

# ТЕХНІЧНІ НАУКИ

DOI: <https://doi.org/10.32839/2304-5809/2019-12-76-36>

УДК 681.5.015

Архипова С.А.

Национальный технический университет Украины  
«Киевский политехнический институт имени Игоря Сикорского»

## АНАЛИЗ СТАТИСТИЧЕСКИХ ОЦЕНОК ХАРАКТЕРИСТИК РЕГРЕССИОННЫХ МОДЕЛЕЙ

**Аннотация.** Рассматривается задача оценивания качества параметрической идентификации линейной регрессионной модели в условиях априорной неопределенности относительно свойств погрешности измерения зависимой переменной и структуры модели. При решении таких задач использование традиционных методов является малоэффективным, так как эти методы обычно ориентированы на наличие определенных ограничений в отношении особенностей исходных данных, в частности, случайной ошибки измерений. При нарушении предположения о нормальности разрушается отлаженный механизм проверки адекватности модели, селекции ее структуры. Предложен метод анализа статистических оценок характеристик регрессионных моделей. Метод не требует информации о свойствах помех в исходных данных и базируется на компьютерной имитации серии "реальных" экспериментов со специальной обработкой их результатов.

**Ключевые слова:** метод варьирования данных, регрессионный анализ, ошибка измерения, нормальность распределения.

Arkhipova Sofia

National Technical University of Ukraine  
«Igor Sikorsky Kyiv Polytechnic Institute»

## ANALYSIS OF STATISTICAL EVALUATION OF CHARACTERISTICS OF REGRESSION MODELS

**Summary.** A method of statistical characteristics analysis of parameter estimates for regression models is considered. This method is not required information about disturbance properties of input data and based on computer imitation of the series of the "real" experiments with special processing their results. The task of optimization of the procedure of model parameters estimating by a sample data which is characterized by a some set of properties, in particular, a certain probability model of error, is sufficiently developed and usually solved within the framework of a general approach. Usually, a parameterized representation (model) of the error is used for the synthesis (inference) of the optimal method. This representation of error is often presented in the form of a probability distribution density. However, the actual optimality of data processing is achieved only if the adopted parameterized model sufficiently fully reflects the basic properties and characteristics of the error. That is, in the case of absence of a priori information about the type of distribution, the problem of its a posteriori estimation is arises, in which the error distribution is restored through the distribution of residuals, and it is assumed that those distributions are close to each other. To obtain the residuals, preliminary estimates of the dependent variable are calculated, which, after optimizing the parameter estimation procedure, can be further refined. In this regard, it is advisable to investigate the possibility of selecting of parametrized models of error based on the results of a posteriori analysis of residuals of actual values, which are the estimates of the initial error of measurement. In solving of practical problems, the value of noise or the parameters of their distribution are a priori unknown, which does not allow, in fact, but not formally, to use the means of classical regression analysis.

**Keywords:** data variation method, regression analysis, measurement error, normality of distribution.

**Постановка проблемы.** В общем случае задача параметризации модели по результатам апостериорного анализа оценок погрешностей не может иметь надежного решения. Более того, можно полагать, что введение любого модельного описания погрешности сопровождается потерей части информации о ее свойствах и характеристиках [4; 5].

Поэтому ориентированный на принятую модель данных оптимальный метод оценивания параметров на деле приводит к субоптимальным результатам, причем характеристики оценок, рассчитанные в рамках постулированной модели данных могут совершенно не отражать реальное качество оценок [6; 7].

**Анализ последних исследований и публикаций.** Для построения аппроксимативных моде-

лей по имеющимся экспериментальным данным существует хорошо развитый мощный теоретический и алгоритмический аппарат [1, с. 15–61; 2, с. 71–104; 3, с. 48–74]. Эффективность и несмещенность оценок параметров модели в традиционном регрессионном анализе зависит от нормальности распределения погрешности исходных данных [1, с. 167–179; 3, с. 147–152].

