УДК 519.862.6
DOI:10.38028/ESI.2022.27.3.015

Исследование критериев нелинейности аддитивных степенных регрессий

Базилевский Михаил Павлович, Караулова Анна Витальевна
Иркутский государственный университет путей сообщения,
Россия, Иркутск, *mik2178@yandex.ru*

Аннотация. Статья посвящена, в первую очередь, исследованию поведения критерия нелинейности «по площади» для аддитивных степенных регрессий с одной объясняющей переменной в зависимости от степени. Этот критерий исследован на монотонность и доказано, что его пределы на бесконечности равны единице. На основе доказанной теоремы введено понятие «воронка нелинейности». Показано, как с помощью «воронок нелинейности» можно контролировать при оценивании степень нелинейности аддитивных степенных регрессий. Предложен новый критерий нелинейности, вычисляемый как разность между единицей и значением коэффициента детерминации в линейной регрессии зависимости объясняющей переменной от той же переменной, возведенной в степень. Установлено, что предложенный критерий корректно использовать только на выборках большого объема.

Ключевые слова: аддитивная степенная регрессия, критерий нелинейности, коэффициент детерминации, «воронка нелинейности»

Цитирование: Базилевский М.П. Исследование критериев нелинейности аддитивных степенных регрессий / М.П. Базилевский, А.В. Караулова // Информационные и математические технологии в науке и управлении. – 2022. – № 3(27). – С. 162-173. – DOI: 10.38028/ESI.2022.27.3.015.

Введение. Проблема оценивания степени нелинейности различных математических функций является весьма актуальной. Её решению посвящены множество научных работ. Так, в [1] предложен подход к определению степени нелинейности дискретных функций, заданных на циклической группе примарного порядка. Работа [2] посвящена оценке нелинейности координатных полиномов произведения преобразований двоичного векторного пространства. В [3] доказывается верхняя оценка нелинейности булевых функций от четного числа переменных. В [4] проведен анализ нелинейных искажений в радиотракте с применением различных методов оценки нелинейности. В [5] описана методика оценивания степени нелинейности отклика грунта при сильных воздействиях. В [6] в качестве критерия нелинейности выступает индекс нелинейности, основанный на преобразовании Гильберта. Работа [7] посвящена измерению нелинейных свойств функций, применяемых в криптографии. В [8] исследованы критерии максимальной нелинейности функции над конечным полем. В [9] рассмотрено оценивание нелинейности в задаче построения доверительных областей движения потенциально опасных астероидов.

При проведении регрессионного анализа [10,11] исследователи часто прибегают к построению нелинейных моделей [12–15]. Степень нелинейности таких регрессий необходимо тщательно контролировать, иначе их будет проблематично интерпретировать.

Рассмотрим аддитивную степенную регрессию с одной объясняющей переменной x:

$$y_i = \alpha_0 + \alpha_1 x_i^k + \varepsilon_i , \qquad i = \overline{1,n} , \qquad (1)$$

где n – объем выборки; y_i, $i=\overline{1,n}$ – значения объясняемой переменной; x_i, $i=\overline{1,n}$ – значения объясняющих переменных; α_0, α_1 – неизвестные параметры; k – выбранное число из промежутка $(-\infty,0)\cup(0,\infty)$; ε_i, $i=\overline{1,n}$ – ошибки аппроксимации. Будем считать, что $x_i > 0$, $i=\overline{1,n}$.

При $k=1$ модель (1) является линейной, а при $k \in (-\infty,0)\cup(0,1)\cup(1,\infty)$ нелинейной по объясняющей переменной, но линейной по параметрам, поэтому она легко оценивается с помощью метода наименьших квадратов (МНК). Число k в (1) нельзя брать равным нулю из-за возникновения совершенной мультиколлинеарности. Возникает вопрос: как меняется нелинейность модели (1) в зависимости от степени k? Ответ на этот вопрос позволит контролировать нелинейность рассмотренной регрессии при её построении.

Для оценки нелинейности будем использовать критерий нелинейности «по площади», предложенный в [16–18]. Для модели (1) он имеет вид

$$NC = \left| \frac{x_{\max}^k + x_{\min}^k}{x_{\max}^k - x_{\min}^k} - \frac{2\int\limits_{x_{\min}}^{x_{\max}} x^k dx}{\left(x_{\max} - x_{\min}\right)\left(x_{\max}^k - x_{\min}^k\right)} \right|, \tag{2}$$

где $x_{\min}$ и $x_{\max}$ – минимальное и максимальное значение переменной x на выборке. По определению критерий (2) принимает значения от 0 до 1. Если $NC = 0$, то модель (1) является линейной. Очевидно, что $NC = 0$ при $k = 1$, а чем ближе значение NC к единице, тем выше степень нелинейности модели.

Целью данной работы является исследование зависимости критерия нелинейности NC от степени k аддитивной степенной регрессии (1).

1. Исследование зависимости критерия нелинейности NC от k.

Теорема. Если $x_{\max} > 0$, $x_{\min} > 0$ и $x_{\max} > x_{\min}$, то критерий нелинейности NC (2) монотонно убывает на промежутке $(-\infty, 0) \cup (0,1)$ и возрастает на промежутке $(1, \infty)$. При этом $\lim\limits_{k\to-\infty} NC = \lim\limits_{k\to\infty} NC = 1$.

Доказательство.

Заметим, что при $k = 0$ функция (2) является неопределенной.

Сначала исследуем поведение критерия (2) при отсутствии знака модуля. Для этого рассмотрим 2 случая.

