УДК 519.862.6
ISSN 1995-5499
DOI: https://doi.org/
Поступила в редакцию 29.10.2020
Подписана в печать 02.02.2021

ПОСТРОЕНИЕ СТЕПЕННО-ПОКАЗАТЕЛЬНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ И ИХ ИНТЕРПРЕТАЦИЯ

© 2020 М. П. Базилевский✉

Иркутский государственный университет путей сообщения
ул. Чернышевского, 15, 664074 Иркутск, Российская Федерация

Аннотация. Настоящая статья посвящена разработке и исследованию новых структурных спецификаций регрессионных моделей. Предложено два обобщения производственной функции Кобба — Дугласа: степенно-показательная регрессия со степенями в виде линейных комбинаций объясняющих переменных и в виде линейных комбинаций натуральных логарифмов объясняющих переменных. Аппроксимационные качества этих моделей всегда не хуже, чем для производственных функций Кобба — Дугласа. Модели первого типа прекрасно подходят для прогнозирования значений зависимой переменной, но возникает проблема с содержательной интерпретацией их оценок. Для моделей второго типа доказано, что коэффициент эластичности функции регрессии по конкретной объясняющей переменной не зависит от её значения и представляет собой линейную комбинацию натуральных логарифмов оставшихся переменных. Поэтому оценки таких моделей вполне можно интерпретировать. Кроме того, на основе моделей второго типа по заданным значениям коэффициентов эластичности можно определять прогнозные значения объясняющих и объясняемой переменной. Построены четыре регрессионных модели зависимости объемов пассажирских перевозок железнодорожного транспорта Иркутской области от числа собственных легковых автомобилей, цен на проезд в поезде и цен на перелет в самолете. Дана подробная интерпретация степенно-показательной регрессии со степенями в виде линейных комбинаций натуральных логарифмов объясняющих переменных. На основе этой модели построены графики изменения эластичности объема перевозок по каждой объясняющей переменной от времени и получены прогнозные значения объясняющих и объясняемой переменной.
Ключевые слова: регрессионная модель, производственная функция, степенно-показательная регрессия, функция Кобба — Дугласа, коэффициент эластичности, пассажирские перевозки железнодорожного транспорта.

ВВЕДЕНИЕ

Признанным во всем мире инструментом анализа данных является регрессионный анализ [1, 2]. Первоочередной проблемой, с которой сталкивается исследователь при построении регрессионной модели, является выбор её спецификации, т.е. состава переменных и математической формы связи между ними. К настоящему времени разработан значительный арсенал форм связи между переменными, описание большинства из которых можно найти в работах [3–8].

Чаще всего регрессионный анализ начинается с оценивания самой простой зависимости — модели множественной линейной регрессии [1–3]:

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

где y_i, $i = \overline{1,n}$ — наблюдаемые значения объясняемой (выходной) переменной y; x_{ij}, $i = \overline{1,n}$, $j = \overline{1,m}$ — наблюдаемые значения объясняющих (входных) переменных x_1, x_2, …, x_m; ε_i, $i = \overline{1,n}$ — ошибки аппроксимации; α_0, α_1, α_2, …, α_m — неизвестные параметры.

✉ Базилевский Михаил Павлович
e-mail: mik2178@yandex.ru

Оценки параметров $\alpha_1, \alpha_2, \ldots, \alpha_m$ модели (1) легко интерпретировать. Например, оцененный коэффициент $\tilde{\alpha}_s$ $(1 \le s \le m)$ при переменной x_s показывает, что с изменением x_s на 1 единицу (при неизменных значениях остальных переменных), значение выходного показателя y изменится в среднем на $\tilde{\alpha}_s$ единиц.

В [4] приводится описание линейно-мультипликативных регрессий (ЛМР) вида

$$y_i = \alpha_0 + \sum_{j=1}^{2^m-1} \alpha_j \prod_{k=1}^{m} x_{ik}^{\lambda_{jk}} + \varepsilon_i, \quad i = \overline{1,n}, \qquad (2)$$

где λ_{jk} — элементы бинарной матрицы $\Lambda_{(2^m-1)\times m}$, строками которой являются размещения с повторениями из m элементов по 2 (без нулевой строки).

Дать интерпретацию оценкам параметров ЛМР (2) не представляется возможным. Однако можно интерпретировать частные производные её оцененного уравнения по объясняющим переменным.

Модели (1)–(2) являются линейными по параметрам, поэтому без труда оцениваются с помощью метода наименьших квадратов (МНК).

В эконометрике особое внимание уделяется вопросам построения производственных функций (ПФ). В работах [5–8] можно найти описание следующих ПФ: линейной, Леонтьева, Кобба — Дугласа, CES, Солоу, Джири, логарифмической, Реванкара (с линейной эластичностью замещения, LES), Сато, Лу — Флетчера, Лиу — Хильдебранда, Кадияла, Бруно, Сато —Гофмана, Солоу — Мукерджи.

Для ПФ Кобба — Дугласа оценивается степенная модель:

$$y_i = \alpha_0 \prod_{j=1}^{m} x_{ij}^{\alpha_j} \varepsilon_i, \quad i = \overline{1,n}. \qquad (3)$$

Известно [8, 9], что оцененный коэффициент $\tilde{\alpha}_s$ $(1 \le s \le m)$ степенной модели (3) характеризует эластичность ПФ по соответствующей переменной. Иными словами, $\tilde{\alpha}_s$ показывает, что с изменением переменной x_s на 1 %, значение выходного показателя y изменится в среднем на $\tilde{\alpha}_s$ %.

