УДК 519.862.6 doi: 10.26102/2310-6018/2018.23.4.015

М. П. Базилевский

КРИТЕРИИ НЕЛИНЕЙНОСТИ КВАЗИЛИНЕЙНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Иркутский государственный университет путей сообщения, Иркутск, Россия

Часто при выборе наилучшей в некотором заданном смысле регрессионной модели таковой оказывается нелинейная регрессия. Например, при реализации технологии «конкурса» моделей наилучшей может оказаться регрессия, относящаяся к классу квазилинейных. Достоинством квазилинейных регрессий является возможность их оценивания с помощью обычного метода наименьших квадратов. Но полученным при этом оценкам параметров квазилинейной модели редко удается дать какую-либо содержательную интерпретацию. В итоге построенную регрессию при условии, что она обладает высоким качеством аппроксимации, можно использовать только для получения прогнозов, что существенно снижает её практическую значимость. Данная работа посвящена проблеме оценивания степени нелинейности квазилинейных регрессионных моделей. На основе коэффициента Джини разработан критерий нелинейности по площади. Также представлен его аналог – критерий нелинейности по длине. Данные критерии нелинейности позволяют оценивать степень нелинейности как однофакторных, так и многофакторных квазилинейных регрессий. Показано, каким образом при низкой степени нелинейности можно интерпретировать параметры квазилинейной регрессии. Рассмотрен конкретный численный пример оценивания степени нелинейности однофакторных квазилинейных регрессий. Разработанные критерии можно использовать при организации технологии «конкурса» моделей для контроля степени их нелинейности.

Ключевые слова: квазилинейная регрессия, «конкурс» моделей, интерпретация регрессии, кривая Лоренца, коэффициент Джини, критерий нелинейности.

Введение. В регрессионном моделировании одним из ключевых является этап спецификации модели, предполагающий выбор состава и формы связи между переменными в уравнении регрессии [1]. Качественным способом решения этой проблемы является метод построения всех возможных регрессий [2], алгоритм которого предусматривает перебор всех альтернативных вариантов моделей и выбор лучшей из них либо по одному заданному критерию адекватности, либо по нескольким критериям сразу. В последнем случае технология многокритериального выбора получила название «конкурс» моделей, описание которого можно найти в работах [3, 4, 5]. Первый этап алгоритма «конкурса» регрессионных моделей предусматривает формирование множества альтернативных вариантов регрессий, что достигается, например, за счет нелинейных преобразований переменных. Тогда выбранная в результате перебора регрессионная модель, скорее всего, хоть и будет иметь высокое качество аппроксимации, но при этом может оказаться в значительной степени нелинейной, что затрудняет или делает вовсе невозможной интерпретацию её коэффициентов. А такой недостаток, как невозможность интерпретации коэффициентов регрессии, при решении конкретной прикладной задачи иногда чуть ли не полностью лишает модель практической значимости.

Таким образом, при построении регрессионной модели в погоне за высоким качеством аппроксимации не следует забывать о степени её нелинейности, влияющей на интерпретационные свойства. Аналогичную рекомендацию для выбора спецификации регрессии можно найти в работе [1] известного специалиста в области анализа данных С.А. Айвазяна, который утверждает, что: «не следует гнаться за чрезмерной сложностью модели, ориентируясь при этом на минимизацию выборочного критерия адекватности». Иными словами, при построении регрессии следует добиваться компромисса между её сложностью, точностью и трактовкой. На практике, как правило, с увеличением сложности модели её точность повышается, но интерпретационные свойства снижаются: нелинейные регрессии точнее, чем линейные, но их труднее трактовать.