Случай №1. При $k \neq -1$ функция (2) без модуля будет иметь вид:

$$NC^* = \frac{x_{\max}^k + x_{\min}^k}{x_{\max}^k - x_{\min}^k} - \frac{2\left(x_{\max}^{k+1} - x_{\min}^{k+1}\right)}{(k+1)\left(x_{\max} - x_{\min}\right)\left(x_{\max}^k - x_{\min}^k\right)} =$$

$$= \frac{x_{\min}^k\left(\left(\frac{x_{\max}}{x_{\min}}\right)^k + 1\right)}{x_{\min}^k\left(\left(\frac{x_{\max}}{x_{\min}}\right)^k - 1\right)} - 2\frac{x_{\min}^{k+1}\left(\left(\frac{x_{\max}}{x_{\min}}\right)^{k+1} - 1\right)}{(k+1)\,x_{\min}^{k+1}\left(\left(\frac{x_{\max}}{x_{\min}}\right) - 1\right)\left(\left(\frac{x_{\max}}{x_{\min}}\right)^k - 1\right)}.$$

Пусть $\frac{x_{\max}}{x_{\min}} = b > 1$. Тогда

$$NC^* = \frac{b^k + 1}{b^k - 1} - 2\frac{b^{k+1} - 1}{(k+1)(b-1)(b^k - 1)} = \frac{b^k - 1 + 2 - 2\frac{1}{b-1}\frac{b^{k+1} - 1}{k+1}}{(b^k - 1)} =$$

$$= 1 + \frac{2(b-1)(k+1) - 2(b^{k+1} - 1)}{(b^k - 1)(b-1)(k+1)}. \tag{3}$$

Используя (3), нетрудно показать, что $\lim\limits_{k\to-\infty} NC^* = -1$, а $\lim\limits_{k\to\infty} NC^* = 1$.

Также с помощью (3) можно установить, что

$$\lim_{k\to 0} NC^* = 1 - \frac{2b}{b-1} + \frac{2}{\ln b}, \qquad \lim_{k\to -1} NC^* = \frac{b^{-1}+1}{b^{-1}-1} - \frac{2}{b-1}\cdot\frac{\ln b}{b^{-1}-1},$$

поэтому в точках $k=0$ и $k=-1$ функция NC^* имеет устранимые разрывы.

Теперь докажем, что для любого $b>1$ функция NC^* монотонно возрастает на промежутке $(-\infty,-1)\cup(-1,0)\cup(0,\infty)$.

Первая производная функции (3) имеет вид:

$$\left(NC^*\right)' = \frac{2b^k \ln b\left(b-1\right)}{(b^k-1)^2(k+1)^2}\cdot J, \tag{4}$$

где $J = \dfrac{\left(b^{k+1}-1\right)\left(b^k-1\right)}{b^k \ln b\left(b-1\right)} - k\left(k+1\right)$.

Очевидно, что знак производной (4) будет зависеть от знака функции J. Исследуем функцию J. Её первая, вторая и третья производные имеют вид:

$$J' = \frac{b^{2k+1}-1}{b^k\left(b-1\right)} - \left(2k+1\right), \quad J'' = \frac{\ln b\left(b^{2k+1}+1\right)}{b^k(b-1)} - 2, \quad J''' = \ln^2 b\,\frac{b^{2k+1}-1}{b^k(b-1)}.$$

Заметим, что при $b>1$ $\lim\limits_{k\to-1} J'' = \dfrac{\ln b\left(b^{-1}+1\right)}{b^{-1}(b-1)} - 2 = \dfrac{\ln b\left(1+b\right)}{b-1} - 2 > 0$, $\lim\limits_{k\to-1/2} J'' = \dfrac{2\ln b}{b^{-1/2}(b-1)} - 2 < 0$, $\lim\limits_{k\to 0} J'' = \dfrac{\ln b\left(b+1\right)}{(b-1)} - 2 > 0$. Из этого следует, что $J''=0$ в некоторой точке $k_1 \in \left(-1, -1/2\right)$ и $k_2 \in \left(-1/2, 0\right)$.

Рассмотрим следующие 4 случая.

1. Переменная $k\in(0,\infty)$. В этом случае $J'''>0$, поэтому J'' возрастает. При $k\to 0$ производная $J''>0$, поэтому J' возрастает. При $k\to 0$ производная $J'\to 0$, следовательно, $J'>0$, поэтому J возрастает. При $k\to 0$ функция $J\to 0$, следовательно, функция $J>0$, поэтому и $\left(NC^*\right)'>0$. Это означает, что при $k\in(0,\infty)$ функция NC^* возрастает для любого $b>1$.

2. Переменная $k\in\left(-1/2, 0\right)$. В этом случае $J'''>0$, поэтому J'' возрастает. При $k\to 0$ производная $J''>0$, а при $k\to -1/2$ производная $J''<0$, следовательно, $J''<0$ при $k\in(-1/2, k_2)$ и $J''>0$ при $k\in(k_2, 0)$. Тогда J' убывает при $k\in(-1/2, k_2)$ и возрастает при $k\in(k_2,0)$. При $k\to 0$ производная $J'\to 0$, следовательно, $J'<0$, поэтому J убывает при $k\in(k_2,0)$. При $k\to -1/2$ производная $J'\to 0$, следовательно, $J'<0$, поэтому J убывает при $k\in(-1/2, k_2)$. При $k\to 0$ функция $J\to 0$, следовательно, функция $J>0$, поэтому и $\left(NC^*\right)'>0$. Это означает, что при $k\in\left(-1/2, 0\right)$ функция NC^* возрастает для любого $b>1$.

3. Переменная $k\in\left(-1, -1/2\right)$. В этом случае $J'''<0$, поэтому J'' убывает. При $k\to -1$ производная $J''>0$, а при $k\to -1/2$ производная $J''<0$, следовательно, $J''>0$ при

$k \in (-1, k_1)$ и $J'' < 0$ при $k \in \left(k_1, -\tfrac{1}{2}\right)$. Тогда J' возрастает при $k \in (-1, k_1)$ и убывает при $k \in \left(k_1, -\tfrac{1}{2}\right)$. При $k \to -1$ производная $J' \to 0$, следовательно, $J' > 0$, поэтому J возрастает при $k \in (-1, k_1)$. При $k \to -\tfrac{1}{2}$ производная $J' \to 0$, следовательно, $J' > 0$, поэтому J возрастает при $k \in (k_1, -\tfrac{1}{2})$. При $k \to -1$ функция $J \to 0$, следовательно, функция $J > 0$, поэтому и $\left(NC^*\right)' > 0$. Это означает, что при $k \in \left(-1, -\tfrac{1}{2}\right)$ функция NC^* возрастает для любого $b > 1$.

