УДК 519.862.6
DOI: 10.55648/1998-6920-2022-16-1-89-96

Исследование поведения относительных вкладов переменных в общую детерминацию в оцененном на основе метода выпрямления искаженных коэффициентов регрессионном уравнении

М. П. Базилевский

Для решения проблемы мультиколлинеарности в регрессионном анализе может применяться ранее разработанный автором метод выпрямления искаженных коэффициентов, основанный на построении модели полносвязной линейной регрессии. В статье для оценки степени влияния независимых переменных на зависимую переменную в полученном с помощью этого метода регрессионном уравнении предлагается использовать относительные вклады переменных в общую детерминацию. Доказано, что в таком уравнении в случае линейной функциональной зависимости входных переменных их относительные вклады в общую детерминацию равны. Тогда при сильной корреляции входных переменных их вклады распределяются примерно одинаково. Доказано, что задача оценивания полносвязной регрессии не зависит от выбора связующей переменной. Полученные результаты успешно продемонстрированы на примере моделирования внутреннего валового продукта (ВВП) России.

Ключевые слова: регрессионная модель, мультиколлинеарность, метод выпрямления искаженных коэффициентов, модель полносвязной линейной регрессии, относительные вклады переменных в общую детерминацию, ВВП России.

1. Введение

В регрессионном анализе [1, 2] одной из главных проблем является мультиколлинеарность [3–5]. Она искажает коэффициенты регрессии, что приводит к ошибочной интерпретации влияния тех или иных факторов на зависимую переменную. К настоящему времени проблема мультиколлинеарности не является до конца решенной, поэтому продолжается процесс поиск новых методов [6–8] борьбы с этим негативным явлением. В работе [9] автором была разработана модель полносвязной линейной регрессии (МПЛР), на основе которой в [10] был предложен метод выпрямления искаженных из-за мультиколлинеарности коэффициентов (МВИК). При сильной корреляции независимых переменных знаки коэффициентов в полученном с помощью этого метода регрессионном уравнении согласуются со знаками соответствующих коэффициентов корреляции с зависимой переменной. Поэтому в таком уравнении справедливо оценивать степень влияния независимых переменных на зависимую переменную с помощью известных характеристик – относительных вкладов переменных в коэффициент детерминации. Целью данной работы является исследование поведения этих характеристик и демонстрация полученных результатов на примере моделирования внутреннего валового продукта (ВВП) России.

2. Алгоритм МВИК

МПЛР является обобщением регрессии Деминга и имеет вид:

$$x_{ij} = x_{ij}^{*} + \varepsilon_i^{(x_j)}, \quad i = \overline{1,n}, \quad j = \overline{1,m}, \tag{1}$$

$$x_{ij}^{*} = a_j + b_j x_{im}^{*}, \quad i = \overline{1,n}, \quad j = \overline{1,m-1}, \tag{2}$$

где x_{ij}, $i = \overline{1,n}$, $j = \overline{1,m}$ – наблюдаемые значения m объясняющих (входных, независимых) переменных x_1, x_2, …, x_m; x_{ij}^{*}, $i = \overline{1,n}$, $j = \overline{1,m}$ – их истинные значения; n – объем выборки; a_j, b_j, $j = \overline{1,m-1}$ – неизвестные параметры; $\varepsilon_i^{(x_j)}$, $i = \overline{1,n}$, $j = \overline{1,m}$ – ошибки аппроксимации, которые вызваны неточностями при измерении значений переменных (никаких априорных сведений о вероятностной природе этих ошибок нет).

МПЛР (1), (2) оценивается с помощью взвешенного метода наименьших полных квадратов:

$$\sum_{j=1}^{m-1} \lambda_j \sum_{i=1}^{n} \left(x_{ij} - a_j - b_j x_{im}^{*} \right)^2 + \sum_{i=1}^{n} \left(x_{im} - x_{im}^{*} \right)^2 \to \min, \tag{3}$$

где λ_j, $j = \overline{1,m-1}$ – положительные весовые коэффициенты (лямбда-параметры), представляющие собой с вероятностно-статистической точки зрения отношения дисперсий ошибок переменных, т.е. $\lambda_j = \dfrac{\sigma^2_{\varepsilon^{(x_m)}}}{\sigma^2_{\varepsilon^{(x_j)}}}$, $j = \overline{1,m-1}$.

Переменную x_m^{*} в правой части равенств (2) будем называть связующей. Возникает вопрос: влияет ли выбор связующей переменной в равенствах (2) на решение задачи (3)? Ответ на него дает теорема 1.