Чтобы находить компромисс между сложностью, точностью и трактовкой регрессионной модели необходимо использовать соответствующие критерии адекватности [3, 6] указанных характеристик. Для оценки точности регрессии, т.е. её аппроксимационных качеств, можно использовать, например, хорошо известный критерий детерминации или его аналоги, среднюю относительную ошибку аппроксимации и другие [6]. Для оценки степени трактовки регрессии, т.е. её интерпретационных качеств, в настоящее время никаких критериев не существует. Возможно, что в таких критериях и вовсе нет смысла, поскольку степень трактовки регрессии будет зависеть от критерия её сложности. Стоит отметить, что понятие «сложность» модели достаточно многогранное. Под сложностью модели можно понимать, например, громоздкость спецификации регрессии. В этом случае сложность модели можно оценить с помощью информационных критериев Акаике, Шварца и Ханнана-Куинна [6]. В работе [1], со ссылкой на работу С.А. Смоляка [7], сложность модели формализуется в виде функционалов гладкости, которые устроены так, что чем более гладкой является функция регрессии, тем меньшее числовое значение они принимают. Недостатком этих критериев является то, что они являются абсолютными и справедливы только для однофакторных моделей. Подробное описание принципа максимальной гладкости в задаче интерполяции можно найти в работе [8]. К сожалению, среди существующих критериев сложности регрессионных моделей автору не удалось найти такие, которые бы показывали относительную степень нелинейности многофакторных регрессий.

Целью данной работы является разработка и исследование новых критериев адекватности, позволяющих количественно оценивать степень нелинейности квазилинейных регрессионных моделей.

Критерии нелинейности. Главным назначением относительного критерия нелинейности должно быть оценивание степени нелинейности регрессии, т.е. он должен количественно отражать, насколько сильно данная нелинейная модель разнится с линейной зависимостью. В основу создания такого критерия легла идея, согласно которой в экономике вычисляется, так называемый, коэффициент Джини [9]. Коэффициент Джини был предложен итальянским статистиком и демографом Коррадо Джини еще в 1912 году.

Стоит заметить, что немного ранее в 1905 году американский экономист Макс Отто Лоренц предложил кривую Лоренца [10], которая показывает неравенство распределения доходов населения. Кривая Лоренца представляет собой выпуклую вниз линию, соединяющую точки с координатами (0,0) и (1,1), в системе координат (x, y), где x – численность населения (доли), y – совокупный доход (доли). На Рисунке 1 кривая $g(x)$ есть кривая Лоренца. Например, если бы кривая Лоренца проходила через точку с координатами (0.25,0.1), то эту точку можно было интерпретировать так: 25% жителей получают только 10% совокупного дохода. Понятно, что чем сильнее выражено неравенство в распределении доходов, тем ближе изгиб кривой Лоренца к точке (1,0). А в случае равномерного распределения доходов кривая Лоренца трансформируется в гипотетическую кривую, называемую прямой абсолютного равенства. На Рисунке 1 кривая $\varphi(x)$ есть прямая абсолютного равенства.

Рисунок - 1. Кривая Лоренца и прямая абсолютного равенства

Для количественного измерения степени неравенства доходов по кривой Лоренца и разработан коэффициент Джини, который равен отношению площади S_A фигуры, ограниченной прямой абсолютного равенства $\varphi(x)$ и кривой Лоренца $g(x)$, к площади $S_\Delta = S_A + S_B$ треугольника, образованного прямой абсолютного равенства $\varphi(x)$:

$$G = S_A / S_\Delta . \qquad (1)$$

Коэффициент Джини принимает значения от 0 до 1. Чем выше неравенство в распределении доходов, тем ближе коэффициент Джини к 1. А чем выше равенство в распределении доходов, тем меньше данный коэффициент.

Если отбросить экономическое содержание задачи, то станет ясно, что коэффициент Джини при определенных условиях представляет собой не что иное, как критерий нелинейности функции $g(x)$. Если функция $g(x)$ проходит через точки (0,0) и (1,1), является непрерывной, монотонно возрастающей и выпуклой вниз на отрезке $[0,1]$, то на этом отрезке коэффициент Джини (1) показывает меру близости этой функции с прямой линией $\varphi(x)$, проходящей через те же точки. Применим теперь эту идею в поле регрессионного анализа.

Рассмотрим однофакторную модель квазилинейной регрессии:

$$y_i = \alpha_0 + \alpha_1 f_s(x_i) + \varepsilon_i, \qquad i = \overline{1,n}, \qquad s \in \{1,2,\ldots,m\}, \qquad (2)$$

где y_i, $i = \overline{1,n}$ – значения зависимой (объясняемой, выходной) переменной; $x_i > 0$, $i = \overline{1,n}$ – значения независимой (объясняющей, входной) переменной; ε_i, $i = \overline{1,n}$ – ошибки аппроксимации; α_0, α_1 – неизвестные параметры модели; n – количество наблюдений; f_s – нелинейное преобразование с номером s, выбранное из набора преобразований $F(x) = \{f_1(x),\ldots,f_m(x)\}$, в качестве которых выступают непрерывные и монотонные на отрезке $x \in [x_{\min}, x_{\max}]$ ($x_{\min} = \min\{x_i\}, x_{\max} = \max\{x_i\}, i = \overline{1,n}$) функции, например, элементарные: x, x^{-1}, $\sqrt{x}$, x^2, x^3, $\ln x$ и т.д.

