JEL: C01; C13; C52

ЕКОНОМИЧЕСКИ КОРРЕКТНЫЕ И НЕКОРРЕКТНЫЕ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

Тыжненко А. Г.

Ключевые слова: МНК, мультиколлинеарность, некорректные решения, свойства МНК-матицы

С математической точки зрения рассмотрен вопрос появления некорректных решений в задаче многофакторной линейной регрессии. Показано, что это явление связано не с мультиколлинеарностью [1-5], а с фундаментальными свойствами любой несингулярной матрицы.

Область значений экономической корректности МНК матрицы

Определение: произвольная n-мерная квадратная матрица A в поле действительных чисел имеет все пространство R^n в качестве области значений и состоит из двух частей, D^c и $\overline{D}^c$ ($D^c \cup \overline{D}^c = R^n$), где D^c есть область значений для физически корректных решений матричного уравнения $Ax = b$ ($b \in D^c$), а $\overline{D}^c$ есть область значений для физически некорректных решений матричного уравнения $Ax = b$ ($b \in \overline{D}^c$).

Т.е. мы утверждаем этим определением, что любое решение матричного уравнения $Ax = b$ с квадратной несингулярной матрицей A -го порядка, в зависимости от правой части b, принадлежит к одному из двух типов решений. Каждый тип решений состоит из n-мерных векторов (набора коэффициентов регрессий, например) с одинаковыми знаками одноименных координат для всех возможных решений одного типа, но, возможно, с разными знаками координат самих векторов. Решения первого и второго типов различаются тем, что знаки решений (координат) в этих типах обязательно разные, хотя внутри каждого типа они одинаковые. Т.е., решения любого определенного матричного уравнения, как идеально обусловленного, так и плохо-обусловленного, может иметь два принципиально разных решения в зависимости от правой части. Один тип решения соответствует смыслу решаемой задачи, и он назван экономически корректным, а другой тип решений не соответствует смыслу решаемой задачи, и он назван экономически некорректным. Именно с этим свойством МНК-матрицы связано появление некорректных знаков МНК-решений.

На простом примере показано, что с ростом степени коллинеарности данных сужается область корректных решений задачи линейной регрессии и увеличивается вероятность выхода правой части МНК-уравнения из области корректности значений МНК-матрицы за счет случайных ошибок, что и вызывает появление решений другого типа, знаки координат которого отличаются от правильных знаков.

Таким образом, если $b \in D^c$, знаки точного решения уравнения $Ax = b$ согласуются со смыслом задачи, которая исследуется. В этом случае решение этого уравнения является стабильным при любом значении числа обусловленности матрицы A, до тех пор, ясно, пока за счет ошибок в элементах матрицы правая часть не выйдет за пределы D^c.

В общем, некоторые знаки точного решения уравнения $Ax = b$ меняются при переходе вектора правой части из D^c в $\overline{D}^c$ и наоборот. Абсолютные значения компонентов решения зависят от ориентации вектора правой части как при $b \in D^c$, так и при $b \in \overline{D}^c$. Но в последнем случае, они могут очень сильно возрастать в зависимости от ориентации вектора правой части и тем сильнее, чем больше число обусловленности матрицы A.

Следует отметить, что с ростом уровня коллинеарности данных нестабильность решения матричного уравнения $Ax = b$ любым методом возрастает как при $b \in D^c$, так и при $b \in \overline{D}^c$, но при $b \in D^c$ нестабильность решения существенно меньше и знаки решения экономически корректны.

Таким образом, если метод решения уравнения $Ax = b$ обладает повышенной нестабильностью, как, например, метод Крамера при наличии мультиколлинеарности, то нестабильным может как решение при $b \in \overline{D}^c$, так и решение при $b \in D^c$.

Появление больших решений и решений с некорректными знаками геометрически следует из свойств проекций векторов. Если правая часть b выходит за пределы D^c, то ее проекции на базисные векторы меняют некоторые знаки и возрастают по модулю по мере удаления b от $\boldsymbol{D^c}$.

Именно поэтому в решениях задачи линейной регрессии при наличии мультиколлинеарности с помощью МНК появляются как большие по модулю решения, так и решения с некорректными знаками.

Эти проблемы порождены нестабильностью метода Крамера с одной стороны, и сужением области физической корректности МНК-матрицы с ростом уровня коллинеарности данных, с другой стороны.

Сужение области физической корректности МНК-матрицы повышает вероятность выхода вектора правой части из этой области при случайных изменениях векторов данных. Поэтому точное решение задачи линейной регрессии с помощью МНК при наличии мультиколлинеарности будет, скорее всего, экономически неадекватным.

Пример появления экономически некорректных решений

Продемонстрируем новое фундаментальное свойство несингулярных матриц на примере решения простой экономической задачи

Предположим, на торговой точке продают хот-доги и минеральную воду. Каждый хот-дог стоит \$1.50 и каждая бутылка минеральной воды стоит \$0.50. За день выручка составляет \$78.50. Продано вместе 87 единиц товара. В отчете следует указать отдельно, сколько продано хот-догов (x_1) и сколько воды (x_2). Составляем систему уравнений ($Ax = b$):

$$\begin{cases} 1.5x_1 + 0.5x_2 = 78.5 \\ x_1 + x_2 = 87 \end{cases}. \quad (1)$$

Определитель системы равен единице, число обусловленности 4.27. Т.е. система (1) идеально обусловлена. Решение задачи в Matlab: $x = A\backslash b = (35; 52)$. Базисные вектора имеют координаты: $a_1 = (1.5; 1)$, $a_2 = (0.5; 1)$.

Запишем вектор правой части следующим образом: $b = 87(0.9023; 1)$, что позволяет утверждать, что вектор правой части лежит между векторами a_1 и a_2 и принадлежит области физической корректности D^c, поскольку его проекции на a_1 и a_2 положительны, как и должно быть из экономических соображений.

Исследуем решение системы (1) с изменением вектора правой части при фиксированном объеме продаж $b(1) =$ \$78.5. Граничные значения вектора b определяем из условий параллельности $b||a_1$ и $b||a_2$: $b_+ = (78.5; 157)$, $b_- = (78.5; 52.\overline{3})$. В соответствии с экономическим смыслом задачи, берем $b_-(2) = 53$. Тогда, общее число единиц проданной продукции может меняться в пределах [53; 157] и при этом $b \in D^c$.

Если $b = (78.5; 53)$, решением есть $x = (52; 1)$. Это означает, что было продано 52 хот-дога и одна вода. Если $b = (78.5; 157)$, решением есть $x = (0; 157)$. Т.е. продано только 157 бутылок воды. Внутри D^c есть и другие целые решения, например, для $b = (78.5; 109)$ решением есть $x = (24; 109)$.

В другой практической ситуации может продаваться только часть продукции и решения не обязательно будут целыми числами.

Предположим теперь, что вектор b не принадлежит D^c. Этот случай реализуется, если $b(2) < 52.\overline{3}$ или $b(2) > 157$. Например, пусть будет продана 51 единица продукции, т.е., $b(2) = 51$. Тогда решением есть $x = (53;\ -2)$. Мы видим, что знак одного из решений поменялся. При этом и все решение стало некорректным с экономической точки зрения. Продано 53 хот-дога при общем числе проданных единиц продукции 51. Предположим теперь, что $b(2) = 159$ (продано 159 ед. продукции). В этом случае решение $x = (-1;\ 160)$ также некорректно.

Это означает, что мы не можем задавать правую часть уравнения (1) совершенно произвольно, если мы исследуем некую практическую проблему и хотим получить экономически (физически) корректное решение.

Поскольку решение системы линейных уравнений обязательно меняет знаки некоторых компонент решения при переходе правой части с одной части области значений матрицы в другую часть, область корректности легко определяется в том случае, когда исследователь точно знает, какие знаки должно иметь решение. Это имеет место, например, в задаче линейной регрессии.

Появление экономически некорректных решений рассмотрены в работах [6,7]. В работе [6] приведен алгоритм модифицированного метода Крамера, позволяющий получать стабильные экономически корректные решения задачи линейной регрессии при любой степени коллинеарности данных, включая полную идентичность некоторых регрессоров (случай полной мультиколлинеарности). При практическом использовании алгоритма необходимо записать МНК-уравнение с помощью корреляционной матрицы. Это имеет принципиальное значение для погашения ошибок накопления в модифицированном методе Крамера.

Литература

1. Farrar, D. Multicollinearity in regression Analysis: The problem revisited / D. Farrar, R. R. Glauber // Review of Economics and Statistics. – 1967. – 49. – P. 92-107.

2. Hoerl, A. E. Ridge regression: Biased estimation for nonorthogonal problems / A. E. Hoerl, R. W. Kennard // Technometrics. – 1970. – 12(1) . – P. 55–67. 11. Marquardt D.V. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation / D.V. Marquardt // Technometrics. – 1970. – 12. – P. 591-612.

3. Blanchard O.J. Comment. / O.J. Blanchard // Journal of Business and Economic Statistics 5. – 1987. – P. 449–51.

4. Adkins L. C., Collinearity. Companion in Theoretical Econometrics, edited by Badi Baltagi / L. C. Adkins, R. C. Hill. – Oxford: Blackwell Publishers, Ltd. – 2001. – pp. 256-278.

5. Belsley D. A. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity / D. A. Belsley, E. Kun, R. T. Welsh. – New York: Wiley. – 2004. – P. 651.

6. Tyzhnenko A. G. A new stable solution to the linear regression problem under multicollinearity // Economics of Development. – 2018. – Vol. 2 (86). – P. 89–99. URL: http://www.ed.ksue.edu.ua/ER/knt/ee182_86/e182tyz.pdf

7. Tyzhnenko A. G., Ordinary least squares: the Adequacy of Linear Regression Solutions under Multicollinearity and without it / A. G Tyzhnenko, Y. V. Ryeznik // The Problems of Economy. – 2019.- №1. – С. 217–227.

Автор:
Тыжненко Александр Григорьевич, к. ф.-м. н., доцент кафедры высшей математики и экономико-математических методов,
ХНЕУ ім. С. Кузнеця.
oleksandr.tyzhnenko@m.hneu.edu.ua

Тези доповіді надійшли 15 лютого 2020 року.
Опубліковано в авторській редакції.