Как видно, в модель (3) ошибки входят мультипликативно, поэтому её можно линеаризовать с помощью логарифмирования и оценивать на основе МНК.

В работе [10] рассмотрены вопросы построения ПФ Леонтьева через оценивание кусочно-линейных регрессионных моделей. В [11] предложено обобщение ПФ Леонтьева — индексная регрессия. Для оценки моделей в [10] и [11] используется метод наименьших модулей (МНМ). Успешная попытка оценивания двухфакторных кусочно-линейных регрессий с помощью МНК была предпринята в работе [12].

В работе [13] на основе объединения двух производств предложена обобщенная ПФ Кобба — Дугласа:

$$y_i = \alpha_0 \prod_{j=1}^{2} x_{ij}^{\alpha_j} + \beta_0 \prod_{j=3}^{4} x_{ij}^{\beta_{j-2}} + \varepsilon_i, \quad i = \overline{1,n}. \qquad (4)$$

А в работе [14] предложено обобщение ПФ Кобба — Дугласа за счет включения в неё переменной времени:

$$y_i = \left(\alpha_0 + \alpha_1 t\right) x_{i1}^{\beta_0 + \beta_1 t} x_{i2}^{\gamma_0 + \gamma_1 t} \varepsilon_i, \quad i = \overline{1,n}. \qquad (5)$$

Как видно, в настоящее время активно ведутся исследования в области разработки новых и вполне интерпретируемых регрессионных моделей с интересными свойствами. Целью данной работы является обобщение ПФ Кобба — Дугласа на основе использования степенно-показательных функций [15] и его применение для моделирования пассажирских перевозок железнодорожного транспорта Иркутской области.

1. МАТЕРИАЛЫ И МЕТОДЫ

Введем степенно-показательную регрессию (СПР) со степенями в виде линейных комбинаций объясняющих переменных:

$$y_i = \alpha_0 \prod_{j=1}^{m} x_{ij}^{\alpha_{j0} + \sum_{k=1}^{m} \alpha_{jk} x_{ik}} \cdot \varepsilon_i, \quad i = \overline{1,n}, \qquad (6)$$

где α_0, α_{jk}, $j = \overline{1,m}$, $k = \overline{0,m}$ — неизвестные параметры, общее количество которых $(m^2 + m + 1)$.

Очевидно, что ПФ Кобба — Дугласа (3) является частным случаем модели (6) при $\alpha_{jk} = 0$, $j = \overline{1,m}$, $k = \overline{1,m}$.

Модель (6) является нелинейной, однако легко линеаризуется с помощью логарифмирования:

$$\ln y_i = \ln \alpha_0 + \sum_{j=1}^{m} \alpha_{j0} \ln x_{ij} +$$

$$+\sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_{jk} x_{ik} \ln x_{ij} + \ln \varepsilon_i, \quad i = \overline{1,n}. \quad (7)$$

Сделав в (7) замены $z_i = \ln y_i$, $c_0 = \ln \alpha_0$, $v_{ij} = \ln x_{ij}$, $w_{ijk} = x_{ik} \ln x_{ij}$, $u_i = \ln \varepsilon_i$, получим линейную регрессию:

$$z_i = c_0 + \sum_{j=1}^{m} \alpha_{j0} v_{ij} +$$

$$+\sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_{jk} w_{ijk} + u_i, \; i = \overline{1,n}. \quad (8)$$

Неизвестные параметры модели (8) могут быть найдены с помощью метода наименьших квадратов (МНК).

Понятно, что аппроксимационные качества СПР (6) всегда будут не хуже, чем у моделей Кобба — Дугласа (3), поскольку первые содержат на m^2 больше неизвестных параметров, чем последние. Кроме того, за счет меньшего числа степеней свободы, СПР могут быть лучше по качеству, чем линейные регрессии (1). Отсюда можно сделать вывод, что СПР (6) прекрасно подходят для прогнозирования значений зависимой переменной. Но возникает проблема с содержательной интерпретацией оценок их параметров.

Как отмечено выше, оценки параметров α_1, ..., α_m ПФ Кобба — Дугласа (3) равны коэффициентам эластичности по соответствующим переменным. Коэффициент эластичности $Э_{yx_s}$ функции нескольких переменных $y = f(x_1, x_2, ..., x_m)$ по переменной x_s находится по формуле:

$$Э_{yx_s} = x_s \frac{y'_{x_s}}{y} = x_s (\ln y)'_{x_s}. \quad (9)$$

Пусть оцененная с помощью МНК модель (7) имеет вид

$$\ln \tilde{y} = \ln \tilde{\alpha}_0 + \sum_{j=1}^{m} \tilde{\alpha}_{j0} \ln x_j +$$

$$+\sum_{j=1}^{m}\sum_{k=1}^{m} \tilde{\alpha}_{jk} x_k \ln x_j, \quad (10)$$

где $\tilde{\alpha}_0$, $\tilde{\alpha}_{jk}$, $j = \overline{1,m}$, $k = \overline{0,m}$ — оценки параметров.

Определив частную производную функции (10) по переменной x_s и подставив её в (9), получим коэффициент эластичности оцененной СПР (6) по переменной x_s $(1 \le s \le m)$:

$$Э_{\tilde{y}x_s} = \tilde{\alpha}_{s0} + \tilde{\alpha}_{ss} x_s (1 + \ln x_s) +$$

$$+ \sum_{k \in \{1,...,m\} \setminus s} \tilde{\alpha}_{sk} x_k + x_s \sum_{j \in \{1,...,m\} \setminus s} \tilde{\alpha}_{js} \ln x_j. \quad (11)$$

Как видно, выражение (11) оказалось в значительной степени нелинейными, поэтому проблема содержательной интерпретации оценок СПР (6) остается открытой.