Достоинством квазилинейных регрессий (2) является возможность их оценивания с помощью обычного метода наименьших квадратов (МНК). К недостаткам относится сложность интерпретации построенных моделей [6].

Оцененная с помощью МНК модель (2) имеет вид:

$$y^* = \alpha_0^* + \alpha_1^* f_s(x), \qquad (3)$$

где α_0^*, α_1^* – МНК-оценки неизвестных параметров модели (2).

В выражении (3) введем обозначение: $g(x) = \alpha_1^* f_s(x)$. Тогда очевидно, что если функция $f_s(x)$ непрерывна и монотонна на отрезке $x \in [x_{\min}, x_{\max}]$, то функция $g(x)$ на этом отрезке тоже непрерывна и монотонна. При этом график функции $g(x)$ проходит через точки $M\left(x_{\min},\alpha_1^* f_s(x_{\min})\right)$ и $N\left(x_{\max},\alpha_1^* f_s(x_{\max})\right)$. График функции $g(x)$ представляет собой аналог кривой Лоренца.

Проведем через те же точки M и N прямую $\varphi(x)$, являющуюся аналогом прямой абсолютного равенства. Уравнение этой прямой имеет вид:

$$\varphi(x)=kx+b, \tag{4}$$

где $k=\alpha_1^* \dfrac{f_s(x_{\max})-f_s(x_{\min})}{x_{\max}-x_{\min}}$, $b=\alpha_1^*\left(f_s(x_{\min})-x_{\min}\dfrac{f_s(x_{\max})-f_s(x_{\min})}{x_{\max}-x_{\min}}\right)$.

Тогда площадь фигуры S_A, ограниченной прямой $\varphi(x)$ и кривой $g(x)$:

$$S_A=\left|\int_{x_{\min}}^{x_{\max}}\left(\varphi(x)-g(x)\right)dx\right|=\left|\int_{x_{\min}}^{x_{\max}}\left(kx+b-\alpha_1^* f_s(x)\right)dx\right|=$$

$$=\left|k\frac{x_{\max}^2-x_{\min}^2}{2}+b\left(x_{\max}-x_{\min}\right)-\alpha_1^*\int_{x_{\min}}^{x_{\max}} f_s(x)dx\right|; \tag{5}$$

а площадь треугольника, ограниченного на отрезке $x\in[x_{\min},x_{\max}]$ прямой $\varphi(x)$:

$$S_\Delta=\frac{1}{2}\left|\alpha_1^*\left(x_{\max}-x_{\min}\right)\left(f_s(x_{\max})-f_s(x_{\min})\right)\right|. \tag{6}$$

Для квазилинейной регрессии (3), по аналогии с коэффициентом Джини (1), введем критерий нелинейности (Nonlinear Criterion) по площади:

$$NC_S=S_A/S_\Delta, \tag{7}$$

где площади S_A и S_Δ находятся по формулам (5) и (6) соответственно.

Критерий нелинейности (7) принимает значения от 0 до 1. Если $NC_S=0$, то квазилинейная регрессия (3) является линейной. Чем ближе критерий NC_S к 1, тем сильнее квазилинейная регрессия отличается от линейной, т.е. растет степень её нелинейности. Будем считать, что если $NC_S\le 0{,}1$, то квазилинейная регрессия практически не отличается от линейной.

Аналогично введем критерий нелинейности по длине дуги:

$$NC_L=l_\varphi/l_g, \tag{8}$$

где $l_\varphi=\sqrt{\left(x_{\max}-x_{\min}\right)^2+\left(\alpha_1^*\right)^2\left(f_s(x_{\max})-f_s(x_{\min})\right)^2}$ – длина прямой $\varphi(x)$ на отрезке $x\in[x_{\min},x_{\max}]$;

$l_g = \int\limits_{x_{\min}}^{x_{\max}} \sqrt{1+\left(g'(x)\right)^2}dx$ – длина кривой $g(x)$ на отрезке $x \in \left[x_{\min}, x_{\max}\right]$.

