УДК 519.852
ISSN 1995-5499
DOI: https://doi.org/10.17308/sait.2021.2/3511
Поступила в редакцию 08.04.2021
Подписана в печать 19.07.2021

РЕАЛИЗАЦИЯ КОНКУРСА РЕГРЕССИОННЫХ МОДЕЛЕЙ С ПРИМЕНЕНИЕМ КРИТЕРИЯ СОГЛАСОВАННОСТИ ПОВЕДЕНИЯ

© 2021 С. И. Носков✉

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

Аннотация. В работе описываются основные этапы реализации конкурса регрессионных моделей сложных объектов. Он состоит в построении множества альтернативных вариантов модели заданного класса и последующем выборе лучшего варианта на основе использования векторного критерия оценки адекватности. В его состав, в частности, входят критерии Фишера, Стьюдента, Дарбина — Уотсона, множественной детерминации, смещения, информативности набора независимых переменных, остаточная дисперсия, средние относительные ошибки аппроксимации и прогноза, ширина области определения модели. Общее количество потенциальных вариантов модели при этом может составить десятки тысяч для размерностей, соответствующих реальным ситуациям, возникающим при моделировании. Описываются основные этапы организации конкурса моделей: построение множества допустимых вариантов по их соответствию содержательному смыслу факторов и значениям частных критериев адекватности; максимизация на этом множестве обобщенного критерия согласованности поведения расчетных и фактических значений зависимой переменной; в случае неединственности решения этой задачи реализация на сформированном множестве вариантов модели метода идеальной точки. Полученная в результате проведения конкурса модель будет соответствовать смыслу всех входящих в ее состав переменных и обладать высоким качеством.

Ключевые слова: регрессионная модель, критерии адекватности, методы наименьших квадратов, модулей, антиробастного и смешанного оценивания.

ВВЕДЕНИЕ

Весьма эффективным средством научного анализа сложных объектов самой различной природы является регрессионный анализ. При построении математических моделей данного типа исследователям приходится сталкиваться с целым рядом достаточно сложных проблем. К ним, в частности, относятся: выбор подходящей аппроксимирующей функции независимых переменных с учетом преобразований независимых переменных, метода оценивания неизвестных параметров, способов интерпретации полученных результатов (см., в частности, [1–3]). В ряде работ изучаются различные частные аспекты этих проблем, связанные с использованием критериев нелинейности квазилинейных регрессионных моделей [4], построением хорошо интерпретируемых качественных регрессионных моделей [5], применением расчетных схем оценивания параметров аддитивной степенной регрессии [6], многокритериальным выбором модели [7], выявлением лучшей модели по критерию множественной детерминации с регулируемым эффектом мультиколлинеарности между регрессорами [8]. В рамках регрессионного анализа проводятся активные исследования, касающиеся как разработки методов построения конкретных моделей, так и критериев оценки их качества [9–14].

В настоящей работе развивается описанный в работе [1] подход к организации кон-

✉ Носков Сергей Иванович
e-mail: sergey.noskov.57@mail.ru

курса регрессионных моделей, состоящего в построении множества их альтернативных вариантов большой мощности и последующего выбора наиболее приемлемого их них, руководствуясь векторным критерием оценки адекватности каждого варианта.

ПОСТАНОВКА ЗАДАЧИ

Пусть, исходя либо из содержательных соображений, либо основываясь на глубоких теоретических положениях, исследователь полагает, что поведение зависимой переменной y определяется значениями независимых факторов $x_1,...,x_m$, то есть постулирует регрессионную зависимость

$$y_k = F(\alpha; x_{k1},...,x_{km}) + \varepsilon_k,\ k = \overline{1,n}, \qquad (1)$$

где k — номер наблюдения, n — длина выборки, α — вектор оцениваемых параметров, $\varepsilon_1,...,\varepsilon_n$ — ошибки аппроксимации. При этом всякие предположения об их вероятностной природе отсутствуют, что позволяет оставаться в рамках логико-алгебраического подхода к анализу данных. Аппроксимирующая функция F может быть выбрана из разных классов. Мы для определенности будем полагать, что зависимость (1) представима в виде:

$$y_k = \alpha_0 + \sum_{i=1}^{m} \alpha_i g_{j_i}(x_{ki}) + \varepsilon_k,\ k = \overline{1,n}, \qquad (2)$$

где $g_{j_i}(x_{ki})$ — j-е преобразование i-й независимой переменной, выбранное из набора

$$G(z) = \{g_1(z),...,g_L(z)\}. \qquad (3)$$

В качестве преобразований $g_i(z)$ могут, наряду с тождественным преобразованием $g_1(z) = z$, использоваться элементарные функции $1/z$, $\sin(z)$, $\cos(z)$, z^2, z^3, $\exp(z)$, $\ln(z)$, $1/\exp(z)$ и другие.