Теорема 1. *Выбор связующей переменной в МПЛР (1), (2) не влияет на решение оптимизационной задачи (3).*

Доказательство. Пусть переменная x_1^{*} является связующей. Тогда в МПЛР равенства (2) будут иметь вид:

$$x_{ij}^{*} = a_j^0 + b_j^0 x_{i1}^{*}, \quad i = \overline{1,n}, \quad j = \overline{2,m}, \tag{4}$$

где a_j^0, b_j^0, $j = \overline{2,m}$ – неизвестные параметры.

В таком случае требуется решить оптимизационную задачу

$$\sum_{i=1}^{n} \left(x_{i1} - x_{i1}^{*} \right)^2 + \sum_{j=2}^{m} \lambda_j^0 \sum_{i=1}^{n} \left(x_{ij} - a_j^0 - b_j^0 x_{i1}^{*} \right)^2 \to \min, \tag{5}$$

где $\lambda_j^0 = \dfrac{\sigma^2_{\varepsilon^{(x_1)}}}{\sigma^2_{\varepsilon^{(x_j)}}}$, $j = \overline{2,m}$.

Выразим из (4) при $j = m$ значения связующей переменной $x_{i1}^{*} = -\dfrac{a_m^0}{b_m^0} + \dfrac{1}{b_m^0} x_{im}^{*}$, $i = \overline{1,n}$, и подставим их вместе с соотношениями дисперсий ошибок в (5):

$$\sum_{i=1}^{n} \left(x_{i1} + \frac{a_m^0}{b_m^0} - \frac{1}{b_m^0} x_{im}^{*} \right)^2 + \sum_{j=2}^{m-1} \frac{\sigma^2_{\varepsilon^{(x_1)}}}{\sigma^2_{\varepsilon^{(x_j)}}} \sum_{i=1}^{n} \left(x_{ij} - a_j^0 + b_j^0 \frac{a_m^0}{b_m^0} - \frac{b_j^0}{b_m^0} x_{im}^{*} \right)^2 +$$

$$+\frac{\sigma^2_{\varepsilon^{(x_1)}}}{\sigma^2_{\varepsilon^{(x_m)}}}\sum_{i=1}^{n}\left(x_{im}-x^*_{im}\right)^2\to\min. \qquad (6)$$

Умножив выражение (6) на $\frac{\sigma^2_{\varepsilon^{(x_m)}}}{\sigma^2_{\varepsilon^{(x_1)}}}$ и обозначив $-\frac{a^0_m}{b^0_m}=a_1$, $\frac{1}{b^0_m}=b_1$, $a^0_j-b^0_j\frac{a^0_m}{b^0_m}=a_j$, $\frac{b^0_j}{b^0_m}=b_j$, $j=\overline{2,m-1}$, получим выражение (3). Отсюда следует, что выбор связующей переменной в МПЛР (1), (2) не влияет на решение задачи (3). □

Таким образом, в зависимости от выбора связующей переменной существует m равносильных форм записи МПЛР (1), (2). Дальнейшее изложение ведется исходя из того, что роль связующей переменной играет переменная x^*_m.

Задача (3) при известных лямбда-параметрах решается по следующему алгоритму.

1. Из нелинейной системы

$$b_p\left(D_{x_m}+\sum_{j=1}^{m-1}\lambda_j^2b_j^2D_{x_j}+2\sum_{j_1=1}^{m-2}\sum_{j_2=j_1+1}^{m-1}\lambda_{j_1}\lambda_{j_2}b_{j_1}b_{j_2}K_{x_{j_1}x_{j_2}}+2\sum_{j=1}^{m-1}\lambda_jb_jK_{x_jx_m}\right)=$$
$$=\left(1+\sum_{j=1}^{m-1}\lambda_jb_j^2\right)\left(\sum_{j=1}^{m-1}\lambda_jb_jK_{x_jx_p}+K_{x_mx_p}\right),\ p=\overline{1,m-1},$$

с помощью предложенного в [10] численного метода находятся оценки $\tilde{\imath}_{\ }$, $j=\overline{1,m-1}$.

2. Определяются оценки $\tilde{\imath}$ $\tilde{\ }$, $j=\overline{1,m-1}$.