Критерий нелинейности (8) принимает значения от 0 до 1. Если $NC_L = 1$, то квазилинейная регрессия (3) является линейной. Чем ближе критерий NC_L к 0, тем сильнее квазилинейная регрессия отличается от линейной, т.е. растет степень её нелинейности. Будем считать, что если $NC_L \geq 0{,}9$, то квазилинейная регрессия практически не отличается от линейной.

Нетрудно заметить, что разработанные критерии (7) и (8) можно распространить и на случай многофакторной модели квазилинейной регрессии.

Таким образом, критерии (7) и (8) являются относительными показателями нелинейности квазилинейных регрессий, т.е. они отражают степень отличия данной модели от линейной зависимости. В случае слабой нелинейности, т.е. при $NC_S \approx 0$ или $NC_L \approx 1$, можно считать, что кривая $g(x) = \alpha_1^* f_s(x)$ практически не отличается от прямой $\varphi(x) = kx + b$ на отрезке $x \in \left[x_{\min}, x_{\max}\right]$. Тогда вместо не интерпретируемого коэффициента α_1^* можно интерпретировать угловой коэффициент k прямой (4), который показывает, что с изменением переменной x на 1 единицу, y меняется в среднем на k единиц. Очевидно, что если квазилинейная регрессия (3) является линейной, т.е. критерии нелинейности $NC_S = 0$ или $NC_L = 1$, то коэффициент k прямой (4) будет равен коэффициенту α_1^* линейной регрессии (3), поэтому не будет разницы в их интерпретации. К сожалению, в случае сильно выраженной нелинейности интерпретировать коэффициенты квазилинейной регрессии указанным способом не корректно.

При нарушении условия монотонности функции $g(x)$ пользоваться критерием нелинейности по площади (7) необходимо с осторожностью, поскольку в этом случае кривая $g(x)$ и прямая $\varphi(x)$ на отрезке $x \in \left[x_{\min}, x_{\max}\right]$ могут пересекаться в одной или нескольких точках, поэтому для вычисления определенного интеграла (5) может потребоваться разбиение его на сумму нескольких интегралов. Критерий нелинейности по дуге (8) лишен этого недостатка, но при этом его сложнее вычислять, чем критерий нелинейности по площади (7).

Пример. В Таблице 1 приведены случайно сгенерированные статистические данные по двум переменным y и x.

Таблица 1 – Статистические данные

y	0,66	0,74	1,22	1,75	0,63	1,76	1,94	1,17	0,82	1
x	0,27	0,14	0,91	0,9	0,3	0,95	0,99	0,84	0,07	0,54

Задан набор преобразований переменной x: $F(x)=\{x, x^2, x^3, 1/x, \ln x\}$. Оцененные с помощью МНК линейная и 4 квазилинейных регрессии и их критерии детерминации R^2 имеют вид:

$$y^* = 0{,}4768 + 1{,}1712x, \qquad R^2 = 0{,}7671, \tag{9}$$

$$y^* = 0{,}6489 + 1{,}1053x^2, \qquad R^2 = 0{,}8451, \tag{10}$$

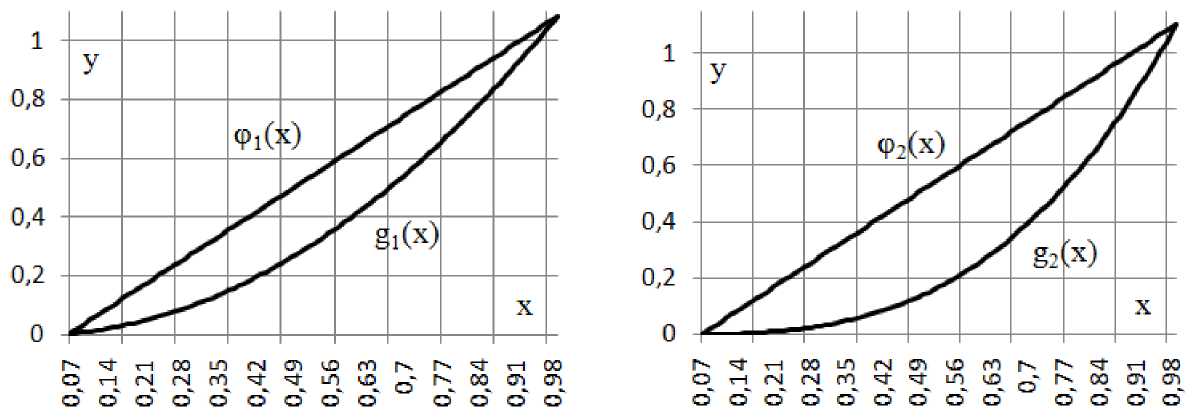
$$y^* = 0{,}7015 + 1{,}1374x^3, \qquad R^2 = 0{,}8735, \tag{11}$$

