DOI: http://doi.org/10.25728/cs.2022.4.1

A METHOD FOR CONSTRUCTING NONELEMENTARY LINEAR REGRESSIONS BASED ON MATHEMATICAL PROGRAMMING

M. P. Bazilevskiy

Irkutsk State Transport University, Irkutsk, Russia

✉ mik2178@yandex.ru

Abstract. This paper is devoted to constructing nonelementary linear regressions consisting of explanatory variables and all possible combinations of their pairs transformed using binary minimum and maximum operations. Such models are formalized through a 0-1 mixed integer linear programming problem. By adjusting the constraints on binary variables, we control the structural specification of a nonelementary linear regression, namely, the number of regressors, their types, and the composition of explanatory variables. In this case, the model parameters are approximately estimated using the ordinary least squares method. The formulated problem has advantages: the number of constraints does not depend on the sample size, and the signs of the estimates for the explanatory variables are consistent with the signs of their correlation coefficients with the dependent variable. Regressors are eliminated at the initial stage to reduce the time for solving the problem and make the model quite interpretable. A nonelementary linear regression of rail freight in Irkutsk oblast is constructed, and its interpretation is given.

Keywords: nonelementary linear regression, ordinary least squares method, 0-1 mixed integer linear programming problem, subset selection, coefficient of determination, interpretation, rail freight.

INTRODUCTION

In regression analysis [1, 2] based on economic data, special attention is paid to constructing production functions (PFs), i.e., mathematical relationships between production volumes (outputs) and production factors. Published in 1986, the monograph [3] was entirely devoted to the theory, methods, and application of PFs. It considered the following PFs: linear, multi-mode, Cobb–Douglas, Leontief, Allen, CES (*Constant Elasticity of Substitution*), LES (*Linear Elasticity of Substitution*), and Solow. At present, new modifications of PFs appear; they are investigated and are actively used in econometric studies [4–6]. In this paper, we construct nonelementary regression models specified on the basis of the well-known Leontief PF

$$y_i = \min\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, ..., \alpha_l x_{il}\} + \varepsilon_i, \; i = \overline{1, n}, \quad (1)$$

with the following notations: n is the sample size; l is the number of explanatory variables; y_i, $i = \overline{1, n}$, are the values of the independent variable y; x_{ij}, $i = \overline{1, n}$, $j = \overline{1, l}$, are the values of the explanatory variables x_1, $x_2, ..., x_l$; α_j, $j = \overline{1, l}$, are unknown parameters; finally, ε_i, $i = \overline{1, n}$, are approximation errors. In the theory of PFs, the variable y in equation (1) is interpreted as the output, whereas $x_1, ..., x_n$ are interpreted as the indicators of production factors.

Note that the monograph [3] also identified the "parallel" Leontief function

$$y_i = \min\{\alpha_{11} x_{i1}, \alpha_{12} x_{i2}, ..., \alpha_{1l} x_{il}\} + \ldots$$
$$\min\{\alpha_{k1} x_{i1}, \alpha_{k2} x_{i2}, ..., \alpha_{kl} x_{il}\} + \varepsilon_i, \; i = \overline{1, n}.$$

This function reflects a process where the output is composed of the outputs of k parallel production processes with fixed proportions of factors using common resources. For two production factors x_1 and x_2, the "parallel" Leontief function is called the linear programming function.

According to the monograph [7], the parameters of the Leontief PF (1) can be estimated using non-smooth optimization methods [8–10], which are often difficult to implement. Therefore, the exact estimation of the PF (1) was reduced in [7] to a 0-1 mixed integer linear programming problem (MILPP) using the least absolute deviations (LAD) method. Note that 0-1 MILPPs are also called partially Boolean linear programming problems. At the same time, the author [7] proposed an approximate estimation method for the Leontief PF based on enumerating the estimates from a preformed domain.

The paper [11] introduced a function with the opposite meaning to the PF (1):

$$y_i = \max\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, ..., \alpha_l x_{il}\} + \varepsilon_i, \quad i = \overline{1, n}, \quad (2)$$

The paper [12] considered the symbiosis of the functions (1) and (2):

$$y_i = \min\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, ..., \alpha_l x_{il}\} + \max\{\beta_1 x_{i1}, \beta_2 x_{i2}, ..., \beta_l x_{il}\} + \varepsilon_i, \; i = \overline{1, n}. \quad (3)$$

In [11] and [12], the exact estimation of the parameters of the regressions (2) and (3) was reduced to corresponding 0-1 MILPPs using the LAD method. In the modern scientific literature, there is increased attention to regression models based on mathematical programming; for example, see the papers [13–15]. An explanation is recent advances in the technology for solving 0-1 MILPPs.

This paper deals with estimating regression models specified based on the Leontief PF using the ordinary least squares (OLS) method [1, 2]. Such a problem was first formulated in [16] for the regression (1) with two explanatory variables. The paper [17] proposed a nonelementary linear regression (NLR) of the form

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \sum_{j=1}^{C_l^2} \alpha_{j+l} \min\{x_{i,\mu_{j1}}, \lambda_j x_{i,\mu_{j2}}\} + \varepsilon_i, \; i = \overline{1, n}, \quad (4)$$

with the following notations: μ_{j1} and μ_{j2}, $j = \overline{1, C_l^2}$, are elements of the first and second columns of the index matrix $\mathbf{M}_{C_l^2 \times 2}$ (its rows contain all possible combinations of index pairs of the variables); α_j, $j = \overline{0, l + C_l^2}$, and λ_j, $j = \overline{1, C_l^2}$, are unknown parameters. By assumption, all variables in equation (4) have strictly positive values.

Obviously, NLR belongs to the class of nonlinear parametric models. But if all parameters λ_j, $j = \overline{1, C_l^2}$, are assigned definite values, the regression becomes linear, and its parameters α_j, $j = \overline{0, l + C_l^2}$, can be easily estimated using the OLS method. As established in the paper [17], the OLS-optimal estimates of the NLR parameters λ_j, $j = \overline{1, C_l^2}$, belong to the intervals

$$\lambda_j \in \left(\lambda_{\min}^{(j)}, \lambda_{\max}^{(j)}\right), \; j = \overline{1, l}, \quad (5)$$

where $\lambda_{\min}^{(j)} = \min\left\{\frac{x_{1,\mu_{j1}}}{x_{1,\mu_{j2}}}, \frac{x_{2,\mu_{j1}}}{x_{2,\mu_{j2}}}, ..., \frac{x_{n,\mu_{j1}}}{x_{n,\mu_{j2}}}\right\}$ and $\lambda_{\max}^{(j)} = \max\left\{\frac{x_{1,\mu_{j1}}}{x_{1,\mu_{j2}}}, \frac{x_{2,\mu_{j1}}}{x_{2,\mu_{j2}}}, ..., \frac{x_{n,\mu_{j1}}}{x_{n,\mu_{j2}}}\right\}$. The points $\lambda_j = \lambda_{\min}^{(j)}$ and $\lambda_j = \lambda_{\max}^{(j)}$ cannot be used because of the perfect multicollinearity of the variables.