4. Переменная $k \in (-\infty, -1)$. В этом случае $J''' < 0$, поэтому J'' убывает. При $k \to -1$ производная $J'' > 0$, поэтому J' возрастает. При $k \to -1$ производная $J' \to 0$, следовательно, $J' < 0$, поэтому J убывает. При $k \to -1$ функция $J \to 0$, следовательно, функция $J > 0$, поэтому и $\left(NC^*\right)' > 0$. Это означает, что при $k \in (-\infty, -1)$ функция NC^* возрастает для любого $b > 1$.

Таким образом, при $k \in (-\infty, -1) \cup (-1, 0) \cup (0, \infty)$ функция NC^* возрастает для любого $b > 1$.

Случай №2. При $k = -1$ функция (2) без модуля будет иметь вид:

$$NC^{**} = \frac{x_{\max}^{-1} + x_{\min}^{-1}}{x_{\max}^{-1} - x_{\min}^{-1}} - \frac{2\ln {x_{\max}}/{x_{\min}}}{\left(x_{\max} - x_{\min}\right)\left(x_{\max}^{-1} - x_{\min}^{-1}\right)} =$$

$$= \frac{b^{-1} + 1}{b^{-1} - 1} - \frac{2\ln b}{(b-1)\left(b^{-1} - 1\right)}.$$

Поскольку $NC^{**} = \lim\limits_{k \to -1} NC^*$, то функция (2) в точке $k = -1$ будет непрерывна.

Учитывая, что функция (2) содержит знак модуля, можно сделать вывод, что для любого $b > 1$ она монотонно убывает при $k \in (-\infty, 0) \cup (0, 1)$ и возрастает при $k \in (1, \infty)$. Теорема доказана.

Из доказанной теоремы следует, что для любой положительной переменной x график функции NC от k будет представлять собой «воронку». Продемонстрируем это на конкретном примере.

Пример. Пусть переменная x принимает значения $x_1 = 7{,}7$, $x_2 = 2{,}9$, $x_3 = 8{,}3$, $x_4 = 4{,}3$, $x_5 = 3$, $x_6 = 4{,}6$, $x_7 = 7{,}6$, $x_8 = 0{,}3$, $x_9 = 4{,}9$, $x_{10} = 1{,}7$. Для этой переменной построим график зависимости критерия нелинейности NC от k (рис. 1).

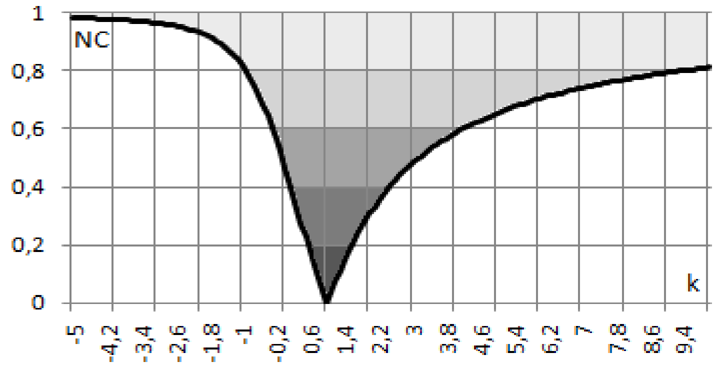


Рис. 1. Зависимость NC от k

Построенный на рис. 1 график будем называть «воронкой нелинейности». По нему видно, что при $k \to \pm\infty$ критерий $NC \to 1$. На промежутке $(-\infty,1)$ функция монотонно убывает, а на промежутке $(1,\infty)$ возрастает. В точке $k=1$ функция достигает своего наименьшего значения $NC=0$. Также можно заметить, что график не является симметричным относительно прямой $k=1$, т.е. изменение отрицательных степеней степенной функции приводит к более резкому изменению нелинейности, чем при изменении положительных степеней. Так, например, на рис. 1 критерий NC достигает значения 0,8 на промежутке $(-\infty,1)$ уже при $k=-0,92$, а на промежутке $(1,\infty)$ только при $k=9,4$.

При неизвестном k аддитивная степенная регрессия (1) является нелинейной, поэтому возникает проблема с её оцениванием. Для приближенного оценивания зачастую назначают некоторую область изменения степени k, разбивают полученный интервал точками, в каждой точке находят оценки параметров α_0 и α_1, и на основе заданного критерия выбирают наилучшие оценки. При этом возникают две проблемы. Первая связана с тем, как на начальном этапе назначать область изменения степени k. Вторая заключается в том, что оцененная модель может оказаться значительно нелинейной, а значит, и плохо интерпретируемой. С помощью предложенных «воронок нелинейности» можно решить обе эти проблемы, контролируя в процессе оценивания степень нелинейности модели. Причём, это работает не только для однофакторных (1), но и для многофакторных аддитивных степенных регрессий.

Рассмотрим возможный вариант контроля степени нелинейности при оценивании аддитивной степенной регрессии. Пусть в распоряжении исследователя имеется m объясняющих переменных. Предположим, что, исходя из своих знаний в данной предметной области, для каждой из этих переменных он может назначить степень нелинейности. Например, одни переменные должны характеризоваться слабой нелинейностью, другие – сильной, а третьи – любой. Для этого можно использовать, например, значения «слабая», «умеренная», «заметная», «высокая», «весьма высокая» из таблицы 1, которая сформирована по аналогии с применяемой в теории корреляции шкалой Чеддока [19].

Таблица 1. Шкала нелинейности

Степень нелинейности	**Значение критерия NC**
Слабая	0 – 0,2
Умеренная	0,2 – 0,4
Заметная	0,4 – 0,6
Высокая	0,6 – 0,8
Весьма высокая	0,8 – 1

Затем для каждой переменной необходимо построить «воронки нелинейности» и на основе таблицы 1 определить области изменения степеней k, после чего внутри полученных промежутков искать наилучшие с точки зрения выбранного критерия аппроксимации оценки.