**Выделение нерешенных ранее частей общей проблемы.** Невыполнение ограничений регрессионного анализа, а в более общем случае неопределенность относительно свойств исходных данных, практически исключает возможность реального контроля качества решения задачи структурно-параметрической идентификации. Поэтому актуальным является развитие подходов к идентификации, мало чувствитель-

ных к уровню априорного описания свойств исходных данных, в частности свободных от необходимости использования параметризованных моделей случайных ошибок данных [5].

**В статье** предлагается методика структурно-параметрической идентификации аппроксимативных моделей в условиях неполной информации об их свойствах, в частности, о свойствах ошибки измерения.

**Изложение основного материала.** В традиционном регрессионном анализе вектор оценок  $\tilde{A} = [\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_m]^T$  рассчитывается методом наименьших квадратов (МНК) [1] по расширенной матрице исходных данных  $[Z, X] = [Y + E, X]$ , где  $Y$  – вектор значений зависимой переменной,  $E$  – вектор значений погрешности измерений,  $M\{E\} = 0$ ,  $D\{E\} = \sigma_e^2 < \infty$ ,  $X$  – матрица плана эксперимента размером  $m \times n$ ,  $n$  – количество экспериментов. Эффективность полученных МНК-оценок характеризуется матрицей ковариаций оценок [1, с. 15–40; 3, с. 64–74, 273–283].

$$D\{\tilde{A}\} = D\{(X^T X)^{-1} X^T Z\} = (X^T X)^{-1} X^T D\{Z\} X (X^T X)^{-1} = \sigma_e^2 (X^T X)^{-1}, \quad (1)$$

в частности, диагональными элементами этой матрицы – дисперсиями оценок  $\sigma^2\{\tilde{a}_j\}$ ,  $j = \overline{1, m}$  и косвенно – средним квадратом ошибки моделирования  $S^2 = \sum_{i=1}^n (z_i - \tilde{y}_i)^2 / n$ , где  $\tilde{Y} = X\tilde{A}$ .

Использование формулы (1) подразумевает знание дисперсии погрешности  $\sigma_e^2$ , на практике обычно неизвестной и являющейся предметом специального оценивания. Успех последнего в свою очередь определяется наличием априорной информации о точной структуре модели и знанием математического ожидания  $M\{S^2\}$ , которые также отсутствуют.

Обойти возникающие трудности можно, опираясь на изложенный в [6; 7; 8] подход, суть которого в имитации выполнения множества из  $L$  экспериментов, обработке полученной в каждом из них матрицы  $[Z, X]$ ,  $l = \overline{1, L}$  псевдоданных и вычислении на множестве найденных векторов МНК-оценок  $\{\tilde{A}_l\}$  выборочных дисперсий и ковариаций, а также среднего значения  $\bar{S}^2 \approx M\{S^2\}$ . Следует однако отметить, что численные значения этих выборочных характеристик могут оказаться зависимыми от используемого метода построения псевдоданных и будут отличаться от значений одноименных характеристик, рассчитанных по стандартной процедуре линейного регрессионного анализа, причем эти отличия могут носить неслучайный, вполне объективный характер [4; 5; 8].

Один из наиболее простых и надежных способов получения псевдоданных дает метод случайной вариации исходных данных. Формирование матриц псевдоданных в этом способе осуществляется путем случайного выбора строк из матрицы реальных исходных данных  $[Z, X]$ . Процедура случайного выбора продолжается до получения заданного числа  $L$  матриц псевдоданных, каждая размерностью  $m \times n$ :  $[Z_1, X_1], \dots, [Z_l, X_l], \dots, [Z_L, X_L]$ . Такой способ формирования псевдоданных гарантирует их статистическую однородность: все наборы псевдоданных получены случайной выборкой из одной и той же генеральной совокуп-

ности, в качестве которой выступает матрица реальных исходных данных.