Due to these properties, an approximate OLS estimation method was proposed in [17] for the NLR (4). The method enumerates the values of the parameters λ_j, $j = \overline{1, C_l^2}$, from the intervals (5).

Unfortunately, the total number of regressors grows significantly with increasing the number l of explanatory variables in the NLR (4). Therefore, it becomes necessary to select a certain number of the most "informative" regressors [7]. Two strategies were developed for this purpose in [18]. Each strategy forms a set of alternative regressions according to a special algorithm; then the approximate OLS estimation method [17] is implemented for each regression; finally, the model with the smallest sum of the squared residuals is selected. The main disadvantage of the NLR construction approach proposed in [18] is the exhaustive search of all possible alternatives: it can take too much time to select the most informative regressors. A more promising approach involves 0-1 MILPPs; see below.

In the paper [19], the selection of the most informative regressors in linear regression estimation using the OLS method was reduced to a 0-1 MILPP. An open issue in [19] was choosing a large positive number *M* affecting both the speed and solution of the problem. It was successfully settled in the next publication [20]: the 0-1 MILPP formulated therein allows constructing a linear regression with a given number of explanatory variables, in which the signs of the OLS estimates are consistent with the signs of the correlation coefficients between the variables y and x_j, $j = \overline{1, l}$. In the course of computational experiments, the conclusion of the paper [21] was confirmed: such a problem with constraints on the signs of the coefficients is solved an order of magnitude faster than without them. In this paper, the main goal is to reduce the construction of the NLR to the 0-1 MILPP considered in the paper [20], which is efficiently solvable.

1. A METHOD FOR CONSTRUCTING NONELEMENTARY LINEAR REGRESSIONS

The NLR equation (4) contains only one binary operation, the minimum. Hereinafter, the binary minimum (maximum) is a mathematical operation with two arguments that returns their minimum (maximum). Let us supplement this regression model with regressors with the binary maximum:

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \sum_{j=1}^{C_l^2} \alpha_{j+l} \min\{x_{i,\mu_{j1}}, \lambda_j x_{i,\mu_{j2}}\} + \sum_{j=1}^{C_l^2} \alpha_{j+l+C_l^2} \max\{x_{i,\mu_{j1}}, \lambda_j x_{i,\mu_{j2}}\} + \varepsilon_i,\ i=\overline{1,n}. \tag{6}$$

The total number of regressors in equation (6), $l+2C_l^2$, is much greater than in equation (4).

Equation (6) is introduced for the first time. Therefore, we pose the following problem: formalize the procedure of constructing this model as a 0-1 MILPP. This can be done as follows.

For each parameter λ_j, $j=\overline{1, C_l^2}$, from equation (6), we determine the intervals (5). Then we evenly divide each of these intervals by p points, forming a matrix $\Lambda = \left(\lambda^*_{jk}\right)$, $j=\overline{1, C_l^2}$, $k=\overline{1, p}$. The element λ^*_{jk} of this matrix shows the k th value of the parameter λ_j for the j th pair of the variables. Replacing the unknown parameters λ_j in equation (6) with the known elements of the matrix Λ yields

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} \alpha^-_{jk} \min\left\{x_{i,\mu_{j1}}, \lambda^*_{jk} x_{i,\mu_{j2}}\right\} + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} \alpha^+_{jk} \max\left\{x_{i,\mu_{j1}}, \lambda^*_{jk} x_{i,\mu_{j2}}\right\} + \varepsilon_i,\ i=\overline{1,n}, \tag{7}$$

where α^-_{jk}, $j=\overline{1, C_l^2}$, $k=\overline{1, p}$, are the unknown parameters for regressors with the binary minimum and α^+_{jk}, $j=\overline{1, C_l^2}$, $k=\overline{1, p}$, are the unknown parameters for regressors with the binary maximum. In model (7), the total number of regressors is $l+2pC_l^2$, even exceeding that in model (6). For example, for $l=100$ variables and $p=10$ partitions, the regression (7) has 99 100 regressors.

Substituting $z^-_{ijk} = \min\left\{x_{i,\mu_{j1}}, \lambda^*_{jk} x_{i,\mu_{j2}}\right\}$ and $z^+_{ijk} = \max\left\{x_{i,\mu_{j1}}, \lambda^*_{jk} x_{i,\mu_{j2}}\right\}$, $i=\overline{1,n}$, $j=\overline{1, C_l^2}$, $k=\overline{1, p}$, into equation (7) gives the multiple linear regression model

$$y_i = \alpha_0 + \sum_{j=1}^{l} \alpha_j x_{ij} + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} \alpha^-_{jk} z^-_{ijk} + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} \alpha^+_{jk} z^+_{ijk} + \varepsilon_i,\ i=\overline{1,n}. \tag{8}$$

Following [19], let us reduce the selection of the most informative regressors for the linear regression (8) with OLS estimation to a 0-1 MILPP. First, we normalize all variables of equation (8) using the well-known rule: subtract from each value of the variable its arithmetic mean and divide the result by the standard deviation.

For model (8), we write the standardized regression equation

$$w_i = \sum_{j=1}^{l} \beta_j q_{ij} + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} \beta^-_{jk} h^-_{ijk} + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} \beta^+_{jk} h^+_{ijk} + \xi_i,\ i=\overline{1,n}, \tag{9}$$

where: w is the normalized variable y; q_j, $j=\overline{1, l}$, are the normalized variables x_j, $j=\overline{1, l}$; h^-_{jk} and h^+_{jk}, $j=\overline{1,C_l^2}$, $k=\overline{1,p}$, are the normalized variables z^-_{jk} and z^+_{jk}, $j=\overline{1,C_l^2}$, $k=\overline{1,p}$, respectively; β_j, $j=\overline{1,l}$, and β^-_{jk} and β^+_{jk}, $j=\overline{1,C_l^2}$, $k=\overline{1,p}$, are unknown standardized coefficients; finally, ξ_i, $i=\overline{1,n}$, are new approximation errors.