Введем СПР со степенями в виде линейных комбинаций натуральных логарифмов объясняющих переменных:

$$y_i = \alpha_0 \prod_{j=1}^{m} x_{ij}^{\alpha_{j0} + \sum_{k=1}^{m} \alpha_{jk} \ln x_{ik}} \cdot \varepsilon_i, \; i = \overline{1,n}. \quad (12)$$

Прологарифмировав выражение (12), получим:

$$\ln y_i = \ln \alpha_0 + \sum_{j=1}^{m} \alpha_{j0} \ln x_{ij} +$$

$$+\sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_{jk} \ln x_{ik} \ln x_{ij} + \ln \varepsilon_i, \; i = \overline{1,n}. \quad (13)$$

Нетрудно заметить, что в (13) в двойной сумме регрессоры $\ln x_{ik} \ln x_{ij}$ при коэффициентах α_{jk} и α_{kj} $(j \ne k)$ одинаковы. Следовательно, МНК-оценки линейной регрессии (13) невозможно получить из-за совершенной коллинеарности факторов. В этой связи введем ограничения на параметры модели (12).

Будем считать, что в СПР (12) $\alpha_{jk} = \alpha_{kj}$, при $j \ne k$. Тогда эта модель будет иметь $\left(\frac{m^2 - m}{2} + 2m + 1\right)$ неизвестных параметров, а выражение (13) примет вид:

$$\ln y_i = \ln \alpha_0 + \sum_{j=1}^{m} \alpha_{j0} \ln x_{ij} + \sum_{j=1}^{m} \alpha_{jj} \ln^2 x_{ij} +$$

$$+2\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \alpha_{jk} \ln x_{ik} \ln x_{ij} + \ln \varepsilon_i, \; i = \overline{1,n}. \quad (14)$$

Сделав в (14) замены $z_i = \ln y_i$, $c_0 = \ln \alpha_0$, $v_{ij} = \ln x_{ij}$, $d_{ij} = \ln^2 x_{ij}$, $q_{ijk} = 2 \ln x_{ik} \ln x_{ij}$, $u_i = \ln \varepsilon_i$, получим линейную регрессию:

$$z_i = c_0 + \sum_{j=1}^{m} \alpha_{j0} v_{ij} + \sum_{j=1}^{m} \alpha_{jj} d_{ij} +$$

$$+\sum_{j=1}^{m-1}\sum_{k=j+1}^{m} \alpha_{jk} q_{ijk} + u_i, \; i = \overline{1,n}. \quad (15)$$

Пусть оцененная с помощью МНК модель (14) имеет вид

$$\ln \tilde{y} = \ln \tilde{\alpha}_0 + \sum_{j=1}^{m} \tilde{\alpha}_{j0} \ln x_j + \sum_{j=1}^{m} \tilde{\alpha}_{jj} \ln^2 x_j + \\ +2\sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \tilde{\alpha}_{jk} \ln x_k \ln x_j. \quad (16)$$

Определив частную производную функции (16) по переменной x_s и подставив её в (9), получим коэффициент эластичности оцененной СПР (12) по переменной x_s:

$$Э_{\tilde{y}x_s} = \tilde{\alpha}_{s0} + 2\tilde{\alpha}_{ss} \ln x_s + 2 \sum_{k\in\{1,\ldots,m\}\setminus s} \tilde{\alpha}_{sk} \ln x_k . \quad (17)$$

Как видно, коэффициент эластичности (17) оцененной СПР (12) по переменной x_s зависит от её значения, что затрудняет его интерпретацию. Поэтому обнулим в модели (12) параметры α_{jj}, $j = \overline{1,m}$:

$$y_i = \alpha_0 \prod_{j=1}^{m} x_{ij}^{\alpha_{j0} + \sum_{k\in\{1,\ldots,m\}\setminus j} \alpha_{jk} \ln x_{ik}} \cdot \varepsilon_i, \quad i = \overline{1,n}. \quad (18)$$

Эта модель содержит $\left(\dfrac{m^2 - m}{2} + m + 1 \right)$ неизвестных параметров, а её МНК-оценки следует находить в соответствии с (15).

Тогда, с учетом (17), для оцененной СПР (18) коэффициент эластичности по переменной x_s:

$$Э_{\tilde{y}x_s} = \tilde{\alpha}_{s0} + 2 \sum_{k\in\{1,\ldots,m\}\setminus s} \tilde{\alpha}_{sk} \ln x_k . \quad (19)$$

Таким образом, коэффициент эластичности (19) по переменной x_s не зависит от того, какое значение она принимает, а зависят от значений других объясняющих переменных. Поэтому при их фиксации для оцененной СПР (18) коэффициент эластичности по переменной x_s является константой и его можно интерпретировать следующим образом: с изменением x_s на 1 % переменная y меняется с среднем на $Э_{\tilde{y}x_s}$ %.

В уравнении (19) коэффициент $2\tilde{\alpha}_{sk}$ показывает, на сколько изменится коэффициент эластичности функции по переменной x_s, если переменная x_k изменится в e раз.