3. Вычисляются оценки истинных значений переменной x^*_m:

$$\tilde{\ }\sum_{j=1}^{m}A_jx_{ij},\qquad i=\overline{1,n}, \qquad (7)$$

где $A_0=-\left(1+\sum_{j=1}^{m-1}\lambda_j\tilde{\imath}_{\ }\right)^{-1}\tilde{\ }$; $A_j=\lambda_j\tilde{\imath}_{\ }\left(\sum_{j=1}^{\ }\tilde{\ }\right)^{-1}$, $j=\overline{1,m-1}$;

$A_m=\left(1+\sum_{j=1}^{m-1}\lambda_j\tilde{\imath}_{\ }\right)^{-1}$.

В [10] установлено, что при сильной корреляции переменных x_1, x_2, …, x_m знаки коэффициентов уравнения (7) A_j, $j=\overline{1,m-1}$ совпадают со знаками коэффициентов корреляции $r_{x_jx_m}$, $j=\overline{1,m-1}$, а $A_m>0$. На основе этого в [10] предложен метод выпрямления искаженных коэффициентов (МВИК). Его алгоритм «Straight B» заключается в следующем.

1. Оценивается МПЛР (1), (2) при $\lambda_j=D_{x_m}/D_{x_j}$, $j=\overline{1,m-1}$ (в [10] показано, что выбор таких лямбда-параметров гарантирует максимум аддитивного коэффициента детерминации МПЛР).

2. С помощью метода наименьших квадратов (МНК) оценивается модель $y_i=c_0+c_1\tilde{\ }$, $i=\overline{1,n}$, где y_i, $i=\overline{1,n}$ – значения объясняемой (выходной, зависимой) переменной; c_0, c_1 – неизвестные параметры; ε_i, $i=\overline{1,n}$ – ошибки аппроксимации.

3. В оцененную на предыдущем шаге регрессию подставляется выражение (7) и получается регрессионное уравнение:

$$\tilde{\quad} \quad \sum_{j=1}^{m} \theta_j x_j , \tag{8}$$

где $\theta_0 = \tilde{\quad} \quad \tilde{\quad}$; $\theta_j = \tilde{\quad}$, $j = \overline{1,m}$.

3. Относительные вклады переменных уравнения (8) в детерминацию

Коэффициент детерминации R^2 регрессии (8) находится по формуле:

$$R^2 = \sum_{j=1}^{m} \theta_j \frac{\sigma_{x_j}}{\sigma_y} r_{yx_j} . \tag{9}$$

При сильной корреляции переменных x_1, x_2, …, x_m знаки коэффициентов уравнения (8) θ_j, $j = \overline{1,m}$, совпадают со знаками коэффициентов корреляции r_{yx_j}, $j = \overline{1,m}$, т.е. $\theta_j r_{yx_j} > 0$, $j = \overline{1,m}$. Тогда справедливы формулы для относительных вкладов переменных в детерминацию:

$$C_j^{\text{rel}} = \frac{100\theta_j \dfrac{\sigma_{x_j}}{\sigma_y} r_{yx_j}}{R^2}, \qquad j = \overline{1,m} . \tag{10}$$

Теорема 2. *Если все пары объясняющих переменных x_1, x_2, ..., x_m связаны между собой линейными функциональными зависимостями и коэффициент корреляции переменных x_m и y отличен от 0, то относительные вклады C_j^{rel}, $j = \overline{1,m}$, переменных x_1, x_2, ..., x_m в общую детерминацию R^2 для уравнения (8), полученного на основе МВИК при $\lambda_j = D_{x_m} / D_{x_j}$, $j = \overline{1,m-1}$, равны.*

Доказательство. Поделим первые (m–1) равенств (10) на последнее:

$$\frac{C_j^{rel}}{C_m^{rel}} = \frac{\theta_j \sigma_{x_j} r_{yx_j}}{\theta_m \sigma_{x_m} r_{yx_m}}, \qquad j = \overline{1,m-1} . \tag{11}$$

Поскольку $\theta_j = \tilde{\quad}$, $j = \overline{1,m}$, то $\dfrac{\theta_j}{\theta_m} = \dfrac{A_j}{A_m}$, $j = \overline{1,m-1}$. Учитывая, что $A_j = \lambda_j \tilde{\quad}$, $j = \overline{1,m-1}$, а $A_m = \left(1 + \sum_{j=1}^{m-1} \lambda_j \tilde{\quad}\right)^{-1}$, получим $\dfrac{\theta_j}{\theta_m} = \lambda_j \tilde{\quad}$, $j = \overline{1,m-1}$. Тогда (11) примет вид

$$\frac{C_j^{rel}}{C_m^{rel}} = \lambda_j \tilde{\quad} \frac{\sigma_{x_m} r_{yx_m}}{\quad}, \qquad j = \overline{1,m-1} . \tag{12}$$