For model (9), the OLS estimates are given by

$$\tilde{\beta} = R^{-1}_{XX} \cdot R_{YX}, \tag{10}$$

where $R_{XX} = \begin{pmatrix} R_{xx} & R_{xz^-} & R_{xz^+} \\ R_{z^-x} & R_{z^-z^-} & R_{z^-z^+} \\ R_{z^+x} & R_{z^+z^-} & R_{z^+z^+} \end{pmatrix}$ is a correlation block matrix of dimensions $(l+2pC_l^2)\times(l+2pC_l^2)$. This matrix consists of the following blocks:

$$R_{xx} = \left(r_{x_j x_k}\right),\ j=\overline{1,l},\ k=\overline{1,l};$$

$$R_{xz^-} = \left(r_{x_s z^-_{jk}}\right),\ s=\overline{1, l},\ j=\overline{1, C_l^2},\ k=\overline{1, p};$$

$$R_{xz^+} = \left(r_{x_s z^+_{jk}}\right),\ s=\overline{1, l},\ j=\overline{1, C_l^2},\ k=\overline{1, p};$$

$$R_{z^-x} = \left(r_{z^-_{jk} x_s}\right),\ j=\overline{1, C_l^2},\ k=\overline{1, p},\ s=\overline{1, l};$$

$$R_{z^-z^-} = \left(r_{z^-_{s_1s_2} z^-_{kj}}\right),\ s_1=C_l^2,\ s_2=\overline{1, p},\ j=\overline{1, C_l^2},\ k=\overline{1, p};$$

$$R_{z^-z^+} = \left(r_{z^-_{s_1s_2} z^+_{kj}}\right),\ s_1=C_l^2,\ s_2=\overline{1, p},\ j=\overline{1, C_l^2},\ k=\overline{1, p};$$

$$R_{z^+x} = \left(r_{z^+_{jk} x_s}\right),\ j=\overline{1, C_l^2},\ k=\overline{1, p},\ s=\overline{1, l};$$

$$R_{z^+z^-} = \left(r_{z^+_{s_1s_2} z^-_{kj}}\right),\ s_1=C_l^2,\ s_2=\overline{1, p},\ j=\overline{1, C_l^2},\ k=\overline{1, p};$$

$R_{z^+z^+} = \left(r_{z^+_{s_1s_2} z^+_{kj}} \right)$, $s_1 = C_l^2$, $s_2 = \overline{1,\ p}$, $j = \overline{1,\ C_l^2}$, $k = \overline{1,\ p}$;

$R_{YX} = \left(R_{yx} \quad R_{yz^-} \quad R_{yz^+} \right)^{\mathrm{T}}$ is the correlation block vector of dimensions $\left(l + 2pC_l^2 \right) \times 1$ consisting of the blocks

$R_{yx} = \left(r_{yx_j} \right)$, $j = \overline{1,\ l}$; $R_{yz^-} = \left(r_{yz^-_{jk}} \right)$, $j = \overline{1,\ C_l^2}$, $k = \overline{1,\ p}$; $R_{yz^+} = \left(r_{yz^+_{jk}} \right)$, $j = \overline{1,\ C_l^2}$, $k = \overline{1,\ p}$.

The coefficient of determination of model (9) is given by

$$R^2 = \sum_{j=1}^{l} r_{yx_j} \beta_j + \sum_{j=1}^{C_l^2} \sum_{k=1}^{p} r_{yz^-_{jk}} \beta^-_{jk} + \sum_{j=1}^{C_l^2} \sum_{k=1}^{p} r_{yz^+_{jk}} \beta^+_{jk}. \quad (11)$$

Then, using formulas (10) and (11), we state the problem of selecting the most informative regressors for the linear regression (8):

$$R^2 \to \max, \quad (12)$$

$$-\left(1 - \delta_j\right) M \le \sum_{k=1}^{l} r_{x_j x_k} \beta_k + \sum_{s=1}^{C_l^2} \sum_{k=1}^{p} r_{x_j z^-_{sk}} \beta^-_{sk} + \sum_{s=1}^{C_l^2} \sum_{k=1}^{p} r_{x_j z^+_{sk}} \beta^+_{sk} - r_{yx_j} \le (1 - \delta_j) M,\ j = \overline{1,\ l}, \quad (13)$$

$$-\left(1 - \delta^-_{jk}\right) M \le \sum_{s=1}^{l} r_{x_s z^-_{jk}} \beta_s + \sum_{s_1=1}^{C_l^2} \sum_{s_2=1}^{p} r_{z^-_{s_1s_2} z^-_{jk}} \beta^-_{s_1s_2} + \sum_{s_1=1}^{C_l^2} \sum_{s_2=1}^{p} r_{z^+_{s_1s_2} z^-_{jk}} \beta^+_{s_1s_2} - r_{yz^-_{jk}} \le (1 - \delta^-_{jk}) M, \quad (14)$$
$$j = \overline{1,\ C_l^2},\ k = \overline{1,\ p},$$

$$-\left(1 - \delta^+_{jk}\right) M \le \sum_{s=1}^{l} r_{x_s z^+_{jk}} \beta_s + \sum_{s_1=1}^{C_l^2} \sum_{s_2=1}^{p} r_{z^-_{s_1s_2} z^+_{jk}} \beta^-_{s_1s_2} + \sum_{s_1=1}^{C_l^2} \sum_{s_2=1}^{p} r_{z^+_{s_1s_2} z^+_{jk}} \beta^+_{s_1s_2} - r_{yz^+_{jk}} \le (1 - \delta^+_{jk}) M, \quad (15)$$
$$j = \overline{1,\ C_l^2},\ k = \overline{1,\ p},$$

$$-\delta_j M \le \beta_j \le \delta_j M,\ j = \overline{1,\ l}, \quad (16)$$

$$-\delta^-_{jk} M \le \beta^-_{jk} \le \delta^-_{jk} M,\ j = \overline{1,\ C_l^2},\ k = \overline{1,\ p}, \quad (17)$$

$$-\delta^+_{jk} M \le \beta^+_{jk} \le \delta^+_{jk} M,\ j = \overline{1,\ C_l^2},\ k = \overline{1,\ p}, \quad (18)$$

$$\delta_j \in \{0, 1\},\ j = \overline{1,\ l}, \quad (19)$$

$$\delta^-_{jk} \in \{0,1\},\ j = \overline{1,\ C_l^2},\ k = \overline{1,\ p}, \quad (20)$$

$$\delta^+_{jk} \in \{0, 1\},\ j = \overline{1,\ C_l^2},\ k = \overline{1,\ p}, \quad (21)$$

$$\sum_{j=1}^{l} \delta_j + \sum_{j=1}^{C_l^2} \sum_{k=1}^{p} \delta^-_{jk} + \sum_{j=1}^{C_l^2} \sum_{k=1}^{p} \delta^+_{jk} = m, \quad (22)$$

where: m is a given number of regressors; δ_j, $j = \overline{1,\ l}$, are the Boolean variables specified by the rule

$$\delta_j = \begin{cases} 1 \text{ if the } j\text{th variable enters into the regression,} \\ 0 \text{ otherwise;} \end{cases}$$

δ^-_{jk}, $j = \overline{1,\ C_l^2}$, $k = \overline{1, p}$, are Boolean variables specified by the rule

$$\delta^-_{jk} = \begin{cases} 1 \text{ if the } j\text{th binary minimum with} \\ \text{the } k\text{th transformation enters into the regression,} \\ 0 \text{ otherwise;} \end{cases}$$

δ^+_{jk}, $j = \overline{1,\ C_l^2}$, $k = \overline{1, p}$, are Boolean variables specified by the rule

$$\delta^+_{jk} = \begin{cases} 1 \text{ if the } j\text{th binary maximum with} \\ \text{the } k\text{th transformation enters into the regression,} \\ 0 \text{ otherwise;} \end{cases}$$

finally, M is a large positive number.