МНК-оценки  $\tilde{A}_l = (X_l^T X_l)^{-1} X_l^T Z_l$ ,  $l = \overline{1, L}$ , рассчитанные по соответствующим матрицам псевдоданных и называемые псевдооценками, являются несмещенными оценками вектора параметров модели [4; 5]. В силу принятого способа формирования псевдоданных для матриц плана справедливо соотношение  $X_l \neq X_k$  при  $l \neq k$ ,  $l, k = \overline{1, L}$ . Следовательно, будут отличаться друг от друга и соответствующие дисперсионные матрицы  $D\{\tilde{A}_l\} = \sigma_e^2 [X_l^T X_l]^{-1}$ ,  $l = \overline{1, L}$ , а также их диагональные элементы:  $\sigma^2\{\tilde{a}_{jl}\} \neq \sigma^2\{\tilde{a}_{jk}\}$  при  $l \neq k$ ,  $j = \overline{1, m}$ .

В отличие от описанной классическому регрессионному анализу соответствует схема формирования псевдоданных при соблюдении условия  $X_l = const$ . Вариация данных в этой схеме определяется исключительно вариацией состава вектора шумов  $E$ ,  $[Z_l, X] = [Y + E_l, X]$ ,  $l = \overline{1, L}$ , т.е.  $D\{\tilde{A}_l\} = const$ ,  $\sigma^2\{\tilde{a}_{jl}\} = const$  при  $l = \overline{1, L}$ . К сожалению, значения шумов или параметры их распределения априорно неизвестны, что не позволяет практически применить этот способ построения псевдоданных.

Таким образом, отличительной особенностью формирования псевдоданных методом случайного выбора строк из матрицы исходных данных является варьирование состава матриц плана  $X_1, \dots, X_L$ , что обуславливает появление дополнительной вариации (рассеяния) псевдооценок  $\tilde{a}_1, \dots, \tilde{a}_m$  каждого из параметров, увеличение разброса значений оценок выходной переменной  $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n$  (т.к. в их расчете используются псевдооценки параметров:  $\tilde{Y} = X\tilde{A}$ ), а, следовательно, и разброса значений статистики  $S^2$ .

Как отмеченные выше эффекты сказываются на точности определения показателей качества идентификации [7; 9], рассчитываемых из варьированных данных, в частности, на дисперсиях  $\sigma^2\{\tilde{a}_{jl}\}$ ,  $j = \overline{1, m}$  и статистике  $S^2$ ? Ответить на этот вопрос, с учетом ограниченности информации о свойствах обрабатываемых данных, можно путем проведения моделирования исследуемой задачи на тестовых примерах.

Описание теста.

Тестовое уравнение регрессии, структура и коэффициенты которого считаются неизвестными потенциальному обработчику, имеет вид:

$$z = a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2 + e = f(x_1, x_2) + e, \quad (2)$$

где  $A = [1; 1; 0,04; 0,05; 0,003]^T$ ,  $x_1, x_2$  – независимые переменные,  $x_1 \in X_{1don}$ ,  $x_2 \in X_{2don}$ ,  $X_{1don}$  и  $X_{2don}$  – области допустимых значений переменных,  $y = f(x_1, x_2)$  – зависимая переменная,  $e$  – погрешность ее наблюдения, численные значения которой – значения некоторой случайной величины  $E$ , имеющей плотность вероятности  $f(e)$  и дисперсию  $\sigma_e^2$  (априорная информация о величине  $\sigma_e^2$ , виде и параметрах распределения  $f(e)$  для обработчика предполагается неизвестной).