Будем предполагать, что для идентификации неизвестных параметров α_i, $i = \overline{1,m}$ исследователь располагает арсеналом из D методов, в который входят, в частности, методы наименьших квадратов [1, 9, 15] и модулей [1], антиробастного [1, 16] и смешанного [17] оценивания.

Таким образом, может быть построено множество Ω^1 альтернативных вариантов уравнения (2):

$$\Omega^1 = \{M_1,...,M_r\},$$

где r — число размещений с повторениями из L элементов по m, умноженное на D:

$$r = DL^m.$$

При построении реальных моделей сложных объектов число r может быть очень большим. Действительно, при четырех независимых переменных в (1) $(m = 4)$, восьми возможных их преобразований $(L = 8)$ и четырех используемых методов оценивания параметров $(D = 4)$ число всех возможных регрессий (2) составит $r = 4 * 9^4 = 26244$.

Каждый из r вариантов модели следует проверить на адекватность, то есть оценить, допустимы ли для него значения критериев адекватности? При этом надо иметь в виду, что в регрессионном анализе разработана целая система таких критериев (см., например, [1, 9, 15]), каждый из которых описывает какую-либо частную характеристику качества описания реального объекта его математической моделью. К ним, в частности, относятся:

– критерий Фишера;
– критерий Стьюдента;
– величина остаточной дисперсии;
– критерий информативности набора независимых переменных;
– критерий Дарбина — Уотсона;
– критерий множественной детерминации;
– критерий смещения;
– средняя относительная ошибка аппроксимации;
– средняя относительная ошибка прогноза;
– ширина области определения уравнения (2).

Пусть в общем случае качество каждого из *r* вариантов модели оценивается по векторному критерию адекватности, содержащему T частных: $K(M) = (K_1(M),...,K_T(M))$, $M \in \Omega^1$. Будем считать, что все критерии приведены к однородному виду, т. е. значение каждого из них тем лучше, чем оно выше. Приемы, позволяющие произвести такую операцию для величины остаточной дисперсии, критериев Дарбина — Уотсона и смещения, а также для средних относительных ошибок аппроксимации и прогноза, описаны в работе [1]. Отметим, что для некоторых вариантов $M \in \Omega^1$ какие-то компоненты вектора $K(M)$ могут быть не определены. Так, например, использование критериев детерминации, Фишера и Дарбина — Уотсона правомерно лишь для метода наименьших квадратов. Это обстоятельство, однако, не влияет на ход последующих рассуждений.

Таким образом, задача построения лучшего варианта, или определения «победителя» конкурса моделей, может быть сформулирована следующим образом: найти такой вариант $M^* \in \Omega^1$, который бы доставлял максимум векторному критерию $K(M)$:

$$M^* = \arg\max_{M \in \Omega^1} K(M). \qquad (4)$$

Отметим, что задача (4) относится к классу задач многокритериальной оптимизации [18] с конечным множеством альтернатив. Такие задачи, как правило, не имеют классического единственного решения и требуют привлечения различных формально и содержательно обоснованных эвристик для выделения лучшей альтернативы.

ФОРМИРОВАНИЕ ДОПУСТИМОГО МНОЖЕСТВА АЛЬТЕРНАТИВНЫХ ВАРИАНТОВ МОДЕЛИ

Займемся теперь выделением из Ω^1 множества допустимых вариантов. Прежде всего, удалим из Ω^1 варианты, не согласующиеся с содержательным смыслом переменных. Приведем прежде небольшой пример. Пусть модель (2) представляет собой производственную функцию [1] и при этом y — выпуск продукции некоторым промышленным объектом, x_1 — стоимость производственных фондов, x_2 — численность работающих. Пусть один из вариантов модели имеет вид

$$y = 2.7 + 3.1x_1 - 4.3x_2 + \varepsilon. \qquad (5)$$

Уравнение (5) следует признать неприемлемым, поскольку, в соответствии со смыслом переменных, все его параметры, за исключением свободного члена, должны быть положительны.