На рис. 1 разные зоны нелинейности выделены своим оттенком. По этой «воронке» видно, что с ростом критерия NC область изменения параметра k меняется от точки при $NC=0$ до бесконечного промежутка $(-\infty,\infty)$ при $NC \to 1$. В принципе, по этому графику исследователь сам может визуально выбрать подходящую ему зону нелинейности и соответствующий ей промежуток параметра k. Для нашего примера если нелинейность:

1) слабая, то $k \in (0.55, 1.57)$;

2) умеренная, то $k \in (0.15, 0.55) \cup (1.57, 2.46)$;

3) заметная, то $k \in (-0.29, 0) \cup (0, 0.15) \cup (2.46, 4.2)$;

4) высокая, то $k \in (-0.92, -0.29) \cup (4.2, 9.4)$;

5) весьма высокая, то $k \in (-\infty, -0.92) \cup (9.4, \infty)$.

Отметим, что если целью исследователя является интерпретация модели, то степень нелинейности для каждой переменной нужно задавать «слабой».

2. Исследование альтернативного подхода к оценке нелинейности модели. Возникает справедливый вопрос: можно ли использовать для измерения степени нелинейности модели (1) какой-либо другой критерий? Поэтому рассмотрим альтернативный подход к оценке степени её нелинейности. Для этого будем использовать коэффициент детерминации R^2 следующей регрессионной модели:

$$x_i = \beta_0 + \beta_1 x_i^k + \varepsilon_i, \qquad i = \overline{1,n}. \qquad (5)$$

По определению R^2 принимает значения от 0 до 1. Очевидно, что $R^2 = 1$ при $k = 1$, т.е. когда модель (1) является линейной. Как ведет себя этот критерий при больших по абсолютной величине значениях k, неизвестно.

Коэффициент детерминации модели (5) находится по формуле:

$$R^2 = \frac{\left(\overline{x \cdot x^k} - \bar{x} \cdot \overline{x^k}\right)^2}{D_x D_{x^k}},$$

где $\overline{x \cdot x^k} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^k$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{x^k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$, $D_x = \overline{x^2} - \left(\bar{x}\right)^2$, $D_{x^k} = \overline{x^{2k}} - \left(\overline{x^k}\right)^2$.

К сожалению, исследовать функцию R^2 от k на монотонность не так просто, поскольку она будет зависеть от всех наблюдаемых значений переменной x. Например, при одинаковых значениях переменной x функция R^2 будет вовсе неопределенной, поскольку дисперсия $D_x = 0$. Тем не менее, вывод о целесообразности её применения на практике можно сделать на основе следующих рассуждений.

Подставляя в функцию R^2 выражения для дисперсий и средних величин, получим равенство:

$$R^2 = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_i \cdot x_i^k - \frac{1}{n^2}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^k\right)^2}{D_x\left(\frac{1}{n}\sum_{i=1}^{n} x_i^{2k} - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i^k\right)^2\right)}.$$

Найдем предел R^2 при $k \to \infty$:

$$\lim_{k\to\infty} R^2 = \lim_{k\to\infty}\left(x_{\max}^k\left(\frac{1}{n}\frac{\sum_{i=1}^{n} x_i \cdot x_i^k}{x_{\max}^k} - \frac{1}{n^2}\frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^k}{x_{\max}^k}\right)\right)^2\left(D_x x_{\max}^{2k}\left(\frac{1}{n}\frac{\sum_{i=1}^{n} x_i^{2k}}{x_{\max}^{2k}} - \frac{1}{n^2}\left(\frac{\sum_{i=1}^{n} x_i^k}{x_{\max}^k}\right)^2\right)\right)^{-1} =$$

$$= \frac{\left(\frac{1}{n}x_{\max} - \frac{1}{n^2}\sum_{i=1}^{n} x_i\right)^2}{D_x\left(\frac{1}{n} - \frac{1}{n^2}\right)} = \frac{\left(x_{\max} - \bar{x}\right)^2}{D_x(n-1)}.$$

Аналогично можно доказать, что $\lim_{k\to-\infty} R^2 = \frac{\left(x_{\min} - \bar{x}\right)^2}{D_x(n-1)}$.

Тем самым, значения критерия R^2 на плюс и минус бесконечности различны и зависят от объема n выборки. Чем больше n, тем ближе полученные пределы к нулю.

Дополнительно установлено, что $\lim\limits_{k \to 0} R^2 = \dfrac{\left(\overline{x \ln x} - \overline{x} \cdot \overline{\ln x}\right)^2}{D_x\left(\overline{\ln^2 x} - \left(\overline{\ln x}\right)^2\right)}$.

Таким образом, можно предположить, что функция R^2 для любой положительной переменной x монотонно возрастает на промежутке $(-\infty, 0) \cup (0,1)$, преодолевая в точке $k = 0$ устранимый разрыв, и убывает на промежутке $(1, \infty)$. Безусловно, это предположение требует строгих математических доказательств. Однако многочисленные эксперименты подтверждают выдвинутую гипотезу.

На основе функции R^2 введем следующий критерий оценки нелинейности модели (1):

$$\Omega = 1 - R^2. \tag{6}$$

Очевидно, что область значений критерия (6) будет от 0 до 1. При $k = 1$ он достигает своего минимального значения $\Omega = 0$, что соответствует линейной модели. Предположительно, на промежутке $(-\infty, 0) \cup (0,1)$ функция Ω монотонно убывает, а на промежутке $(1, \infty)$ – возрастает. При этом

$$\lim_{k \to -\infty} \Omega = 1 - \frac{\left(x_{\min} - \overline{x}\right)^2}{D_x (n-1)}, \qquad \lim_{k \to \infty} \Omega = 1 - \frac{\left(x_{\max} - \overline{x}\right)^2}{D_x (n-1)}. \tag{7}$$

График зависимости критерия нелинейности Ω от k для нашего примера представлен на рис. 2.