То, что коэффициент эластичности (19) является линейной комбинацией натуральных логарифмов объясняющих переменных, можно использовать для прогнозирования значений объясняющих и объясняемой переменной по следующему сценарию. Предположим, что исследователь может назначить коэффициенты эластичности $\tilde{Э}_{yx_1}, \tilde{Э}_{yx_2}, \ldots, \tilde{Э}_{yx_m}$ по каждой объясняющей переменной, характеризующие состояние исследуемого социально-экономического процесса в будущем. Тогда, с учетом (19), для нахождения прогнозных значений требуется решить систему линейных алгебраических уравнений относительно переменных v_j, $j = \overline{1,m}$:

$$\begin{cases} \sum_{k\in\{1,\ldots,m\}\setminus 1} \tilde{\alpha}_{1k} v_k = \dfrac{1}{2}\left(\tilde{Э}_{yx_1} - \tilde{\alpha}_{10}\right), \\ \sum_{k\in\{1,\ldots,m\}\setminus 2} \tilde{\alpha}_{2k} v_k = \dfrac{1}{2}\left(\tilde{Э}_{yx_2} - \tilde{\alpha}_{20}\right), \\ \ldots \\ \sum_{k\in\{1,\ldots,m\}\setminus m} \tilde{\alpha}_{mk} v_k = \dfrac{1}{2}\left(\tilde{Э}_{yx_m} - \tilde{\alpha}_{m0}\right). \end{cases} \quad (20)$$

Пусть система (20) имеет решение: $v_j = v_j^*$, $j = \overline{1,m}$. Тогда прогнозные значения объясняющих переменных находятся по формулам $x_j = e^{v_j^*}$, $j = \overline{1,m}$.

2. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Моделированию объемов перевозок на железнодорожном транспорте посвящено множество научных работ (см., например, [16–18]). При этом, как показывает анализ литературных источников, очень много работ связано с построением регрессионных моделей грузовых перевозок, а моделированию пассажирских перевозок уделяется гораздо меньше внимания. Поэтому было принято решение для демонстрации предложенного выше математического аппарата построить регрессионную модель отправления пассажиров железнодорожного транспорта Иркутской области.

Сначала для построения модели на официальном сайте Федеральной службы государственной статистики были собраны годовые данные (табл. 1) за период 2000 – 2018 гг. по следующим показателям:

y — отправление пассажиров железнодорожным транспортом общего пользования (тыс. человек);

x_1 — число собственных легковых автомобилей на 1000 человек населения (штук);

x_2 — средние потребительские цены на проезд в пригородном поезде (поездка) за декабрь (руб.);

x_3 — средние потребительские цены на полет в салоне экономического класса самолета (в расчете на 1000 км пути) за декабрь (руб.).

Как видно, объясняющая переменная x_1 связана с автомобильным транспортом, x_2 – с железнодорожным, x_3 – с воздушным.

Построенная по этим данным модель множественной линейной регрессии имеет вид

$$\tilde{y}_1 = 35227{,}8 - \underset{(-3{,}122)}{67{,}145}\, x_1 - \underset{(-3{,}419)}{284{,}329}\, x_2 + \underset{(3{,}458)}{1{,}203}\, x_3, \quad (21)$$

для которой коэффициент детерминации $R^2 = 0{,}8897$, сумма квадратов остатков $\sum e^2 = 71260504$. В скобках под коэффициентами указаны значения t-критериев Стьюдента, указывающие на значимость всех трех переменных.

Таблица 1. Статистические данные
[Table 1. Statistical data]

Год	y	x_1	x_2	x_3
2000	22133	138,8	5,41	1000,63
2001	23216	140	7,6	1058,38
2002	23363	147,2	8,6	1272,93
2003	25122	151,6	11,2	1804,49
2004	26435	156,5	11,2	1962,69
2005	26247	143,3	12,8	2819,53
2006	27817	154,8	12,8	3529,62
2007	28501	169,2	14	4811,21
2008	28757	188,2	16	6590,97
2009	22427	189,8	20	6202,74
2010	19774	202,6	20	5633,74
2011	18059	224,3	23	5863,89
2012	17266	251,5	25	5888,33
2013	14930	271,8	28	3554,45
2014	14161	270,5	30	3641,52
2015	12967	271,3	36	5095,42
2016	12998	242,7	40	5297,99
2017	12003	246,2	44	5129,7
2018	11796	245,6143	46	4669,4

Интерпретация модели (21): с увеличением числа легковых автомобилей на 1000 человек населения x_1 на 1 штуку (при неизменных x_2 и x_3) перевозки пассажиров y снижаются в среднем на 67145 человек; с увеличением цены на проезд в пригородном поезде x_2 на 1 руб. (при неизменных x_1 и x_3) перевозки пассажиров y снижаются в среднем на 284329 человек; с увеличением цены на полет в салоне самолета x_3 на 1 руб. (при неизменных x_1 и x_2) перевозки пассажиров y возрастают в среднем на 1203 человека.

Полученные результаты согласуются с экономическим смыслом задачи. Так, чем больше обеспеченность населения собственным автомобильным транспортом, тем меньше спрос на услуги железнодорожного транспорта. Чем выше цены на проезд в поезде, тем ниже спрос на перевозки. А чем выше цены на полет в самолете, тем выше спрос на железнодорожные перевозки.

Для построения ПФ Кобба — Дугласа сначала была оценена прологарифмированная модель

$$\ln \tilde{y}_2 = 11{,}524 - \underset{(-2{,}606)}{0{,}603} \ln x_1 - \underset{(-4{,}367)}{0{,}47} \ln x_2 + \underset{(6{,}145)}{0{,}356} \ln x_3, \quad (22)$$

для которой $R^2 = 0{,}9236$.