An advantage of the 0-1 MILPP (12)–(22) is that the number of its constraints does not depend on the sample size n.

In the 0-1 MILPP (12)–(22), the strategy for constructing the NLR is regulated by constraints on the binary variables. Consider the following strategies.

Strategy 1. Selecting m regressors in the linear regression (7).

Here, we simply need to solve problem (12)–(22). In this case, the final model may contain several regressors with the same binary operation and with the same pair of variables but with different values of the parameter λ_j.

Strategy 2. Estimating the NLR (6) approximately using the OLS method (without selecting regressors).

Here, we need to solve the problem with the objective function (12), the constraints (13)–(21) and

$$\sum_{k=1}^{p} \delta^-_{jk} = 1,\ \sum_{k=1}^{p} \delta^+_{jk} = 1,\ j = \overline{1,\ C_l^2}.$$

(In other words, for each pair of the variables, each binary operation enters into the model with only one value of the parameter λ_j.)

Strategy 3. Selecting m regressors in the NLR (6).

Here, we need to solve the problem with the objective function (12), the constraints (13)–(22) and

$$\sum_{k=1}^{p} \delta^-_{jk} \le 1,\ \sum_{k=1}^{p} \delta^+_{jk} \le 1,\ j = \overline{1,\ C_l^2}. \quad (23)$$

Note that by adjusting the constraints on the binary variables, we can control the type of regressors in the

NLR (6). For example, adding the constraints $\sum_{j=1}^{C_l^2}\sum_{k=1}^{p}\delta_{jk}^- = 0$ and $\sum_{j=1}^{C_l^2}\sum_{k=1}^{p}\delta_{jk}^+ = 0$ to problem (12)–(22) yields the problem of selecting the most informative regressors for the linear regression; the constraints $\sum_{j=1}^{l}\delta_j = 0$ and $\sum_{j=1}^{C_l^2}\sum_{k=1}^{p}\delta_{jk}^+ = 0$, the same problem for the regression with binary minimum operations only; the constraints $\sum_{j=1}^{l}\delta_j = 0$ and $\sum_{j=1}^{C_l^2}\sum_{k=1}^{p}\delta_{jk}^- = 0$, the same problem for the regression with binary maximum operations only.

Also, it is possible to control the composition of the variables in the model. For this purpose, we introduce a binary matrix $V = \{v_{ij}\}$, $i = \overline{1, l+2pC_l^2}$, $j = \overline{1, l}$, in which

$$v_{ij} = \begin{cases} 1 \text{ if the } j\text{th variable enters into} \\ \text{the } i\text{th regressor of model (7),} \\ 0 \text{ otherwise.} \end{cases}$$

Then integrating the linear constraints

$$\sum_{j=1}^{l} v_{ij}\delta_j + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} v_{i,l+k+p(j-1)}\delta_{jk}^- + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} v_{i,l+pC_l^2+k+p(j-1)}\delta_{jk}^+ \le 1, \; i = \overline{1, l}, \tag{24}$$

into problem (12)–(22) allows constructing the NLR with m regressors into which each explanatory variable enters at most once. In this case, conditions (23) naturally hold.

Unfortunately, for problem (12)–(22), it is not completely clear how to specify large numbers M. To settle this issue, we adopt the approach from [20]. Let us replace the constraints (13)–(18) by the following ones:

$$-\left(1-\delta_j\right)M_{u_j}^- \le \sum_{k=1}^{l} r_{x_j x_k}\,\beta_k + \sum_{s=1}^{C_l^2}\sum_{k=1}^{p} r_{x_j z_{sk}^-}\,\beta_{sk}^- + \sum_{s=1}^{C_l^2}\sum_{k=1}^{p} r_{x_j z_{sk}^+}\,\beta_{sk}^+ - r_{yx_j} \le (1-\delta_j)M_{u_j}^+, \; j = \overline{1, l}, \tag{25}$$

$$-\left(1-\delta_{jk}^-\right)M_{u_{jk}^-}^- \le \sum_{s=1}^{l} r_{x_s z_{jk}^-}\,\beta_s + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^- z_{jk}^-}\,\beta_{s_1 s_2}^- + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^+ z_{jk}^-}\,\beta_{s_1 s_2}^+ - r_{yz_{jk}^-} \le (1-\delta_{jk}^-)M_{u_{jk}^-}^+, \tag{26}$$

$$j = \overline{1, C_l^2}, \; k = \overline{1, p},$$

$$-\left(1-\delta_{jk}^+\right)M_{u_{jk}^+}^- \le \sum_{s=1}^{l} r_{x_s z_{jk}^+}\,\beta_s + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^- z_{jk}^+}\,\beta_{s_1 s_2}^- + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^+ z_{jk}^+}\,\beta_{s_1 s_2}^+ - r_{yz_{jk}^+} \le (1-\delta_{jk}^+)M_{u_{jk}^+}^+, \tag{27}$$

$$j = \overline{1, C_l^2}, \; k = \overline{1, p},$$

$$0 \le \beta_j \le \delta_j M_{\beta_j}, \; j \in J_\beta^+, \tag{28}$$

$$\delta_j M_{\beta_j} \le \beta_j \le 0, \; j \in J_\beta^-, \tag{29}$$

$$0 \le \beta_{jk}^- \le \delta_{jk}^- M_{\beta_{jk}^-}, \; j, k \in J_{\beta^-}^+, \tag{30}$$

$$\delta_{jk}^- M_{\beta_{jk}^-} \le \beta_{jk}^- \le 0, \; j,k \in J_{\beta^-}^-, \tag{31}$$

$$0 \le \beta_{jk}^+ \le \delta_{jk}^+ M_{\beta_{jk}^+}, \; j, k \in J_{\beta^+}^+, \tag{32}$$

$$\delta_{jk}^+ M_{\beta_{jk}^+} \le \beta_{jk}^+ \le 0, \; j, k \in J_{\beta^+}^-, \tag{33}$$