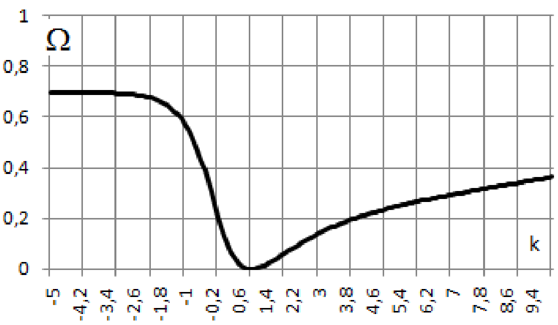


Рис. 2. Зависимость Ω от k

График на рис. 2 согласуется с предположением о поведении функции Ω. Так, на промежутке $(-\infty, 1)$ функция монотонно убывает, а на промежутке $(1, \infty)$ возрастает, принимая своё минимальное значение при $k = 1$. Нельзя не отметить схожесть полученной зависимости с зависимостью NC от k, изображенной на рис. 1, но всё же есть существенная разница. Найденные для нашего примера по формулам (7) пределы $\lim\limits_{k \to -\infty} \Omega = 0,693$, $\lim\limits_{k \to \infty} \Omega = 0,756$. Из этого следует, что, например, при $k = -100000$, т.е. когда аддитивная регрессия значительно нелинейная, критерий Ω не покажет значение больше, чем 0,693. Тем самым, мы имеем усеченную «воронку нелинейности» на рис. 2. В этом плане критерий Ω выглядит несколько ограниченным, по сравнению с NC. Тем не менее, при больших объемах n выборки пределы (7) будут примерно равны 1, поэтому график зависимости Ω от k будет представлять собой полноценную «воронку нелинейности», которую также можно использовать для контроля степени нелинейности аддитивных степенных регрессий.

3. Моделирование. Для демонстрации работы предложенного математического аппарата решалась задача моделирования пассажирских железнодорожных перевозок в Иркутской области. Для этого были использованы статистические данные за 2000–2020 гг. по следующим переменным:

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 полет в салоне экономического класса самолета (в расчете на 1000 км пути) за декабрь (рублей).

Оцененное по этим данным уравнение множественной линейной регрессии имеет вид:

$$\tilde{y} = 42497,5 - \underset{(-8,667)}{134,432}\,x_1 + \underset{(2,602)}{1,152}\,x_2 . \tag{8}$$

В выражении (8) в скобках под коэффициентами указаны значения t-критерия Стьюдента. Как видно, коэффициент при переменной x_1 значим для уровня значимости $\alpha = 0,01$, а коэффициент при переменной x_2 значим для $\alpha = 0,05$.

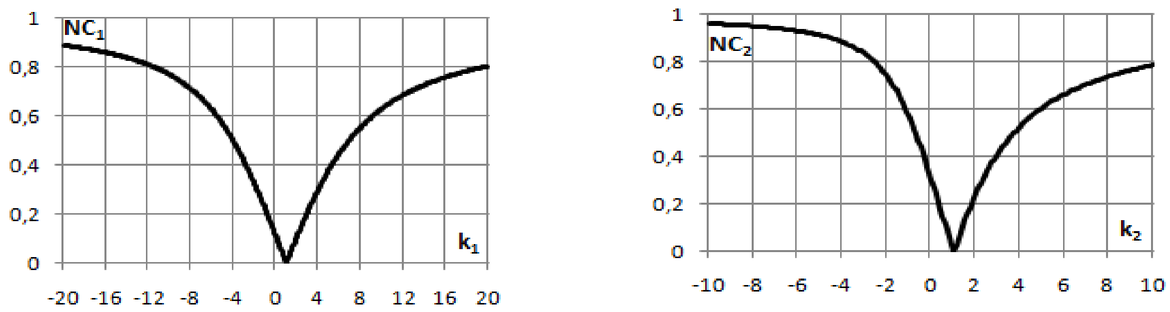
Коэффициент детерминации модели (8) $R^2 = 0,8257$, что позволяет сделать вывод об её хорошем качестве.

Модель (8) интерпретируется следующим образом: с увеличением числа собственных легковых автомобилей на 1000 человек населения x_1 на 1 штуку (при неизменном значении переменной x_2) отправление пассажиров железнодорожным транспортом y уменьшается примерно на 134432 человек; увеличение цены на полет в самолете x_2 на 1000 рублей (при неизменном значении переменной x_1) приводит к увеличению отправления пассажиров железнодорожным транспортом y примерно на 1152 человека.

Для оценивания неизвестных параметров аддитивной степенной регрессии

$$y_i = \alpha_0 + \alpha_1 x_{i1}^{k_1} + \alpha_2 x_{i2}^{k_2} + \varepsilon_i , \qquad i = \overline{1,n} , \tag{9}$$

сначала были построены «воронки нелинейности» для переменных x_1 и x_2 (рис. 3 (а), (б)).



(а) (б)

Рис. 3. «Воронки нелинейности» для переменных x_1 и x_2

Поскольку оцененная нелинейная регрессия (9) должна быть вполне интерпретируемой, то степень нелинейности для каждой переменной была выбрана на уровне между слабой и умеренной, т.е. когда критерий нелинейности NC не превосходит значения 0,3. Для этого по «воронкам нелинейности» (рис. 3 (а), (б)) были найдены области изменения степеней k_1 и k_2:

$$-1,76 \le k_1 \le 4,008 , \qquad -0,01 \le k_2 \le 2,278 . \tag{10}$$

Для численного оценивания параметров аддитивной степенной регрессии (9) на языке программирования hansl эконометрического пакета Gretl была разработана специализированная программа. В её основе лежит следующий алгоритм.

1. Отрезки (10) разбиваются равномерно s точками. В результате получается сеть, состоящая, с учётом точек на концах отрезков, из $(s+2)^2$ узлов.

2. В каждом узле сети с помощью МНК находятся оценки параметров α_0, α_1 и α_2 модели (9).