С учетом того, что $\lambda_j = D_{x_m} / D_{x_j}$, $j = \overline{1,m-1}$, а при линейной функциональной зависимости переменных x_1, x_2, ..., x_m оценка $\tilde{\quad}$ равна МНК-оценке углового коэффициента модели парной линейной регрессии x_j от x_m, т.е. $\tilde{\quad} \quad \dfrac{\sigma_{x_j}}{\sigma_{x_m}}$, $j = \overline{1,m-1}$, формулу (12) можно записать в виде:

$$\frac{C_j^{rel}}{C_m^{rel}} = \frac{D_{x_m}}{D_{x_j}} r_{x_m x_j} \cdot \frac{\sigma_{x_j}}{\sigma_{x_m}} \cdot \frac{\sigma_{x_j} r_{yx_j}}{\sigma_{x_m} r_{yx_m}} = r_{x_m x_j} \cdot \frac{r_{yx_j}}{r_{yx_m}}, \qquad j = \overline{1, m-1}. \tag{13}$$

При линейной функциональной зависимости переменных x_1, x_2, ..., x_m коэффициенты корреляции $r_{x_m x_j}$, $j = \overline{1, m-1}$ могут принимать значения или +1, или –1. Если $r_{x_m x_j} = 1$, то в любом случае $r_{yx_j} = r_{yx_m}$, поэтому отношение (13) принимает значение 1. Если же $r_{x_m x_j} = -1$, то $r_{yx_j} = -r_{yx_m}$, поэтому отношение (13) вновь принимает значение 1. Отсюда следует, что все относительные вклады переменных в детерминацию равны. □

Следствием теоремы 2 является то, что при сильной корреляции переменных x_1, x_2, ..., x_m относительные вклады переменных в детерминацию примерно равны между собой. Таким образом, МВИК приводит к построению регрессионного уравнения (8), в котором общая детерминация примерно равномерно распределяется между всеми входящими в модель переменными.

4. Моделирование ВВП России

Проблема моделирования ключевых показателей экономики России является чрезвычайно актуальной. Одним из таких показателей является валовой внутренний продукт. Для построения регрессионной модели ВВП России были собраны статистические данные (`https://rosstat.gov.ru/`) за период 2000–2020 гг. по следующим переменным:

y – ВВП (в текущих ценах, млрд руб.);

x_1 – среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций в целом по экономике РФ (руб.);

x_2 – численность занятых в возрасте 15–72 лет по РФ (тыс. чел.);

x_3 – численность безработных в возрасте 15–72 лет по РФ (тыс. чел.);

x_4 – наличие основных фондов в РФ на конец отчетного года по полной учетной стоимости (млн руб.);

x_5 – потребление электроэнергии по РФ (млн кВт·ч);

x_6 – продукция сельского хозяйства всех категорий по РФ (в фактических действовавших ценах; млрд руб.);

x_7 – количество введенных зданий жилого и нежилого назначения в РФ (тыс.);

x_8 – грузооборот железнодорожного транспорта (млрд т·км);

x_9 – оборот розничной торговли по РФ (млн руб.);

x_{10} – оборот оптовой торговли по РФ (млрд руб.);

x_{11} – инвестиции в основной капитал в РФ (млн руб.);

x_{12} – средние цены на первичном рынке жилья по РФ (на конец периода, руб. за 1 квадратный метр общей площади);

x_{13} – динамика денежной массы (М2) (млрд руб.);

x_{14} – внешняя торговля (экспорт, млн долл. США);

x_{15} – внешняя торговля (импорт, млн долл. США);

x_{16} – цена на нефть (долл. США);

x_{17} – добыча нефти в России (млн тонн).

Для однородности переменные x_{14}, x_{15}, x_{16}, измеряемые в долларах США, были переведены в рублевый эквивалент.

Найденная для переменных y, x_1, x_2, ..., x_{17} корреляционная матрица показала, что все они очень тесно коррелируют друг с другом. Так, например, самое малое по абсолютной величине значение коэффициента корреляции наблюдается между переменными x_3 и x_7, равное -0,794.