where: J_β^+ and J_β^- are the index sets constructed from the set $\{1, 2, ..., l\}$ so that their elements satisfy the conditions $r_{yx_j} > 0$ and $r_{yx_j} < 0$, respectively; $J_{\beta^-}^+$ and $J_{\beta^-}^-$ are the index sets constructed from the set $\{\{1, 2\}, ..., \{1, p\}, \{2, 1\}, ..., \{2, p\}, ..., \{C_l^2, 1\}, ..., \{C_l^2, p\}\}$ so that their elements satisfy the conditions $r_{yz_{jk}^-} > 0$ and $r_{yz_{jk}^-} < 0$, respectively; $J_{\beta^+}^+$ and $J_{\beta^+}^-$ are the index sets constructed from the set $\{\{1, 2\}, ..., \{1, p\}, \{2, 1\}, ..., \{2, p\}, ..., \{C_l^2, 1\}, ..., \{C_l^2, p\}\}$ so that their elements satisfy the conditions $r_{yz_{jk}^+} > 0$ and $r_{yz_{jk}^+} < 0$; finally, $M_{\beta_j} = 1/r_{yx_j}$, $j = \overline{1, l}$, and $M_{\beta_{jk}^-} = 1/r_{yz_{jk}^-}$ and $M_{\beta_{jk}^+} = 1/r_{yz_{jk}^+}$, $j = \overline{1, C_l^2}$, $k = \overline{1, p}$.

To find $M_{u_j}^-$ in the constraints (25), we need to solve a series of l linear programming problems with the objective functions $u_j \to \min$ subject to the constraints

$$0 \le \beta_j \le M_{\beta_j}, \; j \in J_\beta^+, \tag{34}$$

$$M_{\beta_j} \le \beta_j \le 0, \; j \in J_\beta^-, \tag{35}$$

$$0 \le \beta_{jk}^- \le M_{\beta_{jk}^-}, \; j, k \in J_{\beta^-}^+, \tag{36}$$

$$M_{\beta_{jk}^-} \le \beta_{jk}^- \le 0, \; j, k \in J_{\beta^-}^-, \tag{37}$$

$$0 \le \beta_{jk}^+ \le M_{\beta_{jk}^+}, \; j, k \in J_{\beta^+}^+, \tag{38}$$

$$M_{\beta_{jk}^+} \le \beta_{jk}^+ \le 0, \; j, k \in J_{\beta^+}^-, \tag{39}$$

$$\sum_{k=1}^{l} r_{x_j x_k} \beta_k + \sum_{s=1}^{C_l^2}\sum_{k=1}^{p} r_{x_j z_{sk}^-} \beta_{sk}^- + \sum_{s=1}^{C_l^2}\sum_{k=1}^{p} r_{x_j z_{sk}^+} \beta_{sk}^+ - r_{yx_j} = u_j,\ j=\overline{1,\ l}, \quad (40)$$

$$\sum_{s=1}^{l} r_{x_s z_{jk}^-} \beta_s + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^- z_{jk}^-} \beta_{s_1 s_2}^- + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^+ z_{jk}^-} \beta_{s_1 s_2}^+ - r_{yz_{jk}^-} = u_{jk}^-,\ j=\overline{1,\ C_l^2},\ k=\overline{1,\ p}, \quad (41)$$

$$\sum_{s=1}^{l} r_{x_s z_{jk}^+} \beta_s + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^- z_{jk}^+} \beta_{s_1 s_2}^- + \sum_{s_1=1}^{C_l^2}\sum_{s_2=1}^{p} r_{z_{s_1 s_2}^+ z_{jk}^+} \beta_{s_1 s_2}^+ - r_{yz_{jk}^+} = u_{jk}^+,\ j=\overline{1,\ C_l^2},\ k=\overline{1,\ p}, \quad (42)$$

$$\sum_{j=1}^{l} r_{yx_j} \beta_j + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} r_{yz_{jk}^-} \beta_{jk}^- + \sum_{j=1}^{C_l^2}\sum_{k=1}^{p} r_{yz_{jk}^+} \beta_{jk}^+ \le 1. \quad (43)$$

To find $M_{u_j}^+$, we need to solve a series of l linear programming problems with the objective functions $u_j \to \max$ subject to the constraints (34)–(43). Similarly, the numbers $M_{u_{jk}^-}^-$, $M_{u_{jk}^-}^+$, $M_{u_{jk}^+}^-$, and $M_{u_{jk}^+}^+$ are obtained by solving a series of pC_l^2 linear programming problems with the objective functions $u_{jk}^- \to \min$, $u_{jk}^- \to \max$, $u_{jk}^+ \to \min$, and $u_{jk}^+ \to \max$, respectively, subject to the constraints (34)–(43).

Thus, by solving the 0-1 MILPP with the objective function (12) and the constraints (19)–(22), (25)–(33), we construct the linear regression (7) with m regressors in which the signs of the estimates of the parameters β are consistent with those of the corresponding correlation coefficients of the regressors with the variable y. In other words, the following inequalities hold: $\beta_j r_{yx_j} > 0$, $j=\overline{1,\ l}$; $\beta_{jk}^- r_{yz_{jk}^-} > 0$, $\beta_{jk}^+ r_{yz_{jk}^+} > 0$, $j=\overline{1,\ C_l^2}$, $k=\overline{1,\ p}$. The NLR construction strategy in this problem is still regulated, e.g., by constraints (23) and (24) on the binary variables.

As experimentally established in [20, 21], the 0-1 MILPP (12), (19)–(22), (25)–(33) is solved an order of magnitude faster than problem (12)–(22). Moreover, since the signs of the estimates of the parameters β are consistent with those of the corresponding correlation coefficients, the absolute contributions of the variables to the total determination R^2 are given by

$$C_{x_j}^{\text{abs}} = r_{yx_j} \beta_j,\ j=\overline{1,\ l},\ C_{z_{jk}^-}^{\text{abs}} = r_{yz_{jk}^-} \beta_{jk}^-, \quad C_{z_{jk}^+}^{\text{abs}} = r_{yz_{jk}^+} \beta_{jk}^+,\ j=\overline{1,\ C_l^2},\ k=\overline{1,\ p}. \quad (44)$$

They can be used to assess the effect of each regressor on the variable y.

We make two important remarks about the solution of problem (12), (19)–(22), (25)–(33).

Remark 1. As mentioned, the signs of the estimates of the parameters β in the solution are consistent with those of the corresponding correlation coefficients. Hence, all signs of the correlation coefficients r_{yx_j} must match the physical meaning of the variables. For this purpose, experts from the relevant subject area can be involved. Inconsistent variables should be excluded from consideration. Otherwise, the resulting regression will be difficult to interpret.