3. Выбирается лучшая по величине коэффициента детерминации R^2 модель.

Результаты работы программы для различного числа разбиений s представлены в таблице 2.

Таблица 2. Результаты работы программы

Реализация	s	Число моделей	$\tilde{k}_1$	$\tilde{k}_2$	$\tilde{\alpha}_0$	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	R^2
1	10	144	-0,1869	-0,01	394385	435351	-582521	0,848023
2	50	2704	-0,1766	-0,01	383863	436116	-581243	0,848028
3	100	10404	-0,1609	-0,01	365522	440211	-579296	0,848032
4	200	40804	-0,1530	-0,01	354917	443865	-578309	0,848034
5	500	252004	-0,1367	-0,01	329623	455450	-576284	0,848035
6	1000	1004004	-0,1408	-0,01	336553	451929	-576798	0,848035
7	2000	4008004	-0,1400	-0,01	335229	452583	-576697	0,848035
8	5000	25020004	-0,1400	-0,01	335251	452572	-576699	0,848035

Таким образом, наиболее точные оценки аддитивной степенной регрессии получены при $s = 5000$. При этом её уравнение имеет вид:

$$\tilde{y} = 335251 + \underset{(9,238)}{452572}\, x_1^{-0,140016} - \underset{(-3,803)}{576699}\, x_2^{-0,01}. \tag{11}$$

Как видно по таблице 2, коэффициент детерминации R^2 для этой модели равен 0,848035, т.е. её качество лучше, чем у линейной регрессии (8). Причем, в модели (11) значения t-критерия Стьюдента по абсолютной величине стали выше, чем у регрессии (8), т.е. в (11) усилилась значимость переменных – теперь оба коэффициента значимы для уровня значимости $\alpha = 0,01$.

В уравнении (11) затруднительно интерпретировать коэффициенты при переменных x_1 и x_2. Однако значения критериев нелинейности для этих переменных, как и ожидалось, не превосходят величины 0,3 и составляют $NC_1 = 0,1267$ и $NC_2 = 0,2999$, поэтому вместо оценок $\tilde{\alpha}_1$ и $\tilde{\alpha}_2$ модели (11), как отмечено в [18], можно интерпретировать коэффициенты:

$$\theta_1 = \tilde{\alpha}_1 \frac{x_{1,\max}^{k_1} - x_{1,\min}^{k_1}}{x_{1,\max} - x_{1,\min}}, \quad \theta_2 = \tilde{\alpha}_2 \frac{x_{2,\max}^{k_2} - x_{2,\min}^{k_2}}{x_{2,\max} - x_{2,\min}}, \tag{12}$$

где $x_{1,\min}$, $x_{1,\max}$, $x_{2,\min}$, $x_{2,\max}$ – минимальные и максимальные значения переменных x_1 и x_2.

Найденные по формулам (12) коэффициенты $\theta_1 = -153,166$, $\theta_2 = 1,798$. По абсолютной величине они оказались больше, чем коэффициенты линейной регрессии (8). Это ещё раз подтверждает факт усиления значимости переменных в нелинейной модели (11) по сравнению с линейной (8). Интерпретируются коэффициенты θ_1 и θ_2 таким же образом, как это было сделано выше для линейной регрессии.

Заключение. В статье проведено исследование зависимости критерия нелинейности «по площади» от степени аддитивной степенной регрессии с одной объясняющей переменной. Доказано, что график этой зависимости имеет форму воронки. При этом пределы зависимости на бесконечности равны единице. Введен термин «воронка нелинейности». Показано, каким образом с помощью «воронок нелинейности» можно контролировать при оценивании степень нелинейности аддитивных степенных регрессий. Предложен альтернативный критерий нелинейности, основанный на вычислении коэффициента детерминации модели зависимости между объясняющей переменной и той же переменной, возведенной в степень. Выдвинуто предположение о поведении этого критерия. Установлено, что этот критерий, в отличие от первого, корректно работает только на выборках большого объема. Разработанный математический аппарат может успешно применяться для решения конкретных прикладных задач анализа данных. Перспективным выглядит применение предложенных «воронок нелинейности» при оценивании аддитивных степенных регрессий со степенями в виде дискретных функций времени.

