

УДК [330.55:303.094.5]:311

*А.А. Мухин***НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ОБНАРУЖЕНИЯ ВЗАИМОСВЯЗЕЙ
МАКРОЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ В РОССИЙСКОЙ ФЕДЕРАЦИИ**

В методологии исследования обнаружения взаимосвязей макроэкономических показателей в Российской Федерации используются три параметра: валовый региональный продукт, численность работников, основные фонды. В первой части статьи рассматривается изучение взаимосвязей количественных признаков методами непараметрической статистики, использующей коэффициент конкордации (множественный коэффициент ранговой корреляции). С помощью коэффициента конкордации определена теснота связи между валовым региональным продуктом (трлн руб.), численностью занятых по субъектам Российской Федерации (млн чел.), наличием основных фондов в Российской Федерации (трлн руб.). Значимость коэффициента конкордации оценена по критерию χ^2 . Во второй части статьи построена математическая модель оценки влияния факторов на параметр – валовый региональный продукт РФ. Проведен выбор факторных признаков для построения двухфакторной регрессионной модели за 2007–2013 гг. Рассчитаны параметры модели, проведена оценка параметров регрессии с помощью метода наименьших квадратов. Для оценки качества модели определены линейный коэффициент множественной корреляции и коэффициент детерминации. Осуществлена проверка значимости уравнения и коэффициентов регрессии, уравнения множественной регрессии. Произведена оценка влияния факторов на зависимую переменную по модели с использованием коэффициента эластичности и бета-коэффициентов. Выявлено, что в наибольшей степени на рост валового регионального продукта оказывает влияние такой фактор, как основные фонды.

Ключевые слова: непараметрические методы обнаружения взаимосвязей макроэкономических показателей, математическое моделирование в оценке факторных зависимостей макроэкономических показателей РФ.

Для большинства статистических исследований макроэкономических показателей важно выявить существующие взаимосвязи между протекающими явлениями и процессами. Все наблюдаемые явления экономической жизни общества, какими бы независимыми они ни казались на первый взгляд, – это, как правило, следствие действия определенных факторов. Все статистические методы прогнозирования базируются на факте существования таких зависимостей, которые используются для построения математических моделей, что активно применяется в практике экономических расчетов [1–4].

Часть 1. Методы корреляционного анализа применяются только по отношению к количественным признакам. Большое значение имеют законы распределения исследуемых величин, так как в результате проведения корреляционного анализа обращаются к их основным характеристикам – параметрам распределения: среднему, дисперсии и среднему квадратическому отклонению. Методы статистического анализа, в основе которых лежит расчет параметров распределения величин, называются параметрическими.

Существуют и другие подходы к оценке взаимосвязи явлений – непараметрические методы, их реализация не требует знания законов и параметров распределения показателей. Непараметрические методы статистики применяются в случае, когда законы распределения неизвестны или же различны у показателей, некоторые из них, по отношению к качественным признакам, никогда не имеют количественного выражения.

При изучении взаимосвязей количественных признаков методами непараметрической статистики используются следующие показатели:

- коэффициент корреляции знаков Фехнера;
- коэффициент корреляции рангов Спирмена;
- ранговый коэффициент корреляции Кендэла (ранговые коэффициенты Спирмена и Кендэла применяются также и по отношению к качественным признакам, если их можно каким-то образом ранжировать);
- коэффициент конкордации (множественный коэффициент ранговой корреляции).

В статье рассмотрим применение коэффициента конкордации.

Коэффициент конкордации, или множественный коэффициент ранговой корреляции, применяется для измерения тесноты связи между несколькими признаками (больше двух), если их можно определенным образом ранжировать, поэтому данный коэффициент носит название «множественный».

Формула для вычисления коэффициента конкордации в случае связанных рангов имеет вид:

$$W = \frac{12S}{m^2(n^2 - n)},$$

где m – количество признаков, между которыми устанавливается связь;

n – число наблюдений;

S – разность между суммой квадратов рангов и средним квадратом рангов, то есть

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m R_{ji} \right)^2 - \frac{\left(\sum_{i=1}^n \sum_{j=1}^m R_{ji} \right)^2}{n},$$

где n – число наблюдений ($i = 1, 2, \dots, n$);

m – число признаков ($j = 1, 2, \dots, m$);

R_{ji} – ранг, присвоенный i -му значению j -го признака.

Коэффициент конкордации изменяется от 0 до 1. Чем ближе его значение к 1, тем сильнее связь между признаками.

По данным табл. 1 определим тесноту связи между валовым региональным продуктом (трлн руб.), численностью занятых по субъектам Российской Федерации (млн чел.), наличием основных фондов в Российской Федерации (трлн руб.).

Таблица 1

**Расчет коэффициента конкордации между основными экономическими показателями
Российской Федерации (2007–2013 гг.) [5]**

Год	Валовый региональный продукт, трлн руб. x	Численность работников, млн чел. y	Основные фонды, трлн руб. z	Ранги x R_x	Ранги y R_y	Ранги z R_z	Сумма рангов показателей ($R_x + R_y + R_z$)	Квадрат суммы рангов ($R_x + R_y + R_z$) ²
2007	28	70,8	60,4	1	3	1	5	25
2008	33,9	71	74,4	3	5	2	10	100
2009	32	69,4	82,3	2	1	3	6	36
2010	37,7	69,9	93,2	4	2	4	10	100
2011	45,4	70,9	108	5	4	5	14	196
2012	49,9	71,5	121,2	6	7	6	19	361
2013	54	71,4	133,5	7	6	7	20	400
Σ	–	–	–	–	–	–	84	1218

Определим величину S :

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m R_{ji} \right)^2 - \frac{\left(\sum_{i=1}^n \sum_{j=1}^m R_{ji} \right)^2}{n} = 1218 - \frac{84^2}{7} = 21.$$

Тогда коэффициент конкордации получится следующим:

$$W = \frac{12S}{m^2(n^2 - n)} = \frac{12 \cdot 210}{3^2 \cdot (7^2 - 7)} = 0,833.$$

Значение коэффициента указывает на умеренную связь между тремя исследуемыми показателями. Подобное заключение мы сделали на основе данных семи наблюдений, поэтому естественно предположить, что полученные результаты исследования случайны. Для того чтобы исключить принятие ошибочного решения, проверим коэффициент на значимость.

Значимость коэффициента конкордации оценивается по критерию χ^2 . Наблюдаемое значение $\chi^2_{\text{набл}}$ при отсутствии связанных рангов определяется по формуле

$$\chi^2_{\text{набл}} = \frac{12S}{mn(n+1)}.$$

Если наблюдаемое значение превышает критическое значение, соответствующее принятому уровню значимости α и числу степеней свободы $\nu = n - 1$, то коэффициент W считается значимым.

Критическое значение при $\alpha = 0,05$ и числе степеней $\nu = n - 1 = 7 - 1 = 6$ равно 12,59 (см. табл. 2).

Таблица 2

Значение χ^2 -критерия Пирсона при уровне значимости 0,05 (число степеней свободы ν)

0,05	1	2	3	4	5	6
$df(\nu)$	3,84	5,99	7,81	9,49	11,07	12,59

Поскольку $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, коэффициент W считается значимым.

Можно сделать вывод, комплексный характер дефиниции «валовый региональный продукт» (трлн руб.) вытекает из комплексного взаимодействия численности занятых по субъектам Российской Федерации (млн чел.), наличия основных фондов в Российской Федерации (трлн руб.).

Часть 2. Рассмотрим, какие из факторов оказывают наибольшее влияние на валовый региональный продукт согласно табл. 1.

Построить модель оценки влияния факторов на параметр – валовый региональный продукт.

1. Осуществить выбор факторных признаков для построения двухфакторной регрессионной модели.
2. Рассчитать параметры модели.
3. Для оценки качества модели определить линейный коэффициент множественной корреляции и коэффициент детерминации.
4. Осуществить оценку значимости уравнения регрессии.
5. Оценить с помощью t -критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии.
6. Оценить влияние факторов на зависимую переменную по модели.

Выберем факторные признаки для построения двухфакторной регрессионной модели. Валовый региональный продукт – это зависимая переменная Y . В качестве независимых, объясняющих переменных выбраны: численность работников X_1 , основные фонды X_2 .

Для проведения корреляционного анализа используются инструменты MS Excel.

Таблица 3

Факторные признаки для построения двухфакторной регрессионной модели

№ п.п.	Год	Валовый региональный продукт, трлн руб. Y	Численность работников, млн чел. X_1	Основные фонды, трлн руб. X_2
1	2007	28	70,8	60,4
2	2008	33,9	71	74,4
3	2009	32	69,4	82,3
4	2010	37,7	69,9	93,2
5	2011	45,4	70,9	108
6	2012	49,9	71,5	121,2
7	2013	54	71,4	133,5

В результате перечисленных действий будет получена *матрица коэффициентов парной корреляции* (табл. 4).

Анализ матрицы коэффициентов парной корреляции показывает, что зависимая переменная (то есть валовый региональный продукт) имеет очень тесную связь с основными фондами ($r_{y,x_2} = 0,99$).

Однако факторы X_1 и X_2 имеют умеренную связь ($r_{x_1, x_2} = 0,48$), факторы Y и X_1 также имеют умеренную связь ($r_{y, x_1} = 0,62$).

Таблица 4

Матрица коэффициентов парной корреляции

Страна	Валовый региональный продукт, трлн руб. Y	Численность работников, млн чел. X_1	Основные фонды, трлн руб. X_2
Валовый региональный продукт, трлн руб. Y	1		
Численность работников, млн чел. X_1	0,62	1	
Основные фонды, трлн руб. X_2	0,99	0,48	1

Рассчитаем параметры модели в матричной форме. Оценка параметров регрессии осуществляется методом наименьших квадратов в матричной форме по формуле

$$A = (X'X)^{-1} X'Y$$

с использованием данных, приведенных в табл. 1 (для вычисления a_0 добавлен столбец x_0).

Таблица 5

Оценка параметров регрессии

№ п.п.	Год	Валовый региональный продукт, трлн руб. Y	X_0	Численность работников, млн чел. X_1	Основные фонды, трлн руб. X_2
1	2007	28	1	70,8	60,4
2	2008	33,9	1	71	74,4
3	2009	32	1	69,4	82,3
4	2010	37,7	1	69,9	93,2
5	2011	45,4	1	70,9	108
6	2012	49,9	1	71,5	121,2
7	2013	54	1	71,4	133,5

Итак, используем матричное исчисление:

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{2,1} & \dots & x_{7,1} \\ x_{1,2} & x_{2,2} & \dots & x_{7,2} \end{pmatrix} \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \dots & \dots & \dots \\ 1 & x_{7,1} & x_{7,2} \end{pmatrix}^{-1} \times \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{2,1} & \dots & x_{7,1} \\ x_{1,2} & x_{2,2} & \dots & x_{7,2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_{167} \end{pmatrix},$$

$$(X'X) = \begin{pmatrix} 7 & 494,9 & 673 \\ 494,9 & 3493 & 47639 \\ 1473 & 47639 & 3493 \end{pmatrix},$$

$$(X'X)^{-1} = \begin{pmatrix} 1744,25 & -25,13 & 0,34 \\ -25,13 & 0,36 & -0,01 \\ 0,34 & -0,01 & 0,00 \end{pmatrix},$$

$$A = (X'X)^{-1} X'Y = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -155,16 \\ 2,307 \\ 0,335 \end{pmatrix}.$$

Уравнение регрессии зависимости валового регионального продукта от численности работников и основных фондов можно записать в следующем виде:

$$y = -155,16 + 2,307x_1 + 0,335x_2$$

Расчетные значения Y определяются путем последовательной подстановки в эту модель значений факторов, взятых для каждого наблюдения.

Для проведения регрессионного анализа используем инструменты MSExcel.

В результате перечисленных действий будет получен протокол выполнения регрессионного анализа (рис. 1).

Подведение итогов

Регрессионная статистика	
Множественный R	0,998
R -квадрат	0,997
Нормированный R -квадрат	0,995
Стандартная ошибка	0,647
Наблюдения	7

Дисперсионный анализ				
	df	SS	MS	F
Регрессия	2	571,9	285,94	683,7
Остаток	4	1,67	0,42	
Итого	6	573,55		

	Коэффициенты	Стандартная ошибка	t-статистика
У-пересечение	-155,16	27	-5,74
Численность работников, X_1	2,307	0,39	5,92
Основные фонды, X_2	0,335	0,01	0,01

Вывод остатка

Наблюдение	Предсказанное, Y	Остатки
1	33,97	-0,07
2	32,51	-0,51
3	37,30	0,40
4	44,61	0,79
5	50,39	-0,49
6	54,12	-0,12

Рис. 1. Протокол выполнения регрессионного анализа

Рассмотрим содержание протокола регрессионного анализа (рис. 1), в котором отражены основные итоги расчетов. Пояснения таблицы «Регрессионная статистика» приведены в табл. 6, а «Дисперсионный анализ» – в табл. 7.

В протоколе выполнения регрессионного анализа (рис. 1) отображены коэффициенты уравнения регрессии a_0 , a_1 , a_2 , стандартные ошибки коэффициентов, t -статистика, используемая для проверки значимости коэффициентов уравнения регрессии, индекс корреляции, критерий Фишера и другие параметры.

Данные матричного исчисления и протокол уравнения регрессии совпали. Зависимость валового регионального продукта от численности работников и основных фондов можно записать в следующем виде:

$$y = -155,16 + 2,307x_1 + 0,335x_2$$

Таблица 6

Регрессионная статистика

Наименование показателя в отчете Excel	Принятые наименования	Формула
Множественный R	Коэффициент множественной корреляции, индекс корреляции	$R = \sqrt{R^2}$
R -квадрат	Коэффициент детерминации R^2	$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Нормированный R -квадрат	Скорректированный R^2	$\bar{R}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$
Стандартная ошибка	Среднеквадратическое отклонение от модели	$S_e = \sqrt{\frac{1}{n - k - 1} \sum_{i=1}^n e_i^2}$
Наблюдения	Количество наблюдений n	n

Таблица 7

Дисперсионный анализ

	df – число степеней свободы	SS – сумма квадратов	MS – среднее значение	F -критерий Фишера
Регрессия	$k = 2$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / k$	$F = \frac{R^2 / k}{(1 - R^2)(n - k - 1)}$
Остаток	$n - k - 1 = 4$	$\sum_{i=1}^n e_i^2$	$\sum_{i=1}^n e_i^2 / (n - k - 1)$	
Итого	$n - 1 = 6$	$\sum_{i=1}^n (y_i - \bar{y})^2$		

Оценим качество модели. В таблице «Вывод остатка» (рис. 1) приведены вычисленные (предсказанные) по модели значения зависимой переменной Y (столбец «Предсказанное») и значения остаточной компоненты e_i (столбец «Остатки»).

Значение коэффициентов детерминации и множественной корреляции можно найти в таблице «Регрессионная статистика» (рис. 1).

Коэффициент детерминации:

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - 69,49/365,99 = 0,99.$$

Коэффициент детерминации показывает, что около 99 % вариаций зависимой переменной учтено в модели и обусловлено влиянием включенных факторов.

Коэффициент множественной корреляции:

$$R = \sqrt{R^2} = 0,998.$$

Он показывает высокую тесноту связи зависимой переменной Y с двумя включенными в модель объясняющими факторами.

Оценим значимость уравнения регрессии. Проверку значимости проведем на основе вычисления F -критерия Фишера:

$$F = \frac{R^2/k}{(1-R^2)(n-k-1)} = \frac{0,99/2}{(1-0,99)/(7-2-1)} = 664$$

Табличное значение F -критерия можно найти с помощью функции $F.ОБР.ПХ$ (2010 Excel). При доверительной вероятности 0,95 при $\nu_1 = k = 2$ и $\nu_2 = n - k - 1 = 7 - 2 - 1 = 4$ оно составляет 4,74.

В нашем примере, то есть при $664 > 4,74$, уравнение регрессии следует признать адекватным.

Оценим с помощью t -критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии. Анализ статистической значимости параметров модели (коэффициентов регрессии) проводится с использованием t -статистики путем проверки равенства нулю j -го параметра уравнения (кроме свободного члена).

$$t_{a_j} = \hat{a}_j / S_{a_j},$$

где S_{a_j} – это стандартное (среднеквадратическое) отклонение коэффициента уравнения регрессии a_j .

Величина S_{a_j} представляет собой квадратный корень из произведения несмещенной оценки дисперсии S_e^2 и j -го диагонального элемента матрицы, обратной матрице системы нормальных уравнений:

$$S_{a_j} = S_e \sqrt{b_{jj}}$$

где b_{jj} – диагональный элемент матрицы $(X'X)^{-1}$.

Значения t -критерия для коэффициентов уравнения регрессии a_0, a_1, a_2 вычислим по формуле

$$t_{a_j} = \hat{a}_j / S_{a_j} = \hat{a}_j / (S_e \sqrt{b_{jj}}); \quad (1)$$

$$(X'X)^{-1} = \begin{pmatrix} 1744,25 & -25,13 & 0,34 \\ -25,13 & 0,36 & -0,01 \\ 0,34 & -0,01 & 0,0003 \end{pmatrix}$$

$$b_{11} = 1744,25, b_{22} = 0,36, b_{33} = 0,0003,$$

$$t_{a_0} = -155,165 / (0,662 \sqrt{1744,25}) = -5,615,$$

$$t_{a_1} = 2,307 / (0,662 \sqrt{0,36}) = 5,81,$$

$$t_{a_2} = 15,753 / 2,467 = 0,335 / (0,662 \sqrt{0,0003}) = 28,436.$$

Табличное значение t -критерия Стьюдента можно найти с помощью функции $СТЮДРАСПОБР(0,05; 4)$ (2010 Excel), при 5 %-ном уровне значимости и 4 степенях свободы оно составляет 2,78.

Так как $|t_{расч.}| > t_{табл.}$, то коэффициенты a_1, a_2 существенны (значимы).

Оценим влияние факторов на зависимую переменную по модели. Для каждого коэффициента регрессии вычислим коэффициент эластичности и бета-коэффициент соответственно по формулам (2) и (3):

$$\mathcal{E}_j = \hat{a}_j \bar{x}_j / \bar{y} \quad (2)$$

$$\mathcal{E}_1 = 2,307 \cdot 96,143 / 40,129 = 5,53,$$

$$\mathcal{E}_2 = 0,335 \cdot 96,143 / 40,129 = 0,805$$

$$\beta_j = \hat{a}_j S_{x_j} / S_y, \quad (3)$$

$$\beta_1 = 2,307 \cdot 0,77 / 9,78 = 0,183,$$

$$\beta_2 = 0,335 \cdot 26,19 / 9,78 = 0,9.$$

Коэффициент эластичности ε_1 показывает, что при увеличении численности работников на 1 млн руб. валовый региональный продукт увеличится на 5,53 трлн руб., а коэффициент β_1 показывает, что при увеличении численности работников на 0,77 млн чел, валовый региональный продукт увеличится на 1,79 трлн руб. ($0,183 \cdot 9,78 = 1,79$).

Коэффициент эластичности ε_2 показывает, что при увеличении основных фондов на 1 млн руб. валовый региональный продукт увеличится на 0,805 трлн руб., а коэффициент β_2 показывает, что при увеличении основных фондов на 26,19 млн руб., валовый региональный продукт увеличится на 8,8 трлн руб. ($0,9 \cdot 9,78 = 1,79$).

Уравнение аппроксимации полиномиального тренда 5-й степени ВВП РФ.

$$y = 0,1208x^5 - 2,5568x^4 + 20,072x^3 - 71,075x^2 + 113,13x - 31,671.$$

Надежность уравнения аппроксимации: $R^2 = 0,9996$.

ВВП, трлн руб.

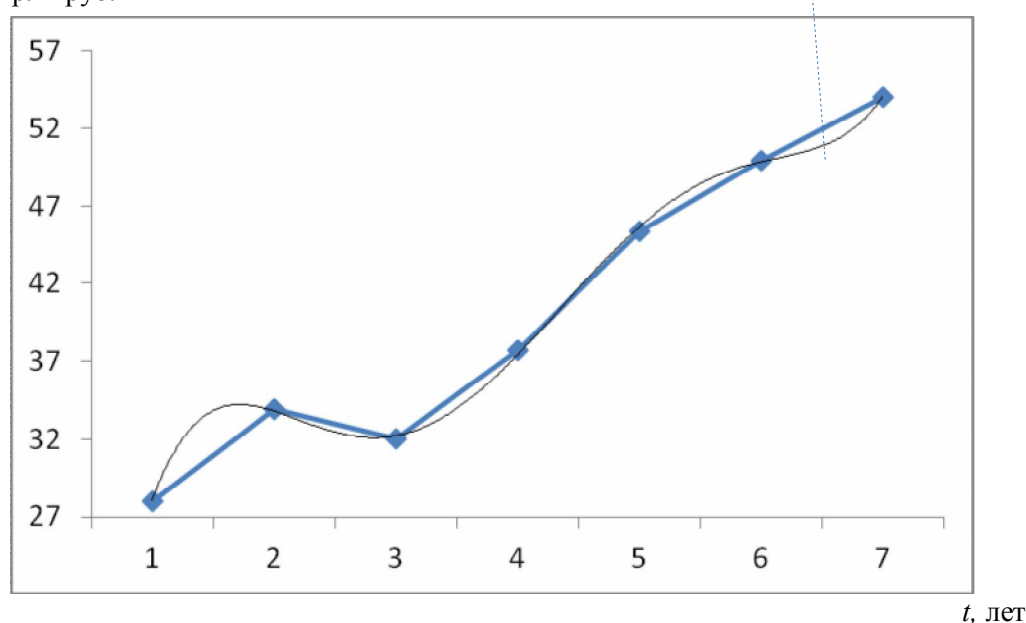


Рис. 2. Динамика валового регионального продукта РФ за 2007–2013 гг., трлн руб

В наибольшей степени на валовый региональный продукт оказывает влияние такой фактор как основные фонды, что требует развития инвестиций в капитальные вложения.

СПИСОК ЛИТЕРАТУРЫ

1. Марковина Е.В., Мухина И.А. Инвестиции: учебное пособие. М.: Флинта, 2011. 200 с.
2. Мухин А.А., Мухина И.А. Статистические методы изучения эффективности государственного и муниципального управления // Право и государство: теория и практика. 2010. № 8. С. 33-38.
3. Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование: учеб. пособие. 3-е изд., перераб. и доп. М.: НИЦ ИНФРА-М, 2013. 389 с.
4. Салин В.Н., Чурилова Э.Ю. Курс теории статистики для подготовки специалистов финансово-экономического профиля: учебник. М.: Финансы и статистика, 2007. 480 с.
5. Постановление Правительства РФ от 02.06.2008 № 420 (ред. 02.11.2013) «О Федеральной службе государственной статистики» // Собрание законодательства РФ, 09.06.2008, № 23, ст. 2710. URL: <http://www.gks.ru>. (дата обращения: 5.10.2015).

Поступила в редакцию 08.10.15

*A.A. Mukhin***NONPARAMETRIC METHODS FOR DETECTING THE INTERACTIONS OF MACROECONOMIC INDICATORS IN THE RUSSIAN FEDERATION**

The research methodology for detecting the relationships of macroeconomic indicators in the Russian Federation uses three parameters: gross regional product, number of employees, fixed assets. In the first part of the article, the author discusses the study of the relationships of quantitative traits by methods of nonparametric statistics using the coefficient of concordance (multiple rank correlation coefficient). Using the coefficient of concordance the author defines the level of communication between gross regional product (billion rubles), the number of persons employed by regions of the Russian Federation (millions of people), availability of fixed assets in the Russian Federation (billion rubles). The significance of the coefficient of concordance is evaluated by the criterion of χ^2 . In the second part of the article, a mathematical model to assess the impact of factors on the parameter – gross regional product of Russia – is constructed. Factor signs have been chosen to build a two-factor regression model for the period 2007–2013. The parameters of the model were calculated; the regression parameters were estimated using the method of least squares. To assess the model quality, the linear coefficient of multiple correlation and the coefficient of determination were calculated. Test of significance of the equation, the regression coefficients and the multiple regression equation was performed. The influence of factors on the dependent variable in the model was estimated using the coefficient of elasticity and beta coefficients. It is revealed that the growth of gross regional product is influenced mainly by such factors as fixed assets.

Keywords: nonparametric methods for detecting the interactions of macroeconomic indicators, mathematical modeling in the evaluation of the factorial dependencies of macroeconomic indicators of the Russian Federation.

Мухин Алексей Арьевич,
кандидат юридических наук, доцент
ФГБОУ ВПО «Удмуртский государственный университет»
426034, Россия, г. Ижевск, ул. Университетская, 1 (корп. 4)

Mukhin A.A.,
Candidate of Law, Associate Professor
Udmurt State University
Universitetskaya st., 1/4, Izhevsk, Russia, 426034