Тогда ПФ Кобба — Дугласа имеет вид

$$\tilde{y}_2 = 101100 \cdot x_1^{-0{,}603} \cdot x_2^{-0{,}47} \cdot x_3^{0{,}356}, \quad (23)$$

для которой $\sum e^2 = 65823179$. Как видно, по этому показателю качество модели (23) выше, чем у (21).

Интерпретация модели (23): с увеличением числа легковых автомобилей на 1000 человек населения x_1 на 1% перевозки пассажиров y снижаются в среднем на 0,603 %; с увеличением цены на проезд в пригородном поезде x_2 на 1% перевозки пассажиров y снижаются в среднем на 0,47 %; с увеличением цены на полет в салоне самолета x_3 на 1% перевозки пассажиров y возрастают в среднем на 0,356 %.

Для построения СПР со степенями в виде линейных комбинаций объясняющих переменных (6) сначала была оценена прологарифмированная модель

$$\ln \tilde{y}_3 = 57{,}143 - \underset{(-0{,}019)}{0{,}067} \ln x_1 -$$
$$-\underset{(-0{,}019)}{0{,}022} \ln x_2 - \underset{(-2{,}25)}{7{,}589} \ln x_3 -$$
$$-\underset{(-2{,}659)}{0{,}104} x_1 \ln x_1 + \underset{(1{,}276)}{0{,}027} x_1 \ln x_2 + \underset{(2{,}349)}{0{,}084} x_1 \ln x_3 -$$
$$-\underset{(-2{,}058)}{0{,}368} x_2 \ln x_1 + \underset{(1{,}324)}{0{,}227} x_2 \ln x_2 + \underset{(1{,}517)}{0{,}098} x_2 \ln x_3 -$$
$$-\underset{(-2{,}366)}{0{,}004} x_3 \ln x_1 - \underset{(-1{,}136)}{0{,}0004} x_3 \ln x_2 +$$
$$+\underset{(2{,}421)}{0{,}002} x_3 \ln x_3, \qquad (24)$$

для которой $R^2 = 0{,}9944$.

Тогда оцененная СПР (6) имеет вид

$$\tilde{y}_3 = 6{,}562 \cdot 10^{24} \cdot x_1^{(-0{,}067-0{,}104x_1-0{,}368x_2-0{,}004x_3)} \times$$
$$\times x_2^{(-0{,}022+0{,}027x_1+0{,}227x_2-0{,}0004x_3)} \times$$
$$\times x_3^{(-7{,}589+0{,}084x_1+0{,}098x_2+0{,}002x_3)}, \qquad (25)$$

для которой $\sum e^2 = 4424523$. Как видно, по этому показателю качество модели (25) гораздо выше, чем у (21) и (23). Даже несмотря на то, что в этой модели присутствует сильная мультиколлинеарность и некоторые коэффициенты незначимы, она прекрасно подходит для прогнозирования. Но интерпретировать её коэффициенты не представляется возможным.

Для построения СПР со степенями в виде линейных комбинаций натуральных логарифмов объясняющих переменных (18) сначала была оценена прологарифмированная модель

$$\ln \tilde{y}_4 = 12{,}7016 - \underset{(-0{,}699)}{2{,}264} \ln x_1 +$$
$$\underset{(1{,}858)}{2{,}357} \ln x_2 + \underset{(0{,}175)}{0{,}254} \ln x_3 +$$
$$+\underset{(0{,}183)}{0{,}0604} \ln x_1 \ln x_2 + \underset{(0{,}591)}{0{,}1993} \ln x_1 \ln x_3 -$$
$$-\underset{(-2{,}955)}{0{,}3889} \ln x_2 \ln x_3, \qquad (26)$$

для которой $R^2 = 0{,}9865$.

Тогда оцененная СПР (18) имеет вид

$$\tilde{y}_4 = 328261 x_1^{(-2{,}264+0{,}0302\ln x_2+0{,}0997\ln x_3)} \times$$
$$\times x_2^{(2{,}357+0{,}0302\ln x_1-0{,}1944\ln x_3)} \times$$
$$\times x_3^{(0{,}254+0{,}0997\ln x_1-0{,}1944\ln x_2)}, \qquad (27)$$

для которой $\sum e^2 = 14127961$. Как видно, по этому показателю качество модели (27) выше, чем у (21) и (23), но ниже, чем у (25).

Для функции (27) с использованием формулы (19) были найдены коэффициенты эластичности:

$$Э_{\tilde{y}_4 x_1} = -2{,}264 + 0{,}0604 \ln x_2 + 0{,}1993 \ln x_3, \quad (28)$$
$$Э_{\tilde{y}_4 x_2} = 2{,}357 + 0{,}0604 \ln x_1 - 0{,}3889 \ln x_3, \quad (29)$$
$$Э_{\tilde{y}_4 x_3} = 0{,}254 + 0{,}1993 \ln x_1 - 0{,}3889 \ln x_2. \quad (30)$$

Зависимости (28)–(30) можно интерпретировать следующим образом: с увеличением числа легковых автомобилей на 1000 человек населения x_1 в e раз (при неизменных x_2 и x_3) коэффициент эластичности $Э_{\tilde{y}_4 x_2}$ увеличивается на 0,0604 %, а $Э_{\tilde{y}_4 x_3}$ – на 0,1993 %; с увеличением цены на проезд в пригородном поезде x_2 в e раз (при неизменных x_1 и x_3) коэффициент эластичности $Э_{\tilde{y}_4 x_1}$ увеличивается на 0,0604%, а $Э_{\tilde{y}_4 x_3}$ уменьшается на 0,3889 %; с увеличением цены на полет в салоне самолета x_3 в e раз (при неизменных x_1 и x_2) коэффициент эластичности $Э_{\tilde{y}_4 x_1}$ увеличивается на 0,1933 %, а $Э_{\tilde{y}_4 x_2}$ уменьшается на 0,3889 %.