Remark 2. For example, suppose that model (8) contains the regressor $z_{11}^- = \min\{x_1, 8x_2\}$ at the parameter α_{11}^-. After the transition to the piecewise representation, the parameter α_{11}^- will have either the variable x_1 or the variable $8x_2$. If $r_{yz_{11}^-} > 0$, the estimate of the parameter α_{11}^- will surely be positive, and the variables x_1 and $8x_2$ will affect y with the plus sign. In this case, the correlation coefficients r_{yx_1} and r_{yx_2} must be positive. (Otherwise, there is a problem with interpreting the model.) On the other hand, if $r_{yz_{11}^-} < 0$, the estimate of the parameter α_{11}^- will surely be negative, and the variables x_1 and $8x_2$ will affect y with the minus sign. In this case, the correlation coefficients r_{yx_1} and r_{yx_2} must be negative. Therefore, after agreeing on the signs of the correlation coefficients r_{yx_j}, $j=\overline{1,\ l}$, with the experts, it is necessary to form the variables z_{jk}^- and z_{jk}^+, $j=\overline{1,\ C_l^2}$, $k=\overline{1,\ p}$, find their correlation coefficients with the variable y, and eliminate those not satisfying the conditions

$$(r_{yz_{jk}^-} > 0 \text{ and } r_{yx_{\mu_{j1}}} > 0 \text{ and } r_{yx_{\mu_{j2}}} > 0) \text{ or } (r_{yz_{jk}^-} < 0 \text{ and } r_{yx_{\mu_{j1}}} < 0 \text{ and } r_{yx_{\mu_{j2}}} < 0), \quad j=\overline{1,\ C_l^2},\ k=\overline{1,\ p}, \quad (45)$$

$$(r_{yz_{jk}^+} > 0 \text{ and } r_{yx_{\mu_{j1}}} > 0 \text{ and } r_{yx_{\mu_{j2}}} > 0) \text{ or } (r_{yz_{jk}^+} < 0 \text{ and } r_{yx_{\mu_{j1}}} < 0 \text{ and } r_{yx_{\mu_{j2}}} < 0), \quad j=\overline{1,\ C_l^2},\ k=\overline{1,\ p}. \quad (46)$$

Removing the contradictory variables will naturally decrease the time to construct the NLR. This time can be considerably reduced further if we supplement the expressions (45) and (46) with the conditions

$$\left| r_{yz_{jk}^-} \right| \ge r,\ \left| r_{yz_{jk}^+} \right| \ge r,\ j=\overline{1,\ C_l^2},\ k=\overline{1,\ p}, \quad (47)$$

where r is a number chosen from the interval [0, 1]. The greater the number r is, the smaller the number of variables will be, and the less time it will take to solve the problem.

2. MODELING

To construct an NLR, we collected annual statistical data on the horizon 2000–2020 for the dependent variable y (freight forward by public railway transport in Irkutsk oblast, million rubles) and 62 variables x_1, x_2, ..., x_{62}, presumably affecting y. First, 6 variables with the absolute value of the correlation coefficient with y not exceeding 0.2 were removed from the list. Then the values of correlation coefficients for the remaining 56 variables were given to 2 experts representing the East Siberian Department of the Russian Railways. They were asked to eliminate the variables for which the signs of the correlation coefficients with y did not correspond to the economic meaning of the problem. After the expertise procedure, 8 factors remained under consideration:

– the percentage of the working-age population, x_2;

– labor force (thousand people), x_3;

– the number of pensioners (thousand people), x_5;

– the number of private cars per 1000 people, x_8;

– the number of enterprises and organizations, x_{18};

– organizations' accounts payable (million rubles), x_{20};

– electricity production (billion kWh), x_{22};

– rail freight tariffs (c. u.), x_{58}.

The value of the variable x_{58} for 2001 was set equal to 1000 c. u. It was used to find the other values of the variable x_{58} using the known tariff indices.

For the selected variables, the correlation coefficients with the variable y were $r_{yx_2} = 0.785$, $r_{yx_3} = 0.543$, $r_{yx_5} = -0.483$, $r_{yx_8} = -0.446$, $r_{yx_{18}} = 0.538$, $r_{yx_{20}} = -0.204$, $r_{yx_{22}} = 0.476$, and $r_{yx_{58}} = -0.465$.

These variables affect the variable y as follows:

• The growth of the labor force x_2 and x_3, as well as the growth of the number of enterprises and organizations x_{18} and electricity production x_{22}, increases the output of products in the region, causing a higher demand for rail freight. On the other hand, an increase in the variable x_5 hinders economic development, reducing the demand for rail freight.

• The surplus of private cars x_8 reduces the demand for rail transportation (passenger and freight).

• The growth of organizations' accounts payable x_{20} has a negative impact on the regional economy: for example, it can lead to imposing various penalties.

• Higher freight tariffs x_{58} naturally reduce the demand for rail freight.

Then, the intervals (5) of the parameters λ_j were determined for each pair of the selected variables. To form the matrix Λ, we uniformly divided each interval by four points. As a result, $4C_8^2 = 112$ variables z_{jk}^-, $j = \overline{1,28}$, $k = \overline{1,4}$, were obtained with the binary minimum operation, and the same number of the variables z_{jk}^-, $j = \overline{1,28}$, $k = \overline{1,4}$, were obtained with the binary maximum operation. From the 224 variables, we excluded those not satisfying conditions (45)–(47) with $r = 0.2$ (140 variables in total). Thus, the final list included 92 variables, of which 8 were explanatory and 84 were transformed using the minimum and maximum operations.

The NLR was constructed by solving the 0-1 MILPP with the objective function (12) and the constraints (19)–(21), (25)–(33). We emphasize that the constraint (22) on the number of regressors was not applied. The constraints (24) were considered to ensure that each explanatory variable entered into the final model at most once. The LPSolve IDE solver was used to solve the 0-1 MILPPs, and a special program in the Delphi environment was developed to form mathematical models of the problems for the solver. First, the unknown numbers in the constraints (25)–(27) were calculated by the program. For that purpose, 184 linear programming problems with the corresponding objective functions and the linear constraints (34)–(43) were solved. Then, the 0-1 MILPP problem (12), (19)–(21), (24)–(33) with 284 constraints, 92 real and 92 binary variables was formulated using the calculated numbers and the developed program for the LPSolve solver. It was solved on a PC with an Intel Core i5 processor (3.40 GHz, 4 cores) and 8 GB RAM. As a result, the following NLR was constructed in approximately 30 s:

$$\begin{aligned}\tilde{y} = -24.5274 + &\underset{(13.98)}{\overset{(0.6427)}{1.1895}} \min\{x_2,\ 0.000933 x_{18}\} - \\ &\underset{(-3.361)}{\overset{(0.1129)}{0.0196}} \min\{x_5,\ 0.006754 x_{20}\} - \\ &\underset{(-2.182)}{\overset{(0.0843)}{0.0323}} \min\{x_8,\ 0.11725 x_{58}\} + \\ &\underset{(3.859)}{\overset{(0.1063)}{0.0254}} \max\{x_3,\ 23.079 x_{22}\}.\end{aligned} \tag{48}$$

Here, the numbers in parentheses below the coefficients are Student's t-test values, and the numbers in parentheses above the coefficients are the absolute contributions of the regressors to the total determination (formulas (44)). All regressors were significant by Student's t-test with the significance level $\alpha = 0.05$.