$$y^* = 1{,}3895 - 0{,}0616/x, \qquad R^2 = 0{,}2826, \tag{12}$$

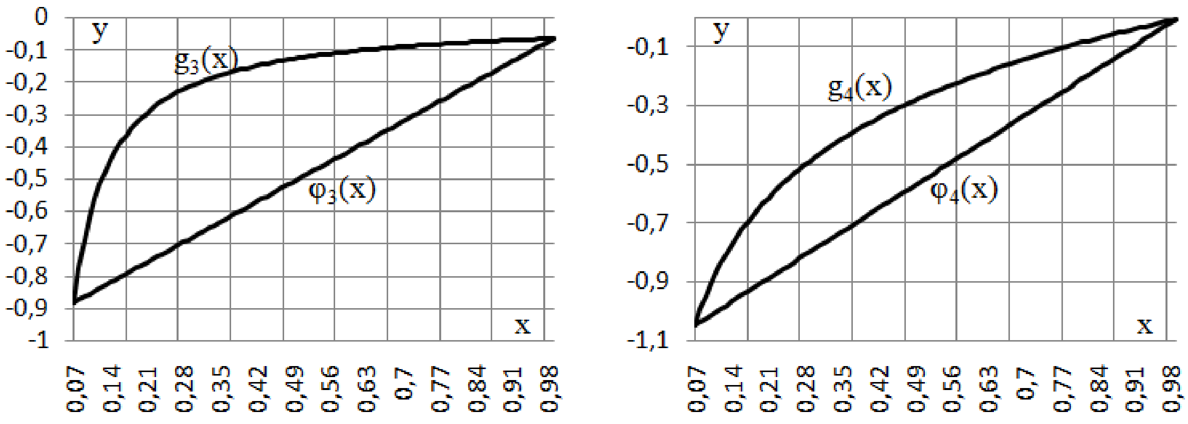
$$y^* = 1{,}4920 + 0{,}3944\ln x, \; R^2 = 0{,}5616. \tag{13}$$

Как видно, лучшей по критерию детерминации является квазилинейная модель (11). Но коэффициент при переменной x^3 не подлежит интерпретации.

В квазилинейных регрессиях (10) – (13) обозначим слагаемые при переменной $x: g_1(x) = 1{,}1053x^2$, $g_2(x) = 1{,}1374x^3$, $g_3(x) = -0{,}0616/x$, $g_4(x) = 0{,}3944\ln(x)$. Графики этих функций на отрезке $x \in [0.07, 0.99]$ представлены на Рисунке 2 (а) – (г). Через концы этих функций проведены прямые, уравнения которых найдены по формуле (4): $\varphi_1(x) = -0{,}0766 + 1{,}1716x$, $\varphi_2(x) = -0{,}0836 + 1{,}199x$, $\varphi_3(x) = -0{,}9426 + 0{,}889x$, $\varphi_4(x) = -1{,}1284 + 1{,}1358x$. Графики этих прямых также приведены на Рисунке 2 (а) – (г) соответственно.



(а) (б)



(в) (г)

Рисунок - 2. Степень нелинейности квазилинейных регрессий (10) – (13)

По Рисунку 2 видно, что самой линейной из квазилинейных регрессий (10) – (13) является модель (10), а самой нелинейной – модель (12).