Список источников

1. Черемушкин А.В. Аддитивный подход к определению степени нелинейности дискретной функции на циклической группе примарного порядка / А.В. Черемушкин // Прикладная дискретная математика, 2013. – № 2 (20). – С. 26-38.
2. Фомичёв В.М. О степени нелинейности координатных полиномов произведения преобразований двоичного векторного пространства / В.М. Фомичёв // Дискретный анализ и исследование операций, 2021. – Т. 28. – № 2. – С. 74-91. – DOI:10.33048/daio.2021.28.700.
3. Коломеец Н.А. О верхней оценке нелинейности некоторого класса булевых функций с максимальной алгебраической иммунностью / Н.А. Коломеец // Прикладная дискретная математика, 2013. – № 1 (19). – С. 14-16.
4. Селиванов К.А. Анализ нелинейных искажений в радиотракте с применением различных методов оценки нелинейности / К.А. Селиванов, Н.В. Москалец, Т.В. Никитенко // Проблемы телекоммуникаций, 2011. – № 2 (4). – С. 150-161.
5. Петров А.Г. Методика оценивания степени нелинейности отклика грунта при сильных воздействиях / А.Г. Петров // Инженерная сейсморазведка и сейсмология-2020. Георадар-2020. Теперь вместе, 2020. – С. 176-180.
6. Trujillo-Franco L.G., Silva-Navarro G., Beltran-Carbajal F. Algebraic Parameter Identification of Nonlinear Vibrating Systems and Non Linearity Quantification Using the Hilbert Transformation / L.G. Trujillo-Franco, G. Silva-Navarro, F. Beltran-Carbajal // Mathematical Problems in Engineering, 2021, vol. 2021, DOI: 10.1155/2021/5595453.
7. Tan Y., Gong G., Zhu B. Enhanced criteria on differential uniformity and nonlinearity of cryptographically significant functions / Y. Tan, G. Gong, B. Zhu // Cryptography and Communications, 2016, vol. 8, no. 2, pp. 291-311, DOI: 10.1007/s12095-015-0141-x
8. Рябов В.Г. Критерии максимальной нелинейности функции над конечным полем / В.Г. Рябов // Дискретная математика, 2021. – Т. 33. – № 3. – С. 79-91. – DOI: 10.4213/dm1635
9. Черницов А.М. Оценивание нелинейности в задаче построения доверительных областей движения потенциально опасных астероидов, наблюдаемых в одной оппозиции / А.М. Черницов, О.М. Сюсина, В.А. Тамаров // Известия высших учебных заведений. Физика, 2014. – Т. 57. – № 12.
10. Westfall P., Arias A.L. Understanding regression analysis: a conditional distribution approach / P. Westfall, A.L. Arias – Chapman and Hall/CRC, 2020, 514 p.
11. Arkes J. Regression analysis: a practical introduction / J. Arkes – Routledge, 2019, 362 p.
12. Носков С.И. Построение производственной функции с постоянными пропорциями методом антиробастного оценивания / С.И. Носков // Известия Тульского государственного университета. Технические науки, 2022. – № 3. – С. 383-387.
13. Носков С.И. Анализ регрессионной модели грузооборота железнодорожного транспорта / С.И. Носков, И.П. Врублевский // Вестник транспорта Поволжья, 2020. – № 1 (79). – С. 86-90.
14. Глухов Н.И. Математическая модель динамики компьютерных преступлений / Н.И. Глухов, С.И. Носков, П.Ю. Попов // Информационные технологии и математическое моделирование в управлении сложными системами, 2020. – № 1 (6). – С. 1-8.
15. Носков С.И. Регрессионные модели оценки влияния экономических показателей на объемы услуг в сфере информационных и коммуникационных технологий / С.И. Носков, А.С. Вергасов // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика, 2021. – № 4. – С. 95-103.
16. Базилевский М. П. Критерии нелинейности квазилинейных регрессионных моделей / М.П. Базилевский // Моделирование, оптимизация и информационные технологии, 2018. – Т. 6. – № 4. – С. 185-195.

17. Базилевский М. П. Выбор оптимального соотношения между точностью и нелинейностью при построении квазилинейных регрессионных моделей / М.П. Базилевский, А.В. Караулова // Вестник кибернетики, 2021. – № 4 (44). – С. 63-70.
18. Караулова А.В. Программный комплекс построения квазилинейных регрессий по критериям точности и нелинейности / А.В. Караулова, М.П. Базилевский // Экономика. Информатика, 2022. – Т. 49. – № 1. – С. 121-133.
19. Орлов А.И. Вероятностно-статистические модели корреляции и регрессии / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета, 2020. – № 160. – С. 130-162.

Базилевский Михаил Павлович, *доцент, кандидат технических наук, Иркутский государственный университет путей сообщения, AuthorID: 679277, SPIN-код: 4347-5028, ORCID: 0000-0002-3253-5697, mik2178@yandex.ru, Россия, Иркутск, Чернышевского д. 15.*

Караулова Анна Витальевна, *аспирант, Иркутский государственный университет путей сообщения, AuthorID: 1066413, SPIN-код: 3492-4430, anuta_kav@mail.ru, Россия, Иркутск, Чернышевского д. 15.*

UDC 519.862.6
DOI:10.38028/ESI.2022.27.3.015

Study of nonlinearity criteria for additive power regressions

Mikhail P. Bazilevskiy, Anna V. Karaulova
Irkutsk State Transport University, Russia, Irkutsk, *mik2178@yandex.ru*

Abstract. This article is devoted, first of all, to the study of the behavior of the «area» non-linearity criterion for additive power regressions with one explanatory variable depending on the degree. This criteria was examined for monotonicity and it is proved that its limits at infinity are equal to one. Based on the proved theorem, the concept of «funnel of nonlinearity» is introduced. It is shown how with the help of «nonlinearity funnels» it is possible to control the degree of nonlinearity for additive power regressions when estimating. A new criteria of non-linearity is proposed, calculated as the difference between unity and the value of the coefficient of determination in a linear regression of the dependence of the explanatory variable on the same variable raised to a power. It has been established that the proposed criterion is correct to use only on large samples.

Keywords: additive power regression, criteria of non-linearity, coefficient of determination, “funnel of non-linearity”

References

1. Cheremushkin A.V. Additivnyy podkhod k opredeleniyu stepeni nelineynosti diskretnoy funktsii na tsiklicheskoy gruppe primarnogo poryadka [An additive approach to nonlinear degree of discrete function]. Prikladnaya diskretnaya matematika [Applied Discrete Mathematics], 2013, no. 2 (20), pp. 26-38.
2. Fomichev V.M. O stepeni nelineynosti koordinatnykh polinomov proizvedeniya preob-razovaniy dvoichnogo vektornogo prostranstva [On the degree of nonlinearity of the coordinate polynomials for a product of transformations of a binary vector space]. Diskretnyy analiz i issledovaniye operatsiy [Discrete Analysis and Operations Research], 2021, vol. 28, no. 2, pp. 74-91, DOI: 10.33048/daio.2021.28.700
3. Kolomeets N.A. O verkhney otsenke nelineynosti nekotorogo klassa bulevykh funktsiy s maksimal'noy algebraicheskoy immunnost'yu [On an upper estimate for the nonlinearity of a certain class of Boolean functions with maximum algebraic immunity]. Prikladnaya diskretnaya matematika [Applied Discrete Mathematics], 2013, no. 1 (19), pp. 14-16.
4. Selivanov K.A., Moskalets N.V., Nikitenko T.V. Analiz nelineynykh iskazheniy v ra-diotrakte s primeneniem razlichnykh metodov otsenki nelineynosti [Analysis of non-linear distortions in the radio path using various methods for assessing non-linearity]. Problemy telekommunikatsiy [Problems of telecommunications], 2011, no. 2 (4), pp. 150-161.
5. Petrov A.G. Metodika otsenivaniya stepeni nelineynosti otklika grunta pri sil'nykh vozdeystviyakh [A technique for estimating the degree of nonlinearity of the soil response under strong impacts]. Inzhenernaya seysmorazvedka i seysmologiya-2020. Geora-dar-2020. Teper' vmeste [Engineering seismic exploration and seismology-2020. Georadar-2020. Now together], 2020, pp. 176-180.
6. Trujillo-Franco L.G., Silva-Navarro G., Beltran-Carbajal F. Algebraic Parameter Identification of Nonlinear Vibrating Systems and Non Linearity Quantification Using the Hilbert Transformation. Mathematical Problems in Engineering, 2021, vol. 2021, DOI: 10.1155/2021/5595453.