В общем случае исследователь должен на начальном этапе проведения конкурса задать вектор $\lambda = (\lambda_1, ..., \lambda_m)$, где

$$\lambda_i = \begin{cases} 1, & \textit{если } i\textit{-я переменная должна входить в уравнение с положительным коэффициентом} \\ 0, & \textit{если знак } i\textit{-го коэффициента безразличен} \\ -1, & \textit{если } i\textit{-я переменная должна входить в уравнение с отрицательным коэффициентом.} \end{cases}$$

Тогда множество вариантов $\Omega^2 \subseteq \Omega^1$, допустимых по смыслу входящих в модель переменных, формируется по правилу:

$$M \in \Omega^2 \Leftrightarrow \lambda_i \alpha_i(M) \ge 0, \ i = \overline{1, m},$$

где $\alpha\,(M)$ — вектор параметров уравнения (2), соответствующий варианту M. Разумеется, для преобразования $1/z$ должно выполняться обратное неравенство.

Сформируем теперь множество вариантов $\Omega^3 \subseteq \Omega^2$, допустимых по значениям критериев. Такое требование тоже является вполне естественным, поскольку, например, явно недопустимыми являются варианты с меньшей, чем 0.7, детерминацией или большей, чем 10 %, средней относительной ошибкой аппроксимации. Поэтому также на начальном этапе проведения конкурса исследователь должен задать нижнюю $\underline{K_i}$ и верхнюю $\overline{K_i}$ границы допустимости для каждого варианта, $i = \overline{1, T}$. Тогда множество вариантов, допустимых по значениям критериев, формируется по правилу:

$$M \in \Omega^3 \Leftrightarrow \underline{K_i} \le K_i(M) \le \overline{K_i}, \ i = \overline{1, T}.$$

ОПРЕДЕЛЕНИЕ ЛУЧШЕГО ВАРИАНТА МОДЕЛИ ПО КРИТЕРИЮ СОГЛАСОВАННОСТИ ПОВЕДЕНИЯ

Практически все перечисленные выше критерии адекватности включают в состав соответствующих расчетных формул ошибки аппроксимации ε_k, $k = \overline{1, n}$. Вместе с тем, существует весьма важный и эффективный критерий (подробно рассмотренный в работах [19, 20]), не связанный с ними, а отражающий степень согласованности в поведении фактических y_k и расчетных $y_k^*(M)$ для варианта M значений зависимой переменной, где

$$y_k = y_k^*(M) + \varepsilon_k(M),\ k = \overline{1,n}.$$

Обобщенная форма представления такого критерия согласованности поведения (ОКСП) имеет вид:

$$\Phi(M) = \sum_{k=1}^{n-1} \sum_{s=k+1}^{n} sign[(y_k^*(M) - y_s^*(M))(y_k - y_s)],$$

где операция $sign(a)$ задается по правилу:

$$sign(a) = \begin{cases} 1, & \text{если } a > 0 \\ 0, & \text{если } a = 0 \\ -1, & \text{если } a < 0. \end{cases}$$

Смысл ОКСП поясним на небольшом примере. Пусть для произвольных номеров наблюдений k и s имеют место неравенства:

$$y_k^*(M) < y_s^*(M),\ y_k > y_s,$$

т. е. $sign[(y_k^*(M) - y_s^*(M))(y_k - y_s)] = -1$. Это означает, что на паре номеров наблюдений (k,s) модель (1) плохо объясняет исследуемый процесс даже при возможно малых значениях ошибок аппроксимации $|\varepsilon_k|$ и $|\varepsilon_s|$. Именно подобные ситуации и призван выявлять данный критерий. Легко видеть, что он принимает значения на отрезке $[n(1-n)/2, n(n-1)/2]$. При этом чем ближе положительные значения $\Phi(M)$ к его правой границе, тем в большей степени согласованы фактические и расчетные значения зависимой переменной модели. При отрицательных значениях $\Phi(M)$ можно говорить об их несогласованности.