Для вычисления критериев нелинейности по площади по формулам (5) и (6) были найдены следующие вспомогательные характеристики:

$$S_A^1 = \int\limits_{0,07}^{0,99} \left(-0,0766 + 1,1716x - 1,1053x^2\right) dx = 0,1434,$$

$$S_A^2 = \int\limits_{0,07}^{0,99} \left(-0,0836 + 1,1992x - 1,1374x^3\right) dx = 0,2347,$$

$$S_A^3 = \int\limits_{0,07}^{0,99} \left(\frac{-0,0616}{x} + 0,9426 - 0,8892x\right) dx = 0,2704,$$

$$S_A^4 = \int\limits_{0,07}^{0,99} \left(0,3944\ln(x) + 1,1284 - 1,1358x\right) dx = 0,1909,$$

$$S_\Delta^1 = 0,4958, \qquad S_\Delta^2 = 0,5075, \qquad S_\Delta^3 = 0,3763, S_\Delta^4 = 0,4806.$$

Тогда критерии нелинейности по площади (7) квазилинейных регрессий (10) – (13):

$$NC_S^1 = \frac{0,1434}{0,4958} = 0,2892, \qquad NC_S^2 = \frac{0,2347}{0,5075} = 0,4625,$$

$$NC_S^3 = \frac{0,2704}{0,3763} = 0,7186, \qquad NC_S^4 = \frac{0,1909}{0,4806} = 0,3972.$$

Для вычисления критериев нелинейности по длине дуги были найдены следующие вспомогательные характеристики:

$$l_g^1 = \int_{0,07}^{0,99} \sqrt{1+(2,2106x)^2}\,dx = 1,4657, \qquad l_g^2 = \int_{0,07}^{0,99} \sqrt{1+(3,4122x^2)^2}\,dx = 1,5554,$$

$$l_g^3 = \int_{0,07}^{0,99} \sqrt{1+\left(\frac{0,0616}{x^2}\right)^2}\,dx = 1,4479, \qquad l_g^4 = \int_{0,07}^{0,99} \sqrt{1+\left(\frac{0,3944}{x}\right)^2}\,dx = 1,4702,$$

$$l_\varphi^1 = 1,4171, \qquad l_\varphi^2 = 1,4365, \qquad l_\varphi^3 = 1,2311, \qquad l_\varphi^4 = 1,3922.$$

Критерии нелинейности по длине дуги (8) для квазилинейных регрессий (10) – (13):

$$NC_L^1 = \frac{1,4171}{1,4657} = 0,9668, \qquad NC_L^2 = \frac{1,4365}{1,5554} = 0,9236,$$

$$NC_L^3 = \frac{1,2311}{1,4479} = 0,8503, \qquad NC_L^4 = \frac{1,3922}{1,4702} = 0,9469.$$

В соответствии с найденными значениями критериев нелинейности по площади и по длине дуги, квазилинейные регрессии (10) – (13) можно упорядочить по возрастанию степени их нелинейности следующим образом: (10), (13), (11), (12).

Завершение. В работе на основе коэффициента Джини предложен критерий нелинейности квазилинейных регрессионных моделей по площади и его аналог – критерий нелинейности по длине дуги. Показано, каким образом при низкой степени нелинейности квазилинейных регрессий можно их интерпретировать. Рассмотрен численный пример оценивания нелинейности однофакторной квазилинейной регрессии. Разработанные критерии можно использовать при организации технологии «конкурса» моделей для контроля степени их нелинейности.

ЛИТЕРАТУРА

1. Айвазян С.А. Прикладная статистика: исследование зависимостей / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М. : Финансы и статистика, 1985. – 487 с.
2. Себер Дж. Линейный регрессионный анализ / Дж. Себер. – М. : Мир, 1980. – 456 с.
3. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных / С.И. Носков. – Иркутск : Облинформпечать, 1996. – 321 с.
4. Базилевский М.П. Технология организации конкурса регрессионных моделей / М.П. Базилевский, С.И. Носков // Информационные

технологии и проблемы математического моделирования сложных систем. – Иркутск, 2009. – Вып. 7. – С. 77-84.
5. Базилевский М.П. Методические и инструментальные средства построения некоторых типов регрессионных моделей / М.П. Базилевский,
С.И. Носков // Системы. Методы. Технологии. – 2012. – №1(13). – С. 80-87.
6. Носков С.И. Построение регрессионных моделей с использованием аппарата линейно-булевого программирования / С.И. Носков, М.П. Базилевский. – Иркутск: ИрГУПС, 2018. – 176 с.
7. Смоляк С.А. Оптимальное восстановление функций и связанные с ним геометрические характеристики множеств. – Тр. 3 зимней школы по математическому программированию и смежным вопросам. – М.: ЦЭМИ АН СССР, 1970, вып. 3. – С. 509 – 557.
8. Клейнер Г.Б. Эконометрические зависимости: принципы и методы построения / Г.Б. Клейнер, С.А. Смоляк. – М. : Наука, 2000. – 104 с.
9. Yitzhaki S., Schechtman E. The Gini methodology. A primer on a statistical methodology // Springer series in statistics, 2013. – 548 p.
10. Павлов О.И., Павлова О.Ю. Функция Лоренца и математическое определение среднего класса // Управление экономическими системами: электронный научный журнал. – 2016. – №12 (94).

M. P. Bazilevskiy

NONLINEAR CRITERIA OF QUASILINEAR REGRESSION MODELS

Irkutsk State Transport University,
Irkutsk, Russia

Often, when choosing the best regression model in some given sense, this is can be non-linear regression. For example, with the implementation of the «competition» of models, the best regression may be quasilinear. The advantage of quasilinear regressions is the possibility of estimating them using the ordinary least squares. But the estimates obtained for the parameters of the quasilinear model rarely provide any meaningful interpretation. As a result, the constructed regression, provided that it has a high quality of approximation, can be used only for obtaining forecasts, which significantly reduces its practical significance. This paper is devoted to the problem of estimating the degree of nonlinearity of quasilinear regression models. Based on the Gini coefficient, a nonlinearity criterion for area has been developed. Also presented is its analogue - a nonlinearity criterion in length. These nonlinearity criteria allow us to estimate the degree of nonlinearity of both single-factor and multifactor quasilinear regressions. It is shown how the parameters of quasilinear regression can be interpreted with a low degree of nonlinearity. A specific numerical example of estimating the degree of nonlinearity of single-factor quasilinear regressions is considered.

The developed criteria can be used in organizing the technology of «competition» of models to control the degree of their nonlinearity.

Keywords: quasilinear regression, «competition» of models, interpretation of regression, Lorenz curve, Gini coefficient, nonlinear criterion.

REFERENCES

1. Ajvazyan S.A., Enyukov I.S., Meshalkin L.D. Prikladnaya statistika: issledovanie zavisimostej. Moscow, Finansy i statistika, 1985. 487 p. (in Russian)
2. Seber Dzh. Linejnyj regressionnyj analiz. Moscow, Izdatel'stvo «Mir», 1980. 456 p. (in Russian)
3. Noskov S.I. Tehnologija modelirovanija ob#ektov s nestabil'nym funkcionirovaniem i neopredelennost'ju v dannyh. Irkutsk: RIC GP «Oblinformpechat'», 1996. 321 p. (in Russian)
4. Bazilevskij M.P., Noskov S.I. Tekhnologiya organizacii konkursa regressionnyh modelej. Informacionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnyh sistem. Irkutsk, 2009, no. 7, pp. 77-84. (in Russian)
5. Bazilevskij M.P., Noskov S.I. Metodicheskie i instrumental'nye sredstva postroeniya nekotoryh tipov regressionnyh modelej. Sistemy. Metody. Tekhnologii, 2012, no. 1, vol. 13, pp. 80-87. (in Russian)
6. Noskov S.I., Bazilevskij M.P. Postroenie regressionnyh modelej s ispol'zovaniem apparata linejno-bulevogo programmirovaniya. Irkutsk: IrGUPS, 2018. 176 p. (in Russian)
7. Smolyak S.A. Optimal'noe vosstanovlenie funkcij i svyazannye s nim geometricheskie harakteristiki mnozhestv. Tr. 3 zimnej shkoly po matematicheskomu programmirovaniyu i smezhnym voprosam. Moscow, CEHMI AN SSSR, 1970, vol. 3, pp. 509 – 557. (in Russian)
8. Klejner G.B., Smolyak S.A. Ekonometricheskie zavisimosti: principy i metody postroeniya. Moscow, Nauka, 2000. 104 p. (in Russian)
9. Yitzhaki S., Schechtman E. The Gini methodology. A primer on a statistical methodology. Springer series in statistics, 2013. 548 p.
10. Pavlov O.I., Pavlova O.Yu. Funkciya Lorenca i matematicheskoe opredelenie srednego klassa. Upravlenie ehkonomicheskimi sistemami: ehlektronnyj nauchnyj zhurnal, 2016, no. 12, vol. 94. (in Russian)