The mathematical apparatus proposed in this paper does not control the significance of NLR coefficients by Student's t-test or the absolute contributions of the variables during the regression construction procedure. For significance control, we expect to integrate special linear constraints into the 0-1 MILPP in the future.

The coefficient of determination of the NLR (48) is $R^2 = 0.946183$, indicating of high quality of the model.

The variance inflation factors for the regressors of the model (48) do not exceed 10 (no multicollinearity). Note that multicollinearity in the 0-1 MILPP cannot yet be controlled either.

Thus, the model (48) is quite interpretable.

The model (48) in the piecewise form is presented in the table.

The equations of model (48) for different ranges of variables

The NLR equation	Ranges of variables
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.00013x_{20} - 0.0038x_{58} + 0.0254x_3$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.00013x_{20} - 0.0038x_{58} + 0.5857x_{22}$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.00013x_{20} - 0.0323x_8 + 0.0254x_3$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.00013x_{20} - 0.0323x_8 + 0.5857x_{22}$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.0196x_5 - 0.0038x_{58} + 0.0254x_3$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.0196x_5 - 0.0038x_{58} + 0.5857x_{22}$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.0196x_5 - 0.0323x_8 + 0.0254x_3$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 0.0011x_{18} - 0.0196x_5 - 0.0323x_8 + 0.5857x_{22}$	$\frac{x_2}{x_{18}} \geq 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.00013x_{20} - 0.0038x_{58} + 0.0254x_3$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.00013x_{20} - 0.0038x_{58} + 0.5857x_{22}$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.00013x_{20} - 0.0323x_8 + 0.0254x_3$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.00013x_{20} - 0.0323x_8 + 0.5857x_{22}$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} \geq 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.0196x_5 - 0.0038x_{58} + 0.0254x_3$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.0196x_5 - 0.0038x_{58} + 0.5857x_{22}$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} \geq 0.117, \frac{x_3}{x_{22}} < 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.0196x_5 - 0.0323x_8 + 0.0254x_3$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} \geq 23.08$
$\tilde{y} = -24.527 + 1.1895x_2 - 0.0196x_5 - 0.0323x_8 + 0.5857x_{22}$	$\frac{x_2}{x_{18}} < 0.000933, \frac{x_5}{x_{20}} < 0.00675, \frac{x_8}{x_{58}} < 0.117, \frac{x_3}{x_{22}} < 23.08$

According to the table, the composition of the variables affecting y changes depending on the conditions satisfied, and the parameter estimates $\lambda^{-}_{4,1} = 0.000933$, $\lambda^{-}_{16,2} = 0.00675$, $\lambda^{-}_{22,2} = 0.117$, and $\lambda^{+}_{12,3} = 23.08$ play the role of switching points for the following four automatically generated indicators:

– the ratio of the percentage of the working-age population (x_2) to the number of enterprises and organizations (x_{18});

– the ratio of the number of pensioners (x_5) to organizations' accounts payable (x_{20});

– the ratio of the number of private cars per 1000 people (x_8) to the rail freight tariffs (x_{58});

– the ratio of labor force (x_3) to electricity production (x_{22}).

Then the following interpretation is valid.

• If the indicator x_2/x_{18} is not smaller than 0.000933, the number of enterprises and organizations x_{18} will affect freight forward, whereas the percentage of the working-age population x_2 will have no effect. For example, increasing the number of enterprises and organizations x_{18} by 1 (under fixed values of the other variables) raises the freight forward y by 0.0011 million rubles on average. However, if the indicator x_2/x_{18} is less than 0.000933, the percentage of the working-age population x_2 will affect freight forward, whereas the number of enterprises and organizations x_{18} will have no effect. For example, increasing the percentage of the working-age population x_2 by 1% (under fixed values of the other variables) raises the freight forward y by 1.1895 million rubles on average.

• If the indicator x_5/x_{20} is not smaller than 0.00675, organizations' accounts payable x_{20} will affect freight forward, whereas the number of pensioners x_5 will have no effect. For example, increasing organizations' accounts payable x_{20} by 1 million rubles (under fixed values of the other variables) reduces the freight forward y by 0.00013 million rubles on average. However, if the indicator x_5/x_{20} is less than 0.00675, the number of pensioners x_5 will affect freight forward, whereas organizations' accounts payable x_{20} will have no effect. For example, increasing the number of pensioners x_5 by 1000 people (under fixed values of the other variables) reduces the freight forward y by 0.0196 million rubles on average.

• If the indicator x_8/x_{58} is not smaller than 0.117, the rail freight tariffs x_{58} will affect freight forward, whereas the number of private cars x_8 per 1000 people will have no effect. For example, increasing the rail freight tariffs x_{58} by 1 c.u. (under fixed values of the other variables) reduces the freight forward y by 0.0038 million rubles on average. However, if the indicator x_8/x_{58} is less than 0.117, the number of private cars x_8 per 1000 people will affect freight forward, whereas the rail freight tariffs x_{58} will have no effect. For example, increasing the number of private cars x_8 per 1000 people by 1 (under fixed values of the other variables) reduces the freight forward y by 0.0323 million rubles on average.

• If the indicator x_3/x_{22} is not smaller than 23.08, the labor force x_3 will affect freight forward, whereas the electricity production x_{22} will have no effect. For example, increasing the labor force x_3 by 1 thousand people (under fixed values of the other variables) raises the freight forward y by 0.0254 million rubles on average. However, if the indicator x_3/x_{22} is less than 23.08, the electricity production x_{22} will affect freight forward, whereas the labor force x_3 will have no effect. For example, increasing the electricity production x_{22} by 1 billion kWh (under fixed values of the other variables) raises the freight forward y by 0.5857 million rubles on average.

Thus, the interpretative characteristics of the NLR are richer and more diverse than those of the traditional linear regression model. Moreover, depending on the chosen construction strategy, the approximation characteristics of the NLR should in most cases exceed the same characteristics of linear regressions, which are only a particular case of the NLR. The proposed NLR are valuable: besides forecasting, they extract new interpretable mathematical laws to improve the efficiency of managerial decisions in various sectors of the economy.