Матрица исходных данных для теста формируется следующим образом. Матрица плана  $X = [x_{i1}, x_{i2}]$ ,  $i = \overline{1, 200}$  составляется из значений, случайным образом выбранным (сгенерированным) из соответствующих множеств  $X_{1don}$ ,  $X_{2don}$ . Для каждой пары  $(x_{i1}, x_{i2})$ ,  $i = \overline{1, n}$  по формуле

$y_i = f(x_{i1}, x_{i2})$  рассчитывается вектор значений зависимой переменной  $Y = [y_1, \dots, y_n]$ , с которым поэлементно суммируется вектор погрешностей наблюдения  $E$ . Значения последнего генерируются генератором случайных чисел с законом распределения  $f(e)$ . Суммы  $z_i = y_i + e_i, i = \overline{1, n}$ , составляют вектор результатов наблюдения зависимой переменной  $Z = [z_1, \dots, z_n]^T$ . Тройки  $(z_i, x_{i1}, x_{i2})$  образуют матрицу исходных данных  $[Z, X_1, X_2]$  размером  $200 \times 3$ . Содержимое этой матрицы – практически вся та информация, которой располагает обработчик. Кроме того, ему известен класс моделей, к которому принадлежит искомая регрессия, – линейные по параметрам модели вида

$$y = \sum_j a_j x_1^{\alpha_j} x_2^{\beta_j}, \quad \alpha_j, \beta_j = 0, 1, 2, 3, \dots \quad (3)$$

Для изучения качественных характеристик отмеченного выше роста уровня разброса оценок коэффициентов регрессии и значений статистики  $S^2$  при работе с варьированными данными был проведен следующий тестовый эксперимент.

Формировалось  $L=100$  матриц исходных данных, удовлетворяющих схеме классической регрессии: для заданной и неизменяемой матрицы плана  $X = const$  и соответствующему ей вектору  $Y = const$  путем последовательной генерации  $L$  векторов погрешностей измерений  $E_1, \dots, E_L$  составлялось  $L$  векторов наблюдений зависимой переменной  $Z_l = Y + E_l, l = \overline{1, L}$ , а затем –  $L$  матриц исходных данных  $[Z_l, X_1, X_2], [Z_2, X_1, X_2], \dots, [Z_L, X_1, X_2]$ . Кроме того, путем генерации последовательности значений переменных  $x_1, x_2$  было получено  $L$  поэлементно не совпадающих матриц плана  $[X_{1l}, X_{2l}], l = \overline{1, L}$ . В отличие от рассмотренной выше схемы варьирования данных в этом случае матрицы плана были получены не путем случайного выбора строк из исходной матрицы  $[X_1, X_2]$ , а поэлементным случайным выбором из действительной генеральной совокупности  $\{X_{1don}, X_{2don}\}$ , чем обеспечивался максимально возможный уровень варьирования состава матриц плана. По каждой из матриц плана рассчитывался свой вектор  $Y_l, l = \overline{1, L}$ , генерировалось  $L$  векторов погрешностей наблюдений и рассчитывались век-

тора  $Z_l = Y_l + E_l, l = \overline{1, L}$ . В итоге было сформировано  $L$  матриц исходных данных вида  $[z_{il}, x_{1l}, x_{2l}], l = \overline{1, L}$ , названных вариантом 2 данных, в отличие от варианта 1, которому соответствуют исходные данные, сформированные по схеме классической регрессии.

На каждом из вариантов данных рассчитывались МНК-оценки параметров ряда упорядоченных по сложности моделей (за показатель сложности принималось количество оцениваемых параметров модели). Результатом расчетов по каждой  $g$ -ой модели являлись выборочные совокупности оценок параметров  $\{\tilde{a}_{0,l}\}_g, \{\tilde{a}_{1,l}\}_g, \{\tilde{a}_{2,l}\}_g, \dots$  и выборки значений статистики  $S^2 : \{S_l^2\}_g, l = \overline{1, L}$ . Для каждой из выборок рассчитывались характеристики разброса соответствующих оценок:  $\sigma(\tilde{a}_{0g}), \sigma(\tilde{a}_{1g}), \sigma(\tilde{a}_{2g}), \dots$ , представленные в табл. 1, а также среднее значение  $\bar{S}^2$  и выборочная оценка  $\sigma\{S^2\}$ . Для компактизации записей уравнения линейных регрессионных моделей вида

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + \dots + a_k x_1^g x_2^q + \dots \quad (4)$$

даны в сокращенном представлении:  $y = (1, x_1, x_2, x_1 x_2, \dots, x_1^g x_2^q, \dots)$ .