Интерпретация модели (27): с увеличением числа легковых автомобилей на 1000 человек населения x_1 на 1 % (при неизменных x_2 и x_3) перевозки пассажиров y меняются в среднем на $Э_{\tilde{y}_4 x_1}$ %; с увеличением цены на проезд в пригородном поезде x_2 на 1 % (при неизменных x_1 и x_3) перевозки пассажиров y меняются в среднем на $Э_{\tilde{y}_4 x_2}$ %; с увеличением цены на полет в салоне самолета x_3 на 1 % (при неизменных x_1 и x_2) перевозки пассажиров y меняются в среднем на $Э_{\tilde{y}_4 x_3}$ %. Динамика изменения коэффициентов эластичности $Э_{\tilde{y}_4 x_1}$, $Э_{\tilde{y}_4 x_2}$ и $Э_{\tilde{y}_4 x_3}$ от времени представлена на рис. 1, рис. 2 и рис. 3 соответственно.

По рис. 1 видно, что в 2000 г. эластичность перевозок пассажиров y по числу легковых автомобилей x_1 составляла -0,785 % и с течением времени снижалась по абсолютной величине по линейному закону до 2008 г., начиная с которого оставалась примерно на уровне –0,3 %. Иными словами, за 2000–2018 гг. произошло снижение чувствительности переменной y к изменению x_1.

По рис. 2 видно, что в 2000 г. эластичность перевозок пассажиров y по цене на проезд в поезде x_2 была практически нулевой и составляла –0,031 %. Однако с течением време-

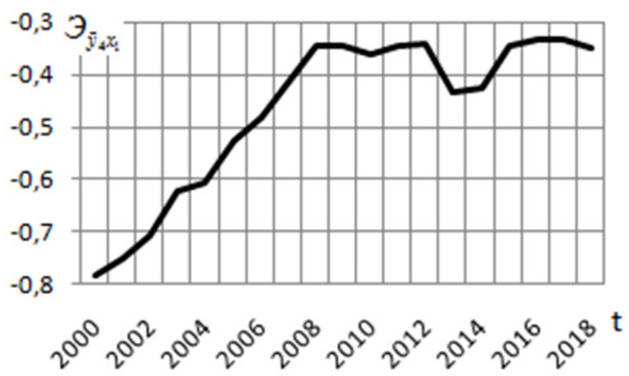

Рис. 1. Зависимость коэффициента эластичности по переменной x_1 от времени [Fig. 1. Dependence of the elasticity coefficient for the variable x_1 on time]

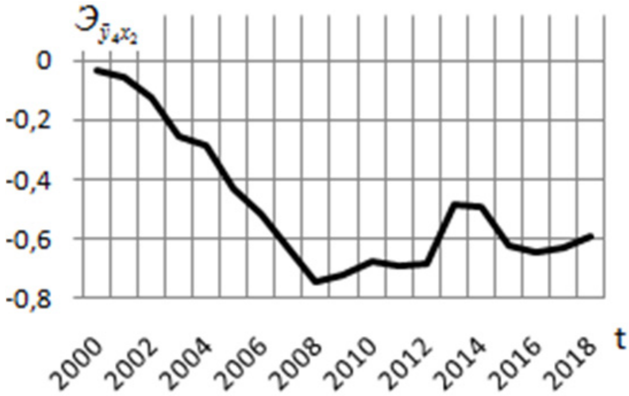

Рис. 2. Зависимость коэффициента эластичности по переменной x_2 от времени [Fig. 2. Dependence of the elasticity coefficient for the variable x_2 on time]

Рис. 3. Зависимость коэффициента эластичности по переменной x_3 от времени [Fig. 3. Dependence of the elasticity coefficient for the variable x_3 on time]

ни росла по абсолютной величине по линейному закону до 2008 г., начиная с которого оставалась примерно на уровне –0,6 %. На основании этого можно заключить, что за 2000–2018 гг. произошло повышение чувствительности переменной y к изменению x_2.

По рис. 3 видно, что в 2000 г. эластичность перевозок пассажиров y по цене на полет в самолете x_3 была наибольшей и составляла 0,58 %. Но с течением времени снижалась по абсолютной величине по линейному закону до 0 % в 2014–2015 гг. После чего эластичность, сменив знак, начала рост по абсолютной величине и в 2018 г. составила –0,138 %. Таким образом, за 2000–2018 гг. произошло снижение чувствительности переменной y к изменению x_3. При этом сменился знак влияния x_3 на y.

Пусть исследователь, исходя, например, из информации, представленной на рис. 1, рис. 2 и рис. 3, может назначить значения коэффициентов эластичности на 2019 г.: $\tilde{Э}_{\tilde{y}_4 x_1} = -0,34$ %, $\tilde{Э}_{\tilde{y}_4 x_2} = -0,61$ %, $\tilde{Э}_{\tilde{y}_4 x_3} = -0,14$ %. Решив систему линейных уравнений (20), получим

$$v_1^* = 5,529,\ v_2^* = 3,846,\ v_3^* = 8,489,$$

откуда

$$\tilde{x}_1^{2019} = 251,913,\ \tilde{x}_2^{2019} = 46,826,\ \tilde{x}_3^{2019} = 4860,969.$$

Реальные значения этих переменных за 2019 г.:

$$x_1^{2019} = 254,9,\ x_2^{2019} = 48,\ x_3^{2019} = 4416,62.$$

Используя значения $\tilde{x}_1^{2019}$, $\tilde{x}_2^{2019}$, $\tilde{x}_3^{2019}$, по формуле (27) получим прогнозное значение пассажирских перевозок на 2019 г.: $\tilde{y}_{2019} = 11434,322$ тыс. человек.