Also, note that the NLR better suits modeling under multicollinearity conditions than the traditional linear regression. The more binary operations the NLR has, the higher the number of its degrees of freedom will be as compared to the linear regression. This means that the NLR can "accommodate" more variables with fewer regressors than the linear regression. For example, the NLR (48) contains only 4 regressors but 8 variables, so the chance of its multicollinearity is a priori lower compared to a linear regression with all 8 variables.

CONCLUSIONS

This paper has considered the NLR with the binary minimum and maximum operations. We have proposed an NLR construction method based on solving a 0-1 MILPP. The solution of this problem yields the structural specification of the NLR and its approximate OLS estimates. As shown, the structural specification of the NLR is regulated through constraints on the binary variables. The contradictory variables have been eliminated at the initial stage to reduce the solution time of the problem and make the NLR quite interpretable. The proposed method has been applied to model rail freight in Irkutsk oblast; the resulting NLR has revealed new rail freight regularities not available within classical linear regression analysis.

The method proposed above is universal and can be used to construct NLRs in any subject area based on statistical data with positive variables only. The parameter partitioning procedure forms a 0-1 MILPP; for a sufficiently large number of partitions, its optimal solution gives estimates slightly differing from the optimal OLS estimates of the NLR. Naturally, increasing the number of partitions requires more time to solve the problem. Nevertheless, as demonstrated in [20, 21] on the linear regression example, such a 0-1 MILPP is solved an order of magnitude faster compared to standard enumeration procedures. The speed of constructing NLRs for different-size samples using the proposed method will be tested in subsequent publications.

REFERENCES

1. Arkes, J., *Regression Analysis: A Practical Introduction*, Routledge, 2019.
2. Westfall, P.H. and Arias, A.L., *Understanding Regression Analysis: A Conditional Distribution Approach*, Chapman and Hall/CRC, 2020.
3. Kleiner, G.B., *Proizvodstvennye funktsii: Teoriya, metody, primenenie* (Production Functions: Theory, Methods, and Application), Moscow: Finansy i Statistika, 1986. (In Russian.)
4. Onalan, O. and Basegmez, H., Estimation of Economic Growth Using Grey Cobb–Douglas Production Function: An Application for US Economy, *Journal of Business Economics and Finance*, 2018, vol. 7, no. 2, pp. 178–190.
5. Yankovyi, O., Koval, V., Lazorenko, L., et al., Modeling Sustainable Economic Development Using Production Functions, *Studies of Applied Economics*, 2021, vol. 39, no. 5.
6. Ishikawa, A., Why Does Production Function Take the Cobb-Douglas Form?, in *Statistical Properties in Firms' Large-scale Data*, Springer, Singapore, 2021, pp. 113–135.
7. Noskov, S.I., *Tekhnologiya modelirovaniya ob''ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh* (A Technology for Modeling Objects with Unstable Operation and Data Uncertainty), Irkutsk: Oblinformpechat', 1996. (In Russian.)
8. Shor, N.Z., *Metody minimizatsii nedifferentsiruemykh funktsii i ikh prilozheniya* (Minimization Methods for Nondifferentiable Functions and Their Applications), Kiev: Naukova dumka, 1979. (In Russian.)
9. Scaman, K., Bach, F., Bubeck, S., et al., Optimal Algorithms for Non-smooth Distributed Optimization in Networks, *Advances in Neural Information Processing Systems*, 2018, vol. 31.
10. Khamaru, K. and Wainwright, M.J., Convergence Guarantees for a Class of Non-convex and Non-smooth Optimization Problems, *Journal of Machine Learning Research*, 2019, vol. 20, no. 154, pp. 1–52.
11. Ivanova, N.K., Lebedeva, S.A., and Noskov, S.I., Parameter Identification for Some Nonsmooth Regressions, *Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2016, no. 17, pp. 107–110. (In Russian.)
12. Noskov, S.I. and Khonyakov, A.A., A Software Complex for Constructing Some Types of Piecewise Linear Regressions, *Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2019, no. 3 (4), pp. 47–55. (In Russian.)
13. Park, Y.W. and Klabjan, D., Subset Selection for Multiple Linear Regression via Optimization, *Journal of Global Optimization*, 2020, vol. 77, pp. 543–574.
14. Chung, S., Park, Y.W., and Cheong, T., A Mathematical Programming Approach for Integrated Multiple Linear Regression Subset Selection and Validation, *Pattern Recognition*, 2020, vol. 108, p. 107565.
15. Bertsimas, D. and Li, M.L., Scalable Holistic Linear Regression, *Operations Research Letters*, 2020, vol. 48, no. 3, pp. 203–208.
16. Bazilevskiy, M.P., OLS-Estimation of Two-Factor Regression Models Specified on Leontiev Functions, *South-Siberian Scientific Bulletin*, 2019, no. 2 (26), pp. 66–70. (In Russian.)
17. Bazilevskiy, M.P., Estimation Linear Nonelementary Regression Models Using Ordinary Least Squares, *Modeling, Optimization, and Information Technology*, 2020, vol. 8, no. 4 (31). (In Russian.)
18. Bazilevskiy, M.P., Selection of Informative Operations in the Construction of Linear Nonelementary Regression Models, *International Journal of Open Information Technologies*, 2021, vol. 9, no. 5, pp. 30–35. (In Russian.)
19. Bazilevskiy, M.P., Reduction the Problem of Selecting Informative Regressors When Estimating a Linear Regression Model by the Method of Least Squares to the Problem of Partial-Boolean Linear Programming, *Modeling, Optimization, and Information Technology*, 2018, vol. 6, no. 1 (20), pp. 108–117. (In Russian.)
20. Bazilevskiy, M.P., Method for the M Parameter Determination in 0-1 Mixed-Integer Linear Programming Problem for Subset Selection in Linear Regression, *Bulletin of the Technological University*, 2022, vol. 25, no. 2, pp. 62–66. (In Russian.)
21. Konno, H. and Yamamoto, R., Choosing the Best Set of Variables in Regression Analysis Using Integer Programming, *Journal of Global Optimization*, 2009, vol. 44, pp. 273–282.

This paper was recommended for publication by M.I. Geraskin, a member of the Editorial Board.

Received April 24, 2022, and revised August 3, 2022.
Accepted August 31, 2022

Author information

Bazilevskiy, Mikhail Pavlovich. Cand. Sci. (Eng.), Irkutsk State Transport University, Irkutsk, Russia
✉ mik2178@yandex.ru

Cite this paper

Bazilevskiy, M.P. A Method for Constructing Nonelementary Linear Regressions Based on Mathematical Programming. *Control Sciences* **4,** 2–11 (2022). http://doi.org/10.25728/cs.2022.4.1

Translated into English by *Alexander Yu. Mazurov*,
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com