Оцененная по исходной выборке с помощью МНК традиционная модель множественной линейной регрессии имеет вид:

$$\tilde{\quad} \cdot 10^6 + \underset{(2.457)}{67.734}\,x_1 + \underset{(2.676)}{14.73}\,x_2 + \underset{(2.423)}{15.508}\,x_3 - \underset{(-2.459)}{0.00841}\,x_4 + \underset{(2.609)}{1.54}\,x_5 +$$
$$+ \underset{(1.323)}{6.209}\,x_6 + \underset{(2.493)}{1205.95}\,x_7 + \underset{(2.09)}{115.685}\,x_8 - \underset{(-2.432)}{0.0319}\,x_9 + \underset{(2.519)}{10.77}\,x_{10} + \underset{(2.774)}{0.0085}\,x_{11} -$$
$$- \underset{(-2.545)}{12.128}\,x_{12} - \underset{(-2.161)}{10.81}\,x_{13} - \underset{(-2.347)}{0.0197}\,x_{14} - \underset{(-1.588)}{0.0108}\,x_{15} - \underset{(-1.948)}{18.08}\,x_{16} - \underset{(-2.097)}{415.949}\,x_{17}, \qquad (14)$$

для которой $R^2 = 0.999976$. В скобках под коэффициентами указаны соответствующие значения t-критерия Стьюдента.

Как и ожидалось, из-за мультиколлинеарности знаки коэффициентов при переменных x_3, x_4, x_9, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17} в уравнении (11) не соответствуют экономическому смыслу задачи.

Как следует из теоремы 2, в полученном на основе наших исходных данных с помощью МВИК регрессионном уравнении все относительные вклады переменных должны быть примерно равны $\frac{100}{17}\%$ (5.88%).

Оцененная по исходной выборке на основе МВИК регрессионная модель имеет вид:

$$\tilde{\quad} \qquad \tilde{\quad}, \qquad (15)$$

где

$$\tilde{\quad} \quad 51 + 0.000296x_1 + 0.00154x_2 - 0.0033x_3 + 6.236\cdot 10^{-8}x_4 + 4.99\cdot 10^{-5}x_5 +$$
$$+0.0025x_6 + 0.0587x_7 + 0.0112x_8 + 3.925\cdot 10^{-7}x_9 + 0.00017x_{10} + 7.208\cdot 10^{-7}x_{11} +$$
$$+0.000214x_{12} + 0.000277x_{13} + 5.361\cdot 10^{-7}x_{14} + 8.6\cdot 10^{-7}x_{15} + 0.00329x_{16} + 0.0537x_{17}. \qquad (16)$$

Поставляя (16) в (15), получим регрессионное уравнение:

$$\tilde{\quad} \quad 7.795 + \underset{(6.27\%)}{0.146}\,x_1 + \underset{(5.58\%)}{0.761}\,x_2 - \underset{(4.75\%)}{1.649}\,x_3 + \underset{(6.17\%)}{3.088\cdot 10^{-5}}\,x_4 + \underset{(5.98\%)}{0.0247}\,x_5 +$$
$$+ \underset{(6.1\%)}{1.237}\,x_6 + \underset{(5.61\%)}{29.044}\,x_7 + \underset{(5.86\%)}{5.557}\,x_8 + \underset{(6.28\%)}{0.000194}\,x_9 + \underset{(6.2\%)}{0.0846}\,x_{10} + \underset{(6.3\%)}{0.000357}\,x_{11} +$$
$$+ \underset{(5.23\%)}{0.106}\,x_{12} + \underset{(6.19\%)}{0.137}\,x_{13} + \underset{(6.19\%)}{0.000265}\,x_{14} + \underset{(6.32\%)}{0.000426}\,x_{15} + \underset{(5.86\%)}{1.631}\,x_{16} + \underset{(5.11\%)}{26.592}\,x_{17}, \qquad (17)$$

для которого $R^2 = 0.97764$. В уравнении (17) под коэффициентами в скобках указаны относительные вклады переменных в общую детерминацию. Как видно, все объясняющие переменные в (17) поделили общую детерминацию примерно в равных долях, что подтверждает справедливость следствия из теоремы 2.

В модели (17) знаки абсолютно всех коэффициентов при объясняющих переменных соответствуют содержательному смыслу задачи. При этом качество модели (17) лишь незначительно ниже, чем у регрессии (14), поэтому её можно использовать не только для интерпретации, но и для прогнозирования.

Доказанные в работе теоремы могут быть использованы для обобщения МВИК для регрессионных моделей с различной степенью корреляции объясняющих переменных.

