressii // Informatsionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnykh sistem. — 2016. — No. 17. — P. 107—110. (In Russian)]
9. *Носков С.И., Хоняков А.А.* Программный комплекс построения некоторых типов кусочно-линейных регрессий // Информационные технологии и математическое моделирование в управлении сложными системами. — 2019. — № 3 (4). — С. 47—55. [*Noskov, S.I., Khonyakov, A.A.* Programmnyi kompleks postroeniya nekotorykh tipov kusochno-lineinykh regressii // Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami. — 2019. — Vol. 4, No. 3. — P. 47—55. (In Russian)]
10. *Базилевский М.П., Носков С.И.* Оценивание индексных моделей регрессии с помощью метода наименьших модулей // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. — 2020. — № 1. — С. 17—23. [*Bazilevskii, M.P., Noskov, S.I.* Otsenivanie indeksnykh modelei regressii s pomoshch'yu metoda naimen'shikh modulei // Vestnik Rossiiskogo novogo universiteta. Seriya: Slozhnye sistemy: modeli, analiz i upravlenie. — 2020. — No. 1. — P. 17—23. (In Russian)]
11. *Базилевский М.П., Вергасов А.С., Носков С.И.* Групповой отбор информативных переменных в регрессионных моделях // Южно-Сибирский научный вестник. — 2019. — № 4-1 (28). — С. 36—39. [*Bazilevskii, M.P., Vergasov, A.S., Noskov, S.I.* Gruppovoi otbor informativnykh peremennykh v regressionnykh modelyakh // Yuzhno-Sibirskii nauchnyi vestnik. — 2019. — Vol. 28, No. 4-1. — P. 36—39. (In Russian)]
12. *Елисеева И.И., Курышева С.В., Костеева Т.В.* и др. Эконометрика. — М.: Финансы и статистика, 2007. — 576 с. [*Eliseeva, I.I., Kurysheva, S.V., Kosteeva, T.V.*, et al. Econometrics. — Moscow: Finance and statistics, 2007. — 576 p. (In Russian)]
13. *Доугерти К.* Введение в эконометрику. — М.: ИНФРА-М, 2009. — 465 с. [*Dougherti, K.* Introduction to Econometrics. — Moscow: INFRA-M, 2009. — 465 p. (In Russian)]
14. *Горидько Н.П., Нижегородцев Р.М.* Построение лаговых регрессионных моделей типа Кобба-Дугласа на долгосрочных временных горизонтах // Проблемы управления. — 2012. — № 3. — С. 55—63. [*Goridko, N.P., Nizhegorodtsev, R.M.* Elaborating of Long-Run Lag Regressional Models of Cobb-Douglas Type // Control Sciences. — 2012. — No. 3. — P. 55—63. (In Russian)]
15. *URL*: https://ru.coursera.org/lecture/ekonometrika/3-1-3-intierprietatsiia-koeffitsiienta-pri-logharifmirovanii-DdROH.

16. *Park, Y.W., Klabjan, Y.W.* Subset selection for multiple linear regression via optimization // Journal of Global Optimization. — 2020. — Vol. 77. — P. 543—574.
17. *Chung, S., Park, Y.W., Cheong, T.* A mathematical programming approach for integrated multiple linear regression subset selection and validation // Pattern Recognition. — 2020. — Vol. 108. — Article no. 107565.
18. *Bertsimas, D., Li, M.L.* Scalable holistic linear regression // Operations Research Letters. — 2020. — Vol. 48, no. 3. — P. 203—208.
19. *Базилевский М.П., Врублевский И.П., Носков С.И.* и др. Среднесрочное прогнозирование эксплуатационных показателей функционирования Красноярской железной дороги // Фундаментальные исследования. — 2016. — № 10-3. — С. 471—476. [*Bazilevskii, M.P., Vrublevskii, I.P., Noskov, S.I.*, et al. Srednesrochnoe prognozirovanie ekspluatatsionnykh pokazatelei funktsionirovaniya Krasnoyarskoi zheleznoi dorogi // Fundamental'nye issledovaniya. — 2016. — No. 10-3. — P. 471—476. (In Russian)]
20. *Носков С.И., Врублевский И.П.* Регрессионная модель динамики эксплуатационных показателей функционирования железнодорожного транспорта // Современные технологии. Системный анализ. Моделирование. — 2016. — № 2 (50). — С. 192—197. [*Noskov, S.I., Vrublevskii, I.P.* Regressionnaya model' dinamiki ekspluatatsionnykh pokazatelei funktsionirovaniya zheleznodorozhnogo transporta // Sovremennye tekhnologii. Sistemnyi analiz. Modelirovanie. — 2016. — Vol. 50, no. 2. — P. 192—197. (In Russian)]

Статья представлена к публикации членом редколлегии Н.Н. Бахтадзе.

Поступила в редакцию 11.01.2021, после доработки 04.03.2020.
Принята к публикации 10.03.2020.

Базилевский Михаил Павлович — канд. техн. наук, Иркутский государственный университет путей сообщения, ✉ mik2178@yandex.ru.

CONSTRUCTING POWER-EXPONENTIAL AND LINEAR-LOGARITHMIC REGRESSION MODELS

M.P. Bazilevskiy

Irkutsk State Transport University, Irkutsk, Russia
✉ mik2178@yandex.ru

Abstract. When using nonlinear regression models, the estimates of the resulting dependence are often difficult or even impossible to interpret. This paper develops nonlinear regression specifications in which any estimated parameter, except the free term, can always be given some practical interpretation. A multiplicative power-exponential regression generalizing the Cobb—Douglas production function and an additive linear-logarithmic regression are constructed. Three construction strategies are formulated for each of them, and the issues of interpreting their estimates are considered in detail. The construction strategies based on the least absolute deviations method are formalized as linear and partially Boolean linear programming problems. The mathematical apparatus developed in this paper is illustrated by modeling rail freight traffic in Irkutsk oblast.

Keywords: regression model, interpretation, multiplicative power-exponential regression, linear-logarithmic regression, feature selection, least absolute deviations, rail freight traffic.