СИСТЕМНИЙ АНАЛІЗ ТА НАУКА ПРО ДАНІ

DOI: https://doi.org/10.20535/kpisn.2026.1.350247
UDC 519.237.5+336.201.2

Vadim V. Romanuke*, https://orcid.org/0000-0001-9638-9572
Vinnytsia Institute of Trade and Economics of State University of Trade and Economics,
Vinnytsia, Ukraine, https://ror.org/03r6hpz93
*Corresponding author: romanukevadimv@gmail.com

DYNAMIC LOGISTIC REGRESSION MODELING FOR BANKRUPTCY RISK PREDICTION IN UKRAINIAN BUILDING SECTOR

Background. Financial distress and bankruptcy forecasting has gained significant importance in the context of post-war economic recovery and restructuring of Ukrainian industries. Firms in the building-and-construction materials sector operate under high uncertainty, where early detection of insolvency risk is crucial for maintaining financial stability. Logistic regression models, widely used in environmental and risk analytics, can be adapted to represent the nonlinear transition from solvency to bankruptcy as a probabilistic process.

Objective. The paper aims to develop and evaluate both static and dynamic logistic regression models for predicting the potential bankruptcy of a representative Ukrainian building-materials manufacturer. The dynamic extension seeks to capture the temporal persistence in financial performance through lagged predictors.

Methods. A synthetic monthly dataset (5 years, 60 observations) is generated to simulate realistic financial ratios, including liquidity, leverage, profitability, efficiency, and interest coverage (solvency). The models are estimated in MATLAB using maximum-likelihood logistic regression with L2 regularisation (ridge penalty) to retain correlated predictors. The dynamic model incorporated one-period lags of all financial ratios and the one-period-lagged response. Predictive performance is assessed by accuracy, precision, recall, F1-score, and the confusion matrix.

Results. The static logistic model achieved an average accuracy of around 89 %, yet it failed to predict two bankruptcy-risky months out of six ones. The dynamic model improved performance to 94 % accuracy, without missing a bankruptcy-risky month, but falsely labelling a non-risky month as bankruptcy-risky one. The signs of estimated coefficients are consistent with economic logic: higher leverage increases bankruptcy probability, whereas greater liquidity, profitability, efficiency, and solvency reduce it.

Conclusions. Dynamic L2-regularised logistic regression provides an interpretable and computationally efficient framework for early bankruptcy prediction in Ukrainian industrial firms. The inclusion of lagged financial indicators enhances predictive stability and timeliness, enabling practical early-warning applications.

Keywords: bankruptcy prediction; logistic regression; dynamic modelling; financial ratios; L2 regularisation; early-warning system; Ukrainian building sector.

Introduction

The probability of firm bankruptcy remains one of the central concerns in modern financial analytics, especially in economies exposed to structural transformations and unstable market conditions [1, 2]. In Ukraine, the construction and building-materials sectors have faced recurrent disruptions caused by macroeconomic turbulence, exchange rate volatility, and evolving regulatory frameworks. These circumstances

Пропозиція для цитування цієї статті: В.В. Романюк, «Моделі динамічної логістичної регресії для прогнозування ризику банкрутства у будівельній галузі України», *Наукові вісті КПІ*, № 1, с. 7–18, 2026. doi: https://doi.org/10.20535/kpisn.2026.1.350247

Offer a citation for this article: Vadim V. Romanuke, “Dynamic logistic regression modeling for bankruptcy risk prediction in Ukrainian building sector”, *KPI Science News*, No. 1, pp. 7–18, 2026. doi: https://doi.org/10.20535/kpisn.2026.1.350247

increase the risk of financial distress among manufacturing enterprises and, consequently, create the need for quantitative models capable of providing early warning signals about potential bankruptcy. Reliable bankruptcy prediction models not only help firms assess their own financial sustainability but also support creditors, investors, and policymakers in managing credit and investment risks more effectively [3, 4].

A considerable body of research on bankruptcy prediction has emerged since the mid-20th century, ranging from discriminant analysis [5] and logit models [6] to more recent approaches employing machine learning and hybrid ensemble techniques [7, 8]. Among these, logistic regression continues to occupy a prominent position because of its interpretability, robustness, and statistical grounding [9, 10]. Logistic models explicitly connect the probability of bankruptcy to a vector of financial indicators (such as liquidity, leverage, profitability, and efficiency), allowing the analyst to quantify the marginal impact of each ratio on the likelihood of financial failure. Moreover, unlike linear discriminant methods, logistic regression does not impose the assumption of normally distributed predictors, which makes it particularly suitable for financial data often characterised by skewness, outliers, and bounded ratios [11].

While static logistic models have proven useful, they fail to account for the temporal dynamics inherent in financial distress processes. A firm's transition toward insolvency rarely occurs as a sudden event. Instead, it develops gradually as liquidity deteriorates, leverage increases, or profitability declines over successive periods. Therefore, incorporating lagged predictors into the logistic framework enables one to capture persistence and delayed effects of financial indicators on bankruptcy risk. This leads to a dynamic logistic regression model, where the current probability of bankruptcy depends not only on present-period ratios but also on their historical trajectories.

However, the inclusion of multiple correlated and lagged predictors increases the risk of multicollinearity and model instability. To mitigate this, L2 regularisation (ridge penalty) can be introduced into the likelihood function [12]. Regularisation shrinks coefficient magnitudes toward zero without eliminating predictors entirely, thereby preserving all available financial information while controlling overfitting [13, 14]. This makes the model more stable and generalizable, particularly when the number of predictors approaches or exceeds the number of observed periods, which is a typical limitation in firm-level bankruptcy datasets.

Problem statement

The objective of this study is to construct and analyse a regularised dynamic logistic regression model for bankruptcy prediction in the context of a Ukrainian building-materials manufacturer. A synthetic dataset will be generated to emulate realistic financial ratios and their temporal dependencies, reflecting the operational specifics of a mid-sized construction-related enterprise. The model will be estimated by maximum penalised likelihood, with predictive performance evaluated through classification metrics such as accuracy, precision, recall, F1-score, and the confusion matrix. The results will illustrate how the proposed approach can provide interpretable and quantitatively consistent insights into bankruptcy risk even in data-constrained environments.

Notation and dynamic logistic model

Let $b(t)$ be the indicator of bankruptcy at time period t for the firm: $b(t) = 1$ means the bankruptcy in the next period, $b(t) = 0$ means no bankruptcy in the next period. Let

$$\mathbf{X}(t) = [x_k(t)]_{1\times K} \in \mathbb{R}^K \quad (1)$$

be a vector of contemporaneous financial ratios observed at period t. We also include lagged values up to L lags of predictors and N lags of the response. Define the stacked predictor vector:

$$\mathbf{Y}_{(L