Как и следовало ожидать, в целом разброс оценок и значения  $S^2$ , рассчитанные по варианту 1 ниже, чем для варианта 2. Однако существенное превышение характерно для моделей с неточной и загрубленной структурой. По мере уточнения структуры показатели разброса сближаются, а для истинной структуры (модель № 7) – практически совпадают. Это подтверждается информацией, представленной в табл. 2 для модели истинной структуры, содержащей выборочные оценки  $\sigma\{\tilde{a}_j\}, j = \overline{1, 5}, \bar{S}^2$  и  $\sigma\{S^2\}$ , рассчитанные по исходным данным вариантов 1, 2 и по псевдоданным, сформированным путем случайного варьирования исходных данных при использовании в качестве исходной пяти реальных матриц плана, взятых из набора данных варианта 2.

Значительно большей точности в определении показателей качества параметрической идентификации можно достигнуть, используя при построении псевдоданных только варьирование значений погрешности  $E$  с сохранением

Таблица 1

Характеристики разброса оценок параметров для всех выборок

N	Модель	Вариант данных	$\sigma\{\tilde{a}_0\}$	$\sigma\{\tilde{a}_1\}$	$\sigma\{\tilde{a}_2\}$	$\sigma\{\tilde{a}_3\}$	$\sigma\{\tilde{a}_4\}$	$\sigma\{\tilde{a}_5\}$	$\sigma\{S^2\}$
1	y(1)	1	0,849						35,70
		2	0,880						72,00
2	y(x <sub>2</sub> )	1		0,110					25,69
		2		0,161					42,34
3	y(x <sub>1</sub> , x <sub>2</sub> )	1		0,109	0,159				23,60
		2		0,172	0,211				33,27
4	y(x <sub>1</sub> , x <sub>2</sub> , x <sub>2</sub> <sup>2</sup> )	1		0,109	0,159	0,0108			17,86
		2		0,159	0,207	0,0161			19,54
5	y(1, x <sub>1</sub> , x <sub>2</sub> , x <sub>2</sub> <sup>2</sup> )	1	1,080	0,109	0,159	0,0137			17,50
		2	1,162	0,159	0,204	0,0201			19,41
6	y(x <sub>1</sub> , x <sub>2</sub> , x <sub>1</sub> <sup>2</sup> , x <sub>2</sub> <sup>2</sup> )	1		0,110	0,159	0,0084	0,015		15,10
		2		0,122	0,183	0,0076	0,016		15,30
7	y(x <sub>1</sub> , x <sub>2</sub> , x <sub>1</sub> <sup>2</sup> , x <sub>2</sub> <sup>2</sup> , x <sub>1</sub> x <sub>2</sub> <sup>2</sup> )	1		0,112	0,177	0,0088	0,015	0,001	14,97
		2		0,129	0,182	0,0072	0,014	0,001	14,49

Источник: разработка автора

Таблица 2

**Выборочные оценки для модели истинной структуры,  
рассчитанные по исходным данным двух вариантов и по псевдоданным**

Оцениваемая характеристика	Вариант 1	Вариант 2	Случайное варьирование состава данных				
			1	2	3	4	5
$\sigma\{\bar{a}_1\}$	0,1122	0,1287	0,1290	0,1302	0,1132	0,1255	0,1364
$\sigma\{\bar{a}_2\}$	0,1774	0,1818	0,1791	0,1839	0,2000	0,1723	0,1701
$\sigma\{\bar{a}_3\}$	0,0088	0,0072	0,0074	0,0050	0,0054	0,0073	0,0063
$\sigma\{\bar{a}_4\}$	0,0153	0,0140	0,0134	0,0100	0,0134	0,0156	0,0116
$\sigma\{\bar{a}_5\}$	0,0008	0,0009	0,0010	0,0008	0,0011	0,0008	0,0009
$\bar{S}^2$	123,9	124,1	122,1	128,0	141,4	133,8	110,4
$\sigma\{S^2\}$	14,97	14,49	12,28	14,71	12,41	12,12	11,92