Литература

1. *Arkes J.* Regression analysis: a practical introduction. Routledge, 2019. 362 p.
2. *Westfall P. H., Arias A. L.* Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC, 2020. 514 p.
3. *Thompson C.G., Kim R.S., Aloe A.M., Becker B.J.* Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results // Basic and Applied Social Psychology. 2017. V. 39, № 2. P. 81–90.
4. *Yoo C., Cho E.* Effect of multicollinearity on the bivariate frequency analysis of annual maximum rainfall events // Water. 2019. V. 11, № 5. P. 905.
5. *Lindner T., Puck J., Verbeke A.* Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies // Journal of International Business Studies. 2020. V. 51. P. 283–298.
6. *Giacalone M., Panarello D., Mattera R.* Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators // Quality & Quantity: International Journal of Methodology. 2018. V. 52, № 4. P. 1831–1859.
7. *Babaie-Kafaki S., Roozbeh M.* A revised Cholesky decomposition to combat multicollinearity in multiple regression models // Journal of Stat. Comp. and Simul. 2017. V. 87. P. 2298–2308.
8. *Dawoud I.* A new improved estimator for reducing the multicollinearity effects // Communication in Statistics – Simulation and Computation. 2021. P. 1–12.
9. *Bazilevskiy M. P.* Multifactor fully connected linear regression models without constraints to the ratios of variables errors variances // Inform. and its Applic. 2020. V. 14, № 2. P. 92–97.
10. *Bazilevskiy M. P.* Method of straightening distorted due to multicollinearity coefficients in regression models // Informatics and its Applications. 2021. V. 15, № 2. P. 60–65.

Статья поступила в редакцию 05.02.2022;
переработанный вариант –13.03.2022.

Базилевский Михаил Павлович
к.т.н., доцент кафедры математики ИрГУПС (664074, Иркутск, ул. Чернышевского, 15), e-mail: mik2178@yandex.ru.

References

1. Arkes J. Regression analysis: a practical introduction. Routledge, 2019, 362 p.
2. Westfall P. H., Arias A. L. Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC, 2020, 514 p.
3. Thompson C.G., Kim R.S., Aloe A.M., Becker B.J. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*. 2017, vol. 39, no. 2, pp. 81-90.
4. Yoo C., Cho E. Effect of multicollinearity on the bivariate frequency analysis of annual maximum rainfall events. *Water*. 2019, vol. 11, no. 5, pp. 905.
5. Lindner T., Puck J., Verbeke A. Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies. *Journal of International Business Studies*. 2020, vol. 51, pp. 283-298.
6. Giacalone M., Panarello D., Mattera R. Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators. *Quality & Quantity: International Journal of Methodology*. 2018, vol. 52, no. 4, pp. 1831-1859.

7. Babaie-Kafaki S., Roozbeh M. A revised Cholesky decomposition to combat multicollinearity in multiple regression models. *Journal of Stat. Comp. and Simul*. 2017, vol. 87, pp. 2298-2308.
8. Dawoud I. A new improved estimator for reducing the multicollinearity effects. *Communication in Statistics - Simulation and Computation*. 2021, pp. 1-12.
9. Bazilevskiy M. P. Multifactor fully connected linear regression models without constraints to the ratios of variables errors variances. *Inform. and its Applic*. 2020, vol. 14, no. 2, pp. 92-97.
10. Bazilevskiy M. P. Method of straightening distorted due to multicollinearity coefficients in regression models. *Informatics and its Applications*. 2021, vol. 15, no. 2, pp. 60-65.

Researching the behavior of variables relative contributions to the total determination in regression equation estimated using the method of distorted coefficients straightening

Mikhail P. Bazilevskiy
Candidate of technical sciences, Docent, Irkutsk State Transport University (Irkutsk, Russia), mik2178@yandex.ru.

To solve the problem of multicollinearity in regression analysis a distorted coefficients straightening method developed by the author and based on the construction of fully connected linear regression model can be used. In the article, to assess the degree of independent variables influence on the dependent variable in the regression equation obtained by using this method, it is proposed to use the variables relative contributions to the total determination. It is proved that in such an equation in the case of linear functional dependence of the input variables their relative contributions to the total determination are equal. Then, with a strong correlation of the input variables, their contributions are distributed approximately in the same way. It is proved that the problem of estimating a fully connected regression does not depend on the choice of connecting variable. The obtained results have been successfully demonstrated using the example of the Russia's GDP modeling.

Keywords: regression model, multicollinearity, method for straightening distorted coefficients, fully connected linear regression model, relative contributions of variables to the total determination, GDP of Russia.