

## Аспект устранения одной из проблем построения эконометрической модели

И. В. Абрамова<sup>1</sup>, З. В. Шилова<sup>2</sup>

<sup>1</sup> кандидат педагогических наук, доцент кафедры математических и естественнонаучных дисциплин, Пермский государственный национальный исследовательский университет. Россия, г. Пермь.

E-mail: irena-leontio@mail.ru

<sup>2</sup> кандидат педагогических наук, доцент кафедры фундаментальной и компьютерной математики, Вятский государственный университет. Россия, г. Киров. E-mail: zoya@soi.su

**Аннотация.** Центральной проблемой любого эконометрического исследования является построение эконометрической модели, этот аспект объясняется тем, что от качества модели зависит достоверность и обоснованность результатов анализа тенденций развития, прогнозов изучаемых явлений и процессов, а также достоверность выводов исследования. В эконометрических исследованиях предполагается, что закономерности моделируемого процесса складываются под влиянием других явлений, факторов. В данной статье рассматривается один из аспектов эффективного построения эконометрической модели – уравнения множественной регрессии. Авторами рассмотрены базовые понятия эконометрики: проблема мультиколлинеарности, мультиколлинеарные независимые переменные, коэффициенты корреляции, коэффициенты эластичности, стандартизованные коэффициенты регрессии. Выделены признаки, свидетельствующие о наличии мультиколлинеарности, достаточно подробно осуществлен обзор каждого признака, позволяющего выявить мультиколлинеарность множественной регрессии. В данной статье рассматривается один из аспектов эффективного построения эконометрической модели – уравнения множественной регрессии. Авторами рассмотрены базовые понятия эконометрики: проблема мультиколлинеарности, мультиколлинеарные независимые переменные, коэффициенты корреляции, коэффициенты эластичности, стандартизованные коэффициенты регрессии. Выделены признаки, свидетельствующие о наличии мультиколлинеарности, достаточно подробно осуществлен обзор каждого признака, позволяющего выявить мультиколлинеарность множественной регрессии. Отбор материала обусловлен необходимостью теоретического и практического обоснования его изучения студентами экономических направлений подготовки, а также будет полезен специалистам, занимающимся построением эконометрических моделей, пригодных для дальнейшего анализа и прогнозирования.

**Ключевые слова:** эконометрическая модель, мультиколлинеарность, коэффициенты корреляции, коэффициенты эластичности, стандартизованные коэффициенты регрессии.

Одним из основных препятствий эффективного построения эконометрической модели – уравнения множественной регрессии – является мультиколлинеарность независимых переменных.

Заметим, что включение в регрессионную модель мультиколлинеарных независимых переменных приводит к тому, что коэффициенты множественной регрессии теряют экономический смысл и их оценки ненадежны, они меняются с изменением объема наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

Выделим признаки, свидетельствующие о наличии мультиколлинеарности. Во-первых, коэффициенты регрессии имеют неправильные с точки зрения экономической теории знаки или неоправданно большие по модулю значения.

Во-вторых, наблюдается при незначительном изменении первоначальных статистических данных существенное изменение оценок коэффициентов регрессии, вплоть до изменения их знаков.

В-третьих, коэффициенты регрессии оказываются статистически незначимыми по  $t$ -критерию Стьюдента, при этом сама модель по  $F$ -критерию Фишера является статистически значимой.

В-четвертых, множественные коэффициенты детерминации одной из независимых переменных и некоторой группой других имеют значения более 0,6. В-пятых, определитель матрицы парных коэффициентов корреляции между независимыми переменными равен нулю или его значение очень близко к нулю:

$$Det_r = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}, \quad Det_r \rightarrow 0$$

В-шестых, коэффициенты парной корреляции имеют по модулю значения большие 0,75.

В-седьмых, коэффициенты частной корреляции высоких порядков имеют по модулю большие значения.

$$r_{yxix1...xp} = \frac{r_{yxix1...x(p-1)} - r_{yxpx1...x(p-1)} \cdot r_{xixpx1...x(p-1)}}{\sqrt{(1-r_{yxpx1...x(p-1)}^2)(1-r_{xixpx1...x(p-1)}^2)}}.$$

В-восьмых, выполняют следующие шаги:

1) под каждым столбцом матрицы парных коэффициентов корреляций вычисляют сумму по модулю, используя математическую функцию **ABS** табличного процессора MS Excel, каждое значение суммы соответствует определенной независимой (объясняющей) переменной;

2) из найденных сумм модулей элементов матрицы парных корреляций выбирают максимальную, она и выявит номер той независимой переменной, который наиболее коррелирует с остальными переменными (зависимость от остальных объясняющих переменных).

При обнаружении мультиколлинеарности необходимо подтвердить ее достоверность посредством проверки статистической гипотезы:

Нулевая гипотеза  $H_0$ :  $Det_r = 1$ , альтернативная гипотеза  $H_1$ :  $Det_r = 0$ .

$$Det_r = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}.$$

Находим эмпирическое значение критерия  $K_{эмп} = n - 1 - \frac{1}{6} \cdot (2p + 5) \cdot \lg(Det_r)$ , имеющее распределение, приближенное к  $\chi^2$ .

Критическое значение критерия находим с помощью статистической функции **ХИ2.ОБР.ПХ** для уровня значимости  $\alpha$  и степени свободы  $k = \frac{p \cdot (p-1)}{2}$  посредством табличного процессора MS Excel.

Если  $K_{эмп} \geq K_{крит}$ , то гипотезу  $H_0$  отклоняем и мультиколлинеарность считаем доказанной. Если  $K_{эмп} < K_{крит}$ , гипотезу  $H_0$  принимаем и утверждаем, что мультиколлинеарность отсутствует.

В случае достоверного подтверждения мультиколлинеарности выделяем наиболее зависимую объясняющую переменную, для этого используем второй шаг восьмого пункта (см. выше) или сравниваем по модулю стандартизованные коэффициенты, коэффициенты эластичности.

Далее осуществляем устранение мультиколлинеарности: первый путь состоит в исключении из модели одного или нескольких независимых переменных; второй связан с преобразованием независимых переменных, при котором уменьшается корреляция между ними; третий представляет собой переход к уравнениям регрессии, отражающим как влияние независимых переменных, так и их взаимодействие.

Рассмотрим пример. Отметим, что данная задача приведена в классическом практикуме по эконометрике без решения [1, с. 96–97]. Пусть по данным, представленным в табл. 1, изучается зависимость индекса человеческого развития  $y$  от переменных:

$x_1$  – ВВП 1997 г., % к 1990 г.;

$x_2$  – расходы на конечное потребление в текущих ценах, % к ВВП;

$x_3$  – расходы домашних хозяйств, % к ВВП;

$x_4$  – валовое накопление, % к ВВП.

Таблица 1

| y     | x <sub>1</sub> | x <sub>2</sub> | x <sub>3</sub> | x <sub>4</sub> |
|-------|----------------|----------------|----------------|----------------|
| 0,904 | 115,0          | 75,5           | 56,1           | 25,2           |
| 0,922 | 123,0          | 78,5           | 61,8           | 21,8           |
| 0,763 | 74,0           | 78,4           | 59,1           | 25,7           |
| 0,923 | 111,0          | 77,7           | 63,3           | 17,8           |
| 0,918 | 113,0          | 84,4           | 64,1           | 15,9           |
| 0,906 | 110,0          | 75,9           | 57,0           | 22,4           |
| 0,905 | 119,0          | 76,0           | 50,7           | 20,6           |
| 0,545 | 146,0          | 67,5           | 57,1           | 25,2           |
| 0,894 | 113,0          | 78,2           | 62,0           | 20,7           |
| 0,900 | 108,0          | 78,1           | 61,8           | 17,5           |
| 0,932 | 113,0          | 78,6           | 58,6           | 19,7           |
| 0,740 | 71,0           | 84,0           | 71,7           | 18,5           |
| 0,701 | 210,0          | 59,2           | 48,0           | 42,4           |
| 0,744 | 94,0           | 90,2           | 63,9           | 23,0           |
| 0,921 | 118,0          | 72,8           | 59,1           | 20,2           |
| 0,927 | 130,0          | 67,7           | 47,5           | 25,2           |
| 0,802 | 127,0          | 82,6           | 65,3           | 22,4           |
| 0,747 | 61,0           | 74,4           | 53,2           | 22,7           |
| 0,927 | 117,0          | 83,3           | 67,9           | 18,1           |
| 0,721 | 46,0           | 83,7           | 61,7           | 20,1           |
| 0,913 | 107,0          | 73,8           | 52,9           | 17,3           |
| 0,918 | 110,0          | 79,2           | 59,9           | 16,8           |
| 0,833 | 99,2           | 71,5           | 51,5           | 29,9           |
| 0,914 | 101,0          | 75,3           | 61,2           | 20,3           |
| 0,923 | 105,0          | 79,0           | 53,1           | 14,1           |

Построим регрессионную модель:

$$y = 1,112 + 0,001 \cdot x_1 + 0,004 \cdot x_2 - 0,007 \cdot x_3 - 0,014 \cdot x_4.$$

С увеличением ВВП 1997 г. на 1% к 1990 г. при неизменных остальных независимых переменных индекс человеческого развития увеличится на 0,001%. С увеличением расходов на конечное потребление в текущих ценах на 1% к ВВП при неизменных остальных независимых переменных индекс человеческого развития увеличится на 0,004%. С уменьшением расходов домашних хозяйств на 1% к ВВП при неизменных остальных независимых переменных индекс человеческого развития уменьшится на 0,007%. С уменьшением валового накопления на 1% к ВВП при неизменных остальных независимых переменных индекс человеческого развития уменьшится на 0,014%. Уменьшение валового накопления идет большими темпами, чем увеличение ВВП и расходов на конечное потребление и уменьшение расходов домашних хозяйств.

Выполним сравнительную оценку силы связи независимых переменных с зависимой переменной с помощью средних (общих) коэффициентов эластичности и стандартизованных коэффициентов (табл. 2) [2].

$$\bar{\varepsilon}_{x_i} = \frac{b_i \cdot \bar{x}_i}{\bar{y}} \quad \varepsilon_{x_i} = b_i x_i / (a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4) \quad \beta_i = b_i \sigma(x_i) / \sigma(y).$$

Таблица 2

| Независимые переменные | Эластичность   |        | Стандартизованные коэффициенты |        |
|------------------------|----------------|--------|--------------------------------|--------|
| x <sub>1</sub>         | Э <sub>1</sub> | 0,172  | β <sub>1</sub>                 | 0,400  |
| x <sub>2</sub>         | Э <sub>2</sub> | 0,340  | β <sub>2</sub>                 | 0,236  |
| x <sub>3</sub>         | Э <sub>3</sub> | -0,474 | β <sub>3</sub>                 | -0,411 |
| x <sub>4</sub>         | Э <sub>4</sub> | -0,346 | β <sub>4</sub>                 | -0,750 |

Стандартизованные коэффициенты (как и эластичность) позволили ранжировать влияние факторов на результат. Наименее весомый коэффициент – β<sub>2</sub>. Наиболее – β<sub>4</sub>. Если перейти к физической интерпретации, то наиболее весомым фактором является валовое накопление, а наименее весомым – расходы на конечное потребление.

Найдем парные коэффициенты корреляции и построим матрицу парных коэффициентов с помощью надстройки **Пакет анализа** табличного процессора MS Excel (см. табл. 3):

Таблица 3

|       | $x_1$   | $x_2$         | $x_3$         | $x_4$   |
|-------|---------|---------------|---------------|---------|
| $x_1$ | 1       | -0,6290       | -0,3651       | 0,5411  |
| $x_2$ | -0,6290 | 1             | <b>0,7650</b> | -0,6671 |
| $x_3$ | -0,3651 | <b>0,7650</b> | 1             | -0,4963 |
| $x_4$ | 0,5411  | -0,6671       | -0,4963       | 1       |
| сумма | 2,5352  | <b>3,0611</b> | 2,6264        | 2,7045  |

Вычислим значение определителя, соответствующего матрице парных коэффициентов корреляции, подтвердим наличие мультиколлинеарности посредством проверки статистической гипотезы.

Аргументы функции

МОПРЕД

Массив: B28:E31 = {1;-0,629;-0,3651;0,5411;-0,629;1;0,765;-0,6671;-0,3651;0,765;1;-0,4963;0,5411;-0,6671;-0,4963;1}

= 0,125793558

Возвращает определитель матрицы (матрица хранится в массиве).

Массив: числовой массив с равным количеством строк и столбцов, диапазон ячеек или массив.

Значение: 0,125793558

[Справка по этой функции](#)

OK Отмена

$Det_r = 0,125794$ . Эмпирическое значение критерия:

$$K_{эмп} = 25 - 1 - \frac{1}{6} \cdot (2 \cdot 4 + 5) \cdot \lg(0,125794) = 25,95074.$$

Критическое значение критерия находим с помощью статистической функции ХИ2.ОБР.ПХ для уровня значимости  $\alpha = 0,05$  и степени свободы  $k = 6$ , его значение равно 12,59158724, и, так как  $K_{эмп} \geq K_{крит}$ , то мультиколлинеарность считаем доказанной.

Из данной матрицы (см. табл. 3) наблюдаем, что две из четырех независимых переменных мультиколлинеарны ( $r_{x_2 x_3} > 0,75$ ), – это  $x_2$  и  $x_3$ . Проведем дополнительные исследования. Из

найденных сумм модулей элементов матрицы парных корреляций выберем максимальную – это 3,0611, она соответствует независимой переменной  $x_2$ . Также сравним стандартизованные коэффициенты, взятые по модулю, при переменных  $x_2$  и  $x_3$ , получаем, что можно отбросить  $x_2$  без ущерба информативности модели, тогда новое уравнение регрессии будет выглядеть следующим образом:

$$y = 1,325 + 0,001 \cdot x_1 - 0,005 \cdot x_3 - 0,015 \cdot x_4.$$

Построим матрицу парных коэффициентов для новой регрессионной модели посредством надстройки **Пакет анализа** табличного процессора MS Excel (табл. 4).

Таблица 4

|       | $y$      | $x_1$    | $x_3$    | $x_4$ |
|-------|----------|----------|----------|-------|
| $y$   | 1        |          |          |       |
| $x_1$ | -0,00434 | 1        |          |       |
| $x_3$ | -0,00433 | -0,36511 | 1        |       |
| $x_4$ | -0,48711 | 0,54107  | -0,49626 | 1     |

В рассматриваемой части нового получившегося поля корреляции нет коэффициентов, больших 0,75. Это значит, что мы избавились от мультиколлинеарности путем упрощения модели (отбросили одну из независимых переменных). В нашем случае это расходы на конечное потребление в текущих ценах.

Найдем стандартизованные коэффициенты для вновь построенной модели (табл. 5).

Таблица 5

| $\beta_1$ | $\beta_3$ | $\beta_4$ |
|-----------|-----------|-----------|
| 0,328     | -0,284    | -0,806    |

Вычислим коэффициент детерминации для вновь построенной регрессионной модели:

$$R^2 = \beta_1 \cdot r_{yx_1} + \beta_3 \cdot r_{yx_3} + \beta_4 \cdot r_{yx_4},$$

$$R^2 = 0,328 \cdot (-0,00434) + (-0,284) \cdot (-0,00433) + (-0,806) \cdot (-0,48711) = 0,39, \text{ что можно}$$

интерпретировать как зависимость индекса человеческого развития на 40% от трех факторов: ВВП 1997 г., расходы домашних хозяйств, валовое накопление. На другие факторы, не учтенные в модели, приходится 60%. Отсюда следует, что модель не совсем адекватно описывает данные выборки. Здесь мы рекомендуем увеличить объем выборки, а также дополнительно исследовать регрессионную модель на гетероскедастичность и автокорреляцию.

Таким образом, если основная задача модели – прогноз будущих значений зависимой переменной, то при достаточно большом коэффициенте детерминации  $R^2$  (больше 0,9) наличие мультиколлинеарности обычно не сказывается на прогнозных качествах модели, при условии, что и в будущем между коррелированными переменными сохраняются те же соотношения.

В свою очередь, если же целью исследования является определение степени влияния каждой из независимых переменных на зависимую переменную, то наличие мультиколлинеарности, скорее всего, исказит истинные зависимости между переменными, в этом случае наличие мультиколлинеарности является серьезной проблемой и требует устранения из регрессионной модели.

### Список литературы

1. Елисеева И. И., Курышева С. В., Гордеенко Н. М. и др. Практикум по эконометрике : учеб. пособие. М. : Финансы и статистика, 2003. 192 с.
2. Шилова З. В. Эконометрика : учеб. пособие. Саратов : Ай Пи ар Букс, 2015. 148 с. URL: <http://www.iprbookshop.ru/33864>.

## Aspect of solving one of the problems of constructing econometric model

I. V. Abramova<sup>1</sup>, Z. V. Shilova<sup>2</sup>

<sup>1</sup> PhD of pedagogical sciences, associate professor of mathematical and natural sciences, Perm State National Research University. Russia, Perm. E-mail: irena-leontio@mail.ru

<sup>2</sup> PhD of pedagogical sciences, associate professor of fundamental and computer mathematics, Vyatka State University. Russia, Kirov. E-mail: zoya@soi.su

**Abstract:** The Central problem of any econometric study is the construction of econometric model, this aspect is due to the fact that the quality of the model depends on the reliability and validity of the results of the analysis of development trends, forecasts of phenomena and processes under study, as well as the reliability of the findings of the study. In econometric studies it is assumed that the regularities of the simulated process are influenced by other phenomena, factors. This article considers one of the aspects of efficient construction of econometric model – multiple regression equation. The authors consider the basic concepts of econometrics: the problem of multicollinearity, multicollinear independent variables, correlation coefficients, elasticity coefficients, standardized regression coefficients. The signs testifying to the presence of multicollinearity are allocated, the review of each sign allowing to reveal multicollinearity of multiple regression is carried out in sufficient detail. The selection of the material is due to the need for theoretical and practical justification of its study by students of economic areas of training, and will be useful to professionals engaged in the construction of econometric models suitable for further analysis and forecasting.

**Keywords:** econometric model, multicollinearity, correlation coefficients, elasticity coefficients, standardized regression coefficients.

### References

1. Eliseeva I. I., Kuryshcheva S. V., Gordeenko N. M. and others. *Praktikum po ehkonometrike : ucheb. posobie* [Workshop on econometrics: tutorial]. M. Finance and statistics. 2003. 192 p.
2. SHilova Z. V. *EHkonometrika : ucheb. posobie* [Econometrics: tutorial]. Saratov. AI PI ar Books. 2015. 148 p. Available at: <http://www.iprbookshop.ru/33864>