Реальное значение этой переменной за 2019 г. $y_{2019} = 11907$.

Относительные ошибки прогноза для переменных x_1, x_2, x_3, y составляют соответственно 1,172 %, 2,45 %, 10,06 %, 3,97 %, что позволяет сделать вывод о вполне хорошем прогнозном качестве построенной модели.

ЗАКЛЮЧЕНИЕ

В работе исследованы СПР со степенями в виде линейных комбинаций объясняющих переменных и со степенями в виде натуральных логарифмов объясняющих переменных. Первые хорошо подходят для прогнозирования, но их затруднительно интерпретировать, а вторые несколько уступают первым по аппроксимационным качествам, но являются вполне интерпретируемыми. Установлено, что СПР второго типа имеют переменные эластичности, в отличие от ПФ Кобба — Дугласа. Предложен способ прогнозирования объясняющих и объясняемой переменной по известным коэффициентам эластичности.

Построено четыре регрессионных модели объемов пассажирских перевозок железнодорожного транспорта Иркутской области: линейная (21), ПФ Кобба — Дугласа (23), и два типа СПР — (25) и (27). Подробно рассмотрена интерпретация оценок моделей (21), (23) и (27). Для регрессии (27) построены зависимости коэффициентов эластичности от времени, демонстрирующие изменение чувствительности объема перевозок по каждой переменной. С помощью предложенного способа получены прогнозы для всех исследованных переменных.

КОНФЛИКТ ИНТЕРЕСОВ

Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Westfall, P. H.* Understanding regression analysis: a conditional distribution approach / P. H. Westfall, A. L. Arias. – Chapman and Hall/CRC, 2020. – 514 p.

2. *Pardoe, I.* Applied regression modeling / I. Pardoe. – Wiley, 2020. – 336 p.

3. *Носков, С. И.* Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных / С. И. Носков. – Иркутск : Облинформпечать, 1996. – 320 с.

4. *Базилевский, М. П.* Формализация задачи построения линейно-мультипликативной регрессии в виде задачи частично-булевого линейного программирования / М. П. Базилевский, С. И. Носков // Современные технологии. Системный анализ. Моделирование. – 2017. – № 3 (55). – С. 101–105.

5. *Клейнер, Г. Б.* Производственные функции: теория, методы, применение / Г. Б. Клейнер. – М. : Финансы и статистика, 1986. – 239 с.

6. *Клейнер, Г. Б.* Экономика. Моделирование. Математика. Избранные труды / Г. Б. Клейнер. – М. : ЦЭМИ РАН, 2016. – 856 с.

7. *Горбунов, В. К.* Производственные функции: теория и построение / В. К. Горбунов. – Ульяновск: УлГУ, 2013. – 84 с.

8. *Хацкевич, Г. А.* Двухфакторные производственные функции с заданной предельной нормой замещения / Г. А. Хацкевич, А. Ф. Проневич, М. В. Чайковский // Экономическая наука сегодня. – 2019. – № 10. – С. 169–181.

9. *Кутенков, Р. П.* Производственные функции: оценки взаимозаменяемости факторов и прогнозирование объемов сельскохозяйственного производства в регионах России / Р.П. Кутеноков // Региональные агросистемы: экономика и социология. – 2019. – № 2. – С. 50–57.

10. *Иванова, Н. К.* Идентификация параметров некоторых негладких регрессий / Н. К. Иванова, С. А. Лебедева, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – № 17. – С. 107–110.

11. *Базилевский, М. П.* Оценивание индексных моделей регрессии с помощью метода наименьших модулей / М. П. Базилевский, С. И. Носков // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. – 2020. – № 1. – С. 17–23.

12. *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.

13. *Добровольская, Л. П.* Псевдослучайный поиск и сложные эконометрические модели / Л.П. Добровольская // Материалы XV Международной конференции Алгебра, теория чисел и дискретная геометрия: современные проблемы и приложения. – Тула: ТГПУ им. Л. Н. Толстого, 2018. – С. 291–293.

14. *Roubalova, L.* The time augmented Cobb-Douglas production function / L. Roubalova, L. Viskotova // Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunesis. – 2019. – Vol. 67, no. 5. – P. 1347–1356.

15. *Базилевский, М. П.* Исследование степенно-показательных регрессионных моделей со степенями в виде линейных комбинаций логарифмов объясняющих переменных / М. П. Базилевский // Транспортная инфраструктура Сибирского региона. – 2018. – Т. 1. – С. 397–401.

16. *Краковский, Ю. М.* Обобщенное прогнозирование показателя погрузки грузов при перевозке железнодорожным транспортом / Ю. М. Краковский, Н. Н. Попова // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии. – 2020. – № 3. – С. 43–50.

17. *Носков, С. И.* Анализ регрессионной модели грузооборота железнодорожного транспорта / С. И. Носков, И. П. Врублевский // Вестник транспорта Поволжья. – 2020. – № 1 (79). – С. 86–90.

18. *Носков, С. И.* Применение интервального регрессионного анализа для моделирования объектов транспорта / С. И. Носков, И. П. Врублевский, В. О. Заянчуковская // Вестник Уральского государственного университета путей сообщения. – 2020. – № 3 (47). – С. 45–52.

Базилевский Михаил Павлович – канд. техн. наук, доц., доцент кафедры математики Иркутского государственного университета путей сообщения.
E-mail: mik2178@yandex.ru
ORCID iD: https://orcid.org/0000-0002-3253-5697

DOI: https://doi.org/
Received 29.10.2020
Accepted 02.02.2021

ISSN 1995-5499

BUILDING AND INTERPRETING POWER-EXPONENTIAL REGRESSION MODELS

© 2020 M. P. Bazilevskiy✉

Irkutsk State Transport University
15, Chernyshevskogo Street, 664074 Irkutsk, Russian Federation

Annotation. In our research, we developed and studied new structural specifications for regression models. The article presents two generalisations of the Cobb-Douglas production function: power exponential regression with powers presented in the form of linear combinations of explanatory variables and in the form of linear combinations of natural logarithms of explanatory variables. The approximation qualities of these models are always on a par with those of the Cobb-Douglas production functions. Models of the first type are perfect for predicting the dependent variable values, but the interpretation of their estimates is problematic. For the second type of models it has been proven that the coefficient of elasticity of the regression function for a specific explanatory variable does not depend on its value and is a linear combination of the natural logarithms of the remaining variables. Therefore, the estimates of such models are quite interpretable. In addition, models of the second type with the given values of elasticity coefficients can be used to determine predicted values of the explanatory and the response variable. In our study, we built four regression models for the dependence of the railway passenger volume in the Irkutsk Region on the number of personal vehicles and the cost of train and airplane tickets. The article presents a detailed interpretation of the power exponential regression with powers given in the form of linear combinations of natural logarithms of explanatory variables. The model was used to make graphs of changes in the elasticity of railway passenger volume for each explanatory variable over time. The predicted values of the explanatory and the response variable were obtained.

✉ Bazilevskiy Mikhail P.
e-mail: mik2178@yandex.ru

Keywords: regression model; production function; power exponential regression; Cobb-Douglas function; coefficient of elasticity; railway passenger transport.

CONFLICT OF INTEREST

The author declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. *Westfall P. H., Arias A. L.* Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC, 2020. 514 p.

2. *Pardoe I.* Applied regression modeling. Wiley, 2020. 336 p.

3. *Noskov S. I.* Technology for modeling objects with unstable functioning and uncertainty in data. Irkutsk: Oblinformpechat', 1996. 320 p.

4. *Bazilevskiy M. P., Noskov S. I.* Formalization of the problem of constructing linear multiplicative regression in the form of a partial boolean linear programming problem. Modern technologies. System analysis. Modeling. 2017. V. 55. No. 3. P. 101–105.

5. *Kleyner G. B.* Production functions: theory, methods, application. Moscow: Finance and Statistics, 1986. 239 p.

6. *Kleyner G. B.* Economy. Modeling. Maths. Selected Works. Moscow: – M. : TsEMI RAN, 2016. 856 p.

7. *Gorbunov V. K.* Production functions: theory and construction. Ulyanovsk: UlGU, 2013. 84 p.

8. *Khatskevich G. A., Pronevich A. F., Chaykovskiy M. V.* Two-factor production functions with a given marginal substitution rate. Economic science today. 2019. No. 10. P. 169–181.

9. *Kutenkov R. P.* Production functions: assessing the interchangeability of factors and forecasting the volume of agricultural production in the regions of Russia. Regional agrosystems: economics and sociology. 2019. No. 2. P. 50–57.

10. *Ivanova N. K., Lebedeva S.A., Noskov S. I.* Identification of parameters of some non-smooth regressions. Information technologies and problems of mathematical modeling of complex systems. 2016. No. 17. P. 107–110.

11. *Bazilevskiy M. P., Noskov S. I.* Estimating index regression models using the least modulus method. Bulletin of the Russian New University. Series: Complex Systems: Models, Analysis and Management. 2020. No. 1. P. 17–23.

12. *Bazilevskiy M. P.* OLS-estimation of parameters of two-factor regression models specified on the basis of Leontiev functions. South Siberian Scientific Bulletin. 2019. V. 26. No. 2. P. 66–70.

13. *Dobrovolskaya L. P.* Pseudo-Random Search and Complex Econometric Models. Materials of the XV International Conference Algebra, Number Theory and Discrete Geometry: Modern Problems and Applications. Tula: TGPU im. L. N. Tolstogo, 2018. P. 291–293.

14. *Roubalova L., Viskotova L.* The time augmented Cobb-Douglas production function. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunesis. 2019. V. 67. No. 5. P. 1347–1356.

15. *Bazilevskiy M. P.* Study of exponential regression models with degrees in the form of linear combinations of logarithms of explanatory variables. Transport infrastructure of the Siberian region. 2018. P. 397–401.

16. *Krakovskiy Yu. M., Popova N. N.* Generalized forecasting of the indicator of cargo loading during transportation by rail. Proceedings of Voronezh State University. Series: Systems Analysis and Information Technology. 2020. No. 3. P. 43–50.

17. *Noskov S. I., Vrublevskiy I. P.* Analysis of the regression model of rail freight turnover. Volga Transport Bulletin. 2020. V. 79, No. 1. P. 86–90.

18. *Noskov S. I., Vrublevskiy I. P., Zayanchukovskaya V. O.* Application of interval regression analysis for modeling transport objects. Bulletin of the Ural State Transport University. 2020. V. 47. No. 3. P. 45–52.

Bazilevskiy Mikhail P. — PhD in Technical Sciences, Associate Professor, Irkutsk State Transport University.
E-mail: mik2178@yandex.ru
ORCID iD: https://orcid.org/0000-0002-3253-5697