7. Tan Y., Gong G., Zhu B. Enhanced criteria on differential uniformity and nonlinearity of cryptographically significant functions. Cryptography and Communications, 2016, vol. 8, no. 2, pp. 291-311, DOI: 10.1007/s12095-015-0141-x.
8. Ryabov V.G. Kriterii maksimal'noy nelineynosti funktsii nad konechnym polem [Criteria for the Maximum Nonlinearity of a function over a Finite Field]. Diskretnaya matematika [Discrete Mathematics], 2021, vol. 33, no. 3, pp. 79-91, DOI: 10.4213/dm1635.
9. Chernitsov A.M., Syusina O.M., Tamarov V.A. Otsenivanie nelineynosti v zadache po-stroeniya doveritel'nykh oblastey dvizheniya potentsial'no opasnykh asteroidov, nablyu-daemykh v odnoy oppozitsii [Nonlinearity estimation in the problem of constructing confidence regions of motion of potentially hazardous asteroids observed in one opposition]. Izvestiya vysshikh uchebnykh zavedeniy. Fizika. [News of higher educational institutions. Physics], 2014, vol. 57, no. 12.
10. Westfall P., Arias A.L. Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC, 2020, 514 p.
11. Arkes J. Regression analysis: a practical introduction, Routledge, 2019, 362 p.
12. Noskov S.I. Postroenie proizvodstvennoy funktsii s postoyannymi proportsiyami metodom antirobastnogo otsenivaniya [Construction of a production function with constant proportions by antirobast estimation method]. Izvestiya Tul'skogo gosudarstvennogo universiteta [Tekhnicheskiye nauki News of the Tula State University. Technical science], 2022, no. 3, pp. 383-387.
13. Noskov S.I., Vrublevskiy I.P. Analiz regressionnoy modeli gruzooborota zheleznodo-rozhnogo transporta [Analysis of the regression model of railway freight turnover]. Vestnik transporta Povolzhya [Bulletin of transport of the Volga region], 2020, no. 1 (79), pp. 86-90.
14. Glukhov N.I., Noskov S.I., Popov P.Yu. Matematicheskaya model' dinamiki komp'yuter-nykh prestupleniy [Mathematical model of dynamics computer crimes]. Informatsionnyye tekhnologii i matematicheskoye modelirovaniye v upravlenii slozhnymi sistemami [Information technology and mathematical modeling in the management of complex systems], 2020, no. 1 (6), pp. 1-8.
15. Noskov S.I., Vergasov A.S. Regressionnye modeli otsenki vliyaniya ekonomicheskikh pokazateley na ob"emy uslug v sfere informatsionnykh i kommunikatsionnykh tekhnologiy [Regression models of assessing influence of economic indicators on amount of services in information and communication technologies]. Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravleniye, vychislitel'naya tekhnika i informatika [Bulletin of Astrakhan State Technical University. Series: Management, Computer science and Informatics], 2021, no. 4, pp. 95-103.
16. Bazilevskiy M. P. Kriterii nelineynosti kvazilineynykh regressionnykh modeley [Nonlinear criteria of quasilinear regression models]. Modelirovaniye, optimizatsiya i informatsionnyye tekhnologii [Modeling, Optimization and Information Technologies], 2018, vol. 6, no. 4, pp. 185-195.
17. Bazilevskiy M. P., Karaulova A. V. Vybor optimal'nogo sootnosheniya mezhdu tochnost'yu i nelineynost'yu pri postroenii kvazilineynykh regressionnykh modeley [Selecting the optimum relationship between accuracy and non-linearity in constructing quasi-linear regression models]. Vestnik kibernetiki [Proceedings in Cybernetics], 2021, no. 4 (44), pp. 63-70.
18. Karaulova A.V., Bazilevskiy M.P. Programmnyy kompleks postroeniya kvazilineynykh regressiy po kriteriyam tochnosti i nelineynosti [Software complex for constructing quasi-linear regressions according to the criteria of accuracy and non-linearity]. Ekonomika. Informatika [Economics. Information technologies], 2022, vol. 49, no. 1, pp. 121-133.
19. Orlov A.I. Veroyatnostno-statisticheskie modeli korrelyatsii i regressii [Probability-statistical models of correlation and regression]. Politematicheskiy setevoy elektronnyy nauchnyy zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta [Polythematic online scientific journal of Kuban State Agrarian University], 2020, no. 160, pp. 130-162.

Mikhail P. Bazilevskiy, *associate Professor, Candidate of Technical Sciences, Irkutsk State Transport University, AuthorID: 679277, SPIN: 4347-5028, ORCID: 0000-0002-3253-5697, mik2178@yandex.ru, Russia, Irkutsk, 15 Chernyshevskogo St.*

Anna V. Karaulova, *postgraduate Student, Irkutsk State Transport University, AuthorID: 1066413, SPIN: 3492-4430, anuta_kav@mail.ru, Russia, Irkutsk, 15 Chernyshevskogo St.*

Статья поступила в редакцию 20.06.2022; одобрена после рецензирования 06.09.2022; принята к публикации 19.09.2022.

The article was submitted 06/20/2022; approved after reviewing 09/06/2022;accepted for publication 09/19/2022.