Победителем конкурса моделей назначается вариант M^*, на котором ОКСП достигает максимума:

$$M^* = \arg\max_{M \in \Omega^3} \Phi(M). \qquad (6)$$

Вместе с тем, поскольку критерий $\Phi(M)$ имеет дискретный характер, могут иметь место ситуации, когда его максимальное значение достигается не на единственном варианте. В этом случае формируется множество Ω^4 по правилу:

$$\Omega^4 = \{M \in \Omega^3 \,|\, \Phi(M) = \Phi(M^*)\}.$$

Для выделения единственного варианта из множества Ω^4 воспользуемся методом идеальной точки (см., например, [1]), состоящий в данном случае в следующем.

Прежде всего, значения векторов $K(M)$, $M \in \Omega^4$ нормируются по правилу:

$$\tilde{K}_i(M) = \frac{K_i(M) - K_i^-}{K_i^+ - K_i^-},\ i = \overline{1,T},$$

где $K_i^- = \min_{M \in \Omega^4} K_i(M)$, $K_i^+ = \max_{M \in \Omega^4} K_i(M)$,

Очевидно, что $0 \le \tilde{K}_i(M) \le 1$, $i = \overline{1,T}$.

Таким образом, «идеальная» точка представляет собой вектор $\tilde{K}^*$, каждая компонента которого равна максимальному значению соответствующего нормированного критерия, то есть единице. Необходимо выделить вариант $M^{**} \in \Omega^4$, образ которого в критериальном пространстве наиболее близок в некоторой метрике (например, евклидовой) к вектору $\tilde{K}^*$:

$$M^{**} = \arg\min \sum_{j=1}^{T} \left(1 - \tilde{K}_j(M)\right)^2.$$

Таким образом, победителем конкурса моделей в указанном классе аппроксимирующих функций вида (2) является вариант M^*для случая, когда задача (6) имеет единственное решение, и вариант M^{**} в противном случае.

ПРИМЕР

Пусть, как и в случае с построением модели (5), исследователь на основе статистической информации за 20 лет $(n = 20)$ разрабатывает производственную функцию для некоторого промышленного объекта в виде регрессионной модели (2), т. е. $m = 2$. При этом используется набор преобразований

$$G(z) = \{z, z^2, \ln(z)\},$$

т. е. $L = 3$ и один метод оценивания параметров — наименьших квадратов $(D = 1)$. Для оценки адекватности модели задействованы три критерия $(T = 3)$:

K_1 — критерий множественной детерминации;

K_2 — средняя относительная ошибка аппроксимации (в %);

K_3 — средняя относительная ошибка прогноза (в %).

Для каждого критерия заданы границы допустимости $\underline{K}_i$ и $\overline{K}_i$, $i = \overline{1,3}$:

$$K_1 \in [0.85, 1],\quad K_2 \in [0, 5],\quad K_2 \in [0, 7].$$

Отметим, что, в соответствии со смыслом независимых переменных, знаки параметров при них в модели (2) должны быть положительны, т. е. $\lambda = (1,1)$.

Было сформировано множество альтернативных вариантов модели (2):

$$\Omega^1 = \{M_1, ..., M_9\},$$

(поскольку $r = 3^2 = 9$), где:

$M_1: \ y = 2.7 + 3.1x_1 - 4.3x_2,$

$K_1(M_1) = 0.82, K_2(M_1) = 4.9,$

$K_3(M_1) = 5.7, \Phi(M_1) = 169,$

$M_2: y = 1.2 + 2.1x_1 + 0.21x_2^2,$

$K_1(M_2) = 0.87, K_2(M_2) = 3.8,$

$K_3(M_2) = 5.4, \Phi(M_2) = 166,$

$M_3: \ y = 5.8 + 3.6x_1 + 12.2\ln(x_2)$

$K_1(M_3) = 0.91, K_2(M_3) = 3.6,$

$K_3(M_3) = 5.7, \Phi(M_3) = 147,$

$M_4: \quad y = 4.1 - 1.6x_1^2 + 6.5x_2,$

$K_1(M_4) = 0.89, K_2(M_4) = 4.3,$

$K_3(M_4) = 3.8, \Phi(M_4) = 174,$

$M_5: \quad y = 3.1 + 10.2\ln(x_1) + 0.4x_2,$

$K_1(M_5) = 0.79, K_2(M_5) = 8.3,$

$K_3(M_5) = 7.1, \Phi(M_5) = 148,$

$M_6: \quad y = 0.9 + 7.7\ln(x_1) + 0.22x_2^2,$

$K_1(M_6) = 0.87, K_2(M_6) = 6.2,$

$K_3(M_6) = 3.9, \Phi(M_6) = 153,$

$M_7: \quad y = 5.4 + 1.7x_1^2 + 6.9\ln(x_2),$

$K_1(M_7) = 0.97, K_2(M_7) = 4.4,$

$K_3(M_7) = 6.1, \Phi(M_7) = 156,$

$M_8: \quad y = 0.8 + 1.6x_1^2 - 2.3x_2^2,$

$K_1(M_8) = 0.92, K_2(M_8) = 4.7,$

$K_3(M_8) = 2.1, \Phi(M_8) = 177,$

$M_9: \quad y = 11.3 - 6.4\ln(x_1) + 5.7\ln(x_2),$

$K_1(M_9) = 0.93, K_2(M_9) = 4.8,$

$K_3(M_9) = 4.1, \Phi(M_9) = 166.$

Сформируем множество вариантов модели Ω^2, удалив из Ω^1 варианты со значениями параметров, не соответствующими смыслу входящих в модель переменных:

$$\Omega^2 = \{M_2, M_3, M_5, M_6, M_7\}.$$

Затем сформируем множество вариантов Ω^3, исключив из Ω^2 варианты с недопустимыми значениями критериев адекватности:

$$\Omega^3 = \{M_2, M_3, M_7\}.$$

Победителем конкурса моделей является вариант M_2, обладающий максимальным значением обобщенного критерия согласованности поведения, $\Phi(M_2) = 166$.

ЗАКЛЮЧЕНИЕ

В работе рассмотрены основные этапы проведения конкурса регрессионных моделей сложных объектов, заключающегося в формировании множества альтернативных вариантов модели в заданном классе аппроксимирующих функций и последующего выбора лучшего варианта на основе использования совокупности частных критериев оценки его адекватности. В качестве таковых могут быть задействованы критерии Фишера, Стьюдента, Дарбина — Уотсона, множественной детерминации, смещения, информативности набора независимых переменных, остаточная дисперсия, средние относительные ошибки аппроксимации и прогноза, ширина области определения модели. Общее количество всех построенных вариантов модели для реальных данных при этом может быть очень большим — до нескольких десятков тысяч. Приводится описание основных этапов проведения конкурса моделей: построение множества допустимых вариантов как по их соответствию содержательному смыслу факторов, так и значениям локальных критериев адекватности; максимизация на этом множестве обобщенной согласованности в поведении расчетных и фактических значений зависимой переменной. В случае, если решение этой задачи не является единственным, предложено на последней стадии конкурса использовать известный в теории принятия решений метод идеальной точки. Полученная в результате проведения конкурса модель удовлетворяет смыслу всех входящих в ее состав переменных и обладает высоким качеством.

Значительный интерес представляет разработка в будущем программного комплекса, в котором будет реализован предлагаемый в работе подход к реализации конкурса регрессионных моделей.

КОНФЛИКТ ИНТЕРЕСОВ

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

СПИСОК ЛИТЕРАТУРЫ

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

2. *Носков С. И., Потороченко Н. А.* Диалоговая система реализации «конкурса» регрессионных зависимостей // Управляющие системы и машины. – 1992. – №2-4. – С. 111–116.

3. *Базилевский М. П., Носков С. И.* Технология организации конкурса регрессионных моделей // Информационные технологии и проблемы математического моделирования сложных систем. – 2009. – № 7. – С. 77–84.

4. *Базилевский М. П.* Критерии нелинейности квазилинейных регрессионных моделей // Моделирование, оптимизация и информационные технологии. – 2018. – Т. 6, № 4 (23). – С. 185–195.

5. *Базилевский М. П.* Фундаментальный блок алгоритмов построения хорошо интерпретируемых качественных регрессионных моделей // Информационные технологии и математическое моделирование в управлении сложными системами. – 2020. – № 3 (8). – С. 1–10.

6. *Базилевский М. П.* Разработка и исследование алгоритмов оценивания параметров аддитивной степенной регрессии // Современные технологии. Системный анализ. Моделирование. – 2017. – № 4 (56). – С. 131–138.

7. *Базилевский М. П.* Программно-математическое обеспечение автоматизации многокритериального выбора регрессионных моделей: автореф… дис. кан. тех. наук. – Иркутск : 2012. – 27 с.

8. *Базилевский М. П.* Алгоритм выбора лучшей по коэффициенту детерминации регрессии с регулируемым эффектом мульти-коллинеарности между регрессорами // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – № 15. – С. 31–38.

9. *Елисеева И. И., Курышева С. В., Костеева Т. В. и др.* Эконометрика. – М. : Финансы и статистика, 2007. – 576 с.

10. *Fox W. P., Bauldry W. C.* Advanced Problem Solving Using Maple. LocationBoca Raton. – 2020. 404 p. DOI https://doi.org/10.1201/9780429469626.

11. *Kissell R. L.* Algorithmic Trading Methods (Second Edition) // Applications Using Advanced Statistics, Optimization, and Machine Learning Techniques. – 2021. – P. 151–173.

12. *Pardoe I.* Applied Regression Modeling. John Wiley & Sons, Inc. – 2021. – 310 p.

13. *Guerard J. B., Gultekin A. S.* Regression Analysis and Estimating Regression Models // Quantitative Corporate Finance. – P. 263–300.

14. *Westfall P. H., Arias A. L.* Understanding Regression Analysis. A Conditional Distribution Approach. LocationBoca Raton. – 2020. – 514 p.

15. *Дрейпер Н, Смит Г.* Прикладной регрессионный анализ, 3-е изд. – Вильямс, 2016. – 912 с.

16. *Носков С. И.* Метод антиробастного оценивания параметров линейной регрессии: число максимальных по модулю ошибок аппроксимации // Южно-Сибирский научный вестник. – 2020. – № 1 (29). – С. 51–54.

17. *Носков С. И.* О методе смешанного оценивания параметров линейной регрессии // Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – № 1 (2). – С. 41–45.

18. *Лотов А. В., Бушенков В. А., Каменев Г. К., Черных О. Л.* Компьютер и поиск компромисса. Метод достижимых целей. – М. : Наука, 1997. – 239 с.

19. *Носков С. И.* Построение эконометрических зависимостей с учетом критерия «согласованность поведения» // Кибернетика и системный анализ. – 1994. – № 1. – С. 177.

20. *Носков С. И.* Обобщенный критерий согласованности поведения в регрессионном

анализе // Информационные технологии и математическое моделирование в управлении сложными системами. – 2018. – № 1 (1). – С. 14–20.

Носков Сергей Иванович — д-р техн. наук, проф., профессор кафедры «Информационные системы и защита информации» Иркутского государственного университета путей сообщения.
E-mail: sergey.noskov.57@mail.ru
ORCID iD: https://orcid.org/0000-0003-4097-2720

DOI: https://doi.org/10.17308/sait.2021.2/3511 ISSN 1995-5499
Received 08.04.2021
Accepted 19.07.2021

IMPLEMENTATION OF THE COMPETITION OF REGRESSION MODELS USING THE CRITERION OF CONSISTENCY OF BEHAVIOR

© 2021 S. I. Noskov✉

Irkutsk State University of Railways
15, Chernyshevsky Street, 664074 Irkutsk, Russian Federation

Annotation. The paper describes the main stages of the implementation of the competition of regression models of complex objects. It consists in constructing a set of alternative options for a model of a given class and then choosing the best option based on the use of a vector criterion for assessing adequacy. It includes, in particular, the Fisher, Student, Durbin-Watson criteria, multiple determination, bias, informativeness of a set of independent variables, residual variance, average relative errors of approximation and forecast, width of the domain of the model. In this case, the total number of potential variants of the model can amount to tens of thousands for dimensions corresponding to real situations that arise during modeling. The main stages of organizing a competition of models are described: construction of a set of admissible options according to their compliance with the meaningful meaning of factors and the values of particular criteria of adequacy; maximization on this set of the generalized criterion for the consistency of the behavior of the calculated and actual values of the dependent variable; in the case of non-uniqueness of the solution to this problem, the implementation of the ideal point method on the generated set of variants of the model. The model obtained as a result of the competition will correspond to the meaning of all its constituent variables and be of high quality.
Keywords: regression model, adequacy criteria, least squares methods, moduli, anti-robust and mixed estimation.

CONFLICT OF INTEREST

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

✉ Noskov Sergey I.
e-mail: sergey.noskov.57@mail.ru

REFERENCES

1. *Noskov S. I.* (1996) Object modeling technology with unstable operation and data uncertainty, Federal State Public Educational Establishment of Higher Training Eastern Siberia Institute of the Ministry of the Interior of the Russian Federation, 110, Lermontov Street, Irkutsk. 320 p.

2. *Noskov S., Potorochenko N.* (1992) System for the implementation of the “competition” of regression dependencies // Control systems and machines. No 2/4. P. 111–116.

3. *Bazilevsky M. P., Noskov S. I.* (2009) Technology of organizing a competition of regression models // Information technologies and problems of mathematical modeling of complex systems. No 7. P. 77–84.

4. *Bazilevsky M. P.* (2018) Nonlinearity Criteria for Quasilinear Regression Models // Modeling, Optimization and Information Technologies. V. 6. No 4 (23). P. 185–195.

5. *Bazilevsky M. P.* (2020) Fundamental block of algorithms for constructing well-interpreted qualitative regression models // Information technologies and mathematical modeling in the management of complex systems. No 3 (8). P. 1–10.

6. *Bazilevsky M. P.* (2017) Development and research of algorithms for estimating the parameters of additive power regression // Modern technologies. System analysis. Modeling. No 4 (56). P. 131–138.

7. *Bazilevsky M. P.* (2012) Software and mathematical support for automation of multicriteria choice of regression models: abstract of ... dis. can. those. sciences. Irkutsk. 27 p.

8. *Bazilevsky M. P.* (2016) Algorithm for choosing the best regression in terms of the coefficient of determination with an adjustable effect of multi-collinearity between regressors // Information technologies and problems of mathematical modeling of complex systems/ No 15. P. 31–38.

9. *Eliseeva I. I., Kurysheva S. V., Kosteeva T. V. and others* (2007) Econometrics. Moscow: Finance and Statistics. 576 p.

10. *Fox W. P., Bauldry W. C.* (2020) Advanced Problem Solving Using Maple. LocationBoca Raton. 404 p. DOI https://doi.org/10.1201/9780429469626.

11. *Kissell R. L.* (2021) Algorithmic Trading Methods (Second Edition) // Applications Using Advanced Statistics, Optimization, and Machine Learning Techniques. P. 151–173.

12. *Pardoe I.* (2021) Applied Regression Modeling. John Wiley & Sons, Inc. 310 p.

13. *Guerard J. B., Gultekin A. S.* Regression Analysis and Estimating Regression Models // Quantitative Corporate Finance. P. 263–300.

14. *Westfall P. H., Arias A. L.* (2020) Understanding Regression Analysis. A Conditional Distribution Approach. LocationBoca Raton. 514 p.

15. *Draper N, Smith G.* (2016) Applied Regression Analysis, 3rd ed. Williams. 912 p.

16. *Noskov S. I.* (2020) Method of antirobast estimation of linear regression parameters: number of maximum on the module of approximation errors // South – Siberian scientific bulletin. No 1(29). P. 51–54.

17. *Noskov S. I.* (2019) About the method of mixed estimation of parameters of linear regression // Information technology and mathematical modeling in the management of complex systems. N. 1. P. 14–20.

18. *Lotov A. V., Bushenkov V. A., Kamenev G. K., Chernykh O. L.* (1997) The computer and the search for a compromise. Method of achievable goals. Moscow : Nauka. 239 p.

19. *Noskov S. I.* (1994) Construction of econometric dependencies taking into account the criterion of “consistency of behavior” // Cybernetics and Systems Analysis. No 1. P. 177.

20. *Noskov S. I.* (2018) Generalized criterion of coordination of behavior inregression analysis // Information technology and mathematical modeling in the management of complex systems. No 1. P. 14–20.

Noskov Sergey I. — Doctor of Technical Sciences, Professor, Professor of the Department of Information Systems and Information Security, Irkutsk State University of Railways.
E-mail: sergey.noskov.57@mail.ru
ORCID iD: https://orcid.org/0000-0003-4097-2720