Источник: разработка автора

Таблица 3

**Характеристики рассеяния дисперсий и корреляций оценок параметров  
при различных способах формирования псевдоданных**

$\bar{\sigma}\{\}$	Вариация		Генерация данных
	строка	значений	
$\bar{\sigma}\{\bar{\sigma}_1\}$	0,02493	0,01117	0,01059
$\bar{\sigma}\{\bar{\sigma}_2\}$	0,03253	0,01663	0,01637
$\bar{\sigma}\{\bar{\sigma}_3\}$	0,00115	0,00064	0,00058
$\bar{\sigma}\{\bar{\sigma}_4\}$	0,00211	0,00132	0,00148
$\bar{\sigma}\{\bar{\sigma}_5\}$	0,00029	0,00009	0,00008
$\bar{\sigma}\{r_{21}\}$	0,1606	0,1033	0,0817
$\bar{\sigma}\{r_{43}\}$	0,0979	0,0648	0,0534
$\bar{\sigma}\{r_{51}\}$	0,1696	0,0964	0,0871
$\bar{\sigma}\{r_{52}\}$	0,1438	0,0827	0,934
$\bar{\sigma}\{r_{53}\}$	0,1558	0,1016	0,0938
$\bar{\sigma}\{r_{54}\}$	0,1716	0,0943	0,0964

Источник: разработка автора

неизменной матрицы плана  $X$ . На практике для этого используется информация об остатках  $\bar{e}_i = z_i - \hat{y}_i$ ,  $i = \overline{1, n}$ . Формирование псевдоданных при  $X = const$  осуществляется за счет построения псевдовыборок  $\{z_{il}\}$ ,  $l = \overline{1, L}$ ,  $z_{il} = \hat{y}_i + \bar{e}_{il}$ , где очередное значение  $\bar{e}_{il}$  получается путем случайной повторной выборки из множества  $\{\bar{e}_i\}$ , используемого в качестве генеральной совокупности, либо генерацией значений случайной величины  $E$ , распределенной по закону  $f(\bar{e})$ , представляющему собой непараметрическую оценку распределения  $f(e)$ . Последний способ обеспечивает максимальную точность оценивания показателей качества идентификации, что подтверждается результатами тестовых исследований. В табл. 3 представлены характеристики рассеяния различных показателей качества параметрической идентификации: дисперсий оценок параметров и корреляционной матрицы  $R$  для этих оценок, найденные при использовании различных способов формирования псевдоданных. Следует отметить, что для третьего из рассмотренных способов высокая точность и работоспособность сохраняется при малых  $n = 10 \div 25$

и практически не зависит от вида распределения  $f(e)$ . К сожалению второй и третий способы применимы только в случае априорно известной структуры модели, так как базируются на требовании достаточно точного выделения оценок погрешности  $\bar{e}_i$ . В случае отсутствия априорной информации о структуре модели необходимо предварительное решение задачи структурной идентификации [8].

**Выводы и предложения.** При отсутствии априорной информации о нормальности распределения ошибки измерения процедура классического регрессионного анализа не позволяет надежно подобрать структуру и оценить параметры модели.

Предлагается методика структурно-параметрической идентификации аппроксимативных моделей в условиях неполной информации об их свойствах, в частности, о свойствах ошибки измерения.

Разработаны и программно-реализованы алгоритмы варьирования данных для решения подобных задач. Разработана методика исследования средств варьирования данных, с помощью которой проанализированы свойства, особенности, области применения различных средств варьирования для оценки качества задач идентификации.

**Список литературы:**

1. Демиденко Е.З. Линейная и нелинейная регрессия. Москва : Финансы и статистика, 1981. 302 с.
2. Льюнг Л. Идентификация систем. Теория для пользования. Москва : Наука, 1991. 432 с.
3. Себер Дж. Линейный регрессионный анализ. Москва : Мир, 1980. 56 с.
4. Архипов А.Е., Архипова С.А. Идентификация аппроксимативных моделей методом варьирования данных. *Адаптивні системи автоматичного управління*. Межвідом. науково-техн. зб. 1998. Вип. 1(21). С. 81–86.
5. Архипов А.Е., Архипова С.А. О некорректности параметризации распределений оценок при решении задачи идентификации. *Адаптивні системи автоматичного управління*. Межвідом. науково-техн. зб. 1999. Вип. 2(22). С. 114–121.
6. Архипов А.Е., Архипова С.А. Принцип варьирования данных в прикладных задачах идентификации. *Радіоелектроніка. Інформатика. Управління*. 2000. № 1. С. 56–60.
7. Архипов А.Е., Архипова С.А. К выбору интегрального показателя качества параметрической идентификации многомерной линейной модели. *Адаптивні системи автоматичного управління*. Межвідом. науково-техн. зб. 2000. Вип. 3(23). С. 3–8.
8. Архипова С.А. Про неможливість відновлення розподілу похибок вихідних даних в регресійних моделях. *Молодий вчений*. 2019. № 6(70). С. 147–150.

**References:**

1. Demidenko, E.Z. (1981). *Linejnaja i nelinejnaja regressija* [Linear and nonlinear regression]. Moscow: Finansy i statistika. (in Russian)
2. Ljung, L. (1991). *Identifikacija sistem. Teorija dlja pol'zovanija* [System identification. Theory to use]. Moscow: Nauka. (in Russian)
3. Seber, Dzh. (1980). *Linejnyj regressionnyj analiz* [Linear regression analysis]. Moscow: Mir. (in Russian)
4. Arkhypov, A.E., & Arhipova, S.A. (1998). Identifikatsiya approksimativnykh modeley metodom var'irovaniya dannykh [Identification of approximative models by data variation method]. *Adaptivni sistemi avtomatichnogo upravlinnya*. Mezhdidom. naukovu-tekh. zb., no. 1(21), pp. 81–86.
5. Arkhypov, A.E., & Arhipova, S.A. (1999). O nekorrektnosti parametrizatsii raspredeleniy otsenok pri reshenii zadachi identifikatsii [On the incorrect parameterization of the distribution of estimates in solving the identification problem]. *Adaptivni sistemi avtomatichnogo upravlinnya*. Mezhdidom. naukovu-tekh. zb., no. 2(22), pp. 114–121.
6. Arkhypov, A.E., & Arhipova, S.A. (2000). Printsip var'irovaniya dannykh v prikladnykh zadachakh identifikatsii [The principle of data variation in identification applied problems]. *Radioelektronika. Informatika. Upravlinnya*, no. 1, pp. 56–60.
7. Arkhypov, A.E., & Arhipova, S.A. (2000). K vyboru integral'nogo pokazatelya kachestva parametriceskoy identifikatsii mnogomernoy lineynoy modeli [On the choice of an integral quality indicator of parametric identification of a multidimensional linear model]. *Adaptivni sistemi avtomatichnogo upravlinnya*. Mezhdidom. naukovu-tekh. zb., no. 3(23), pp. 3–8.
8. Arhipova, S.A. (2019). Pro nemozhlivist' vidnovlennya rozpodilu pokhibok vikhidnykh danikh v regresiynykh modelyakh [The impossibility of restoring the error distribution of the original data in the regression models]. *Molodiy vcheniy*, no. 6(70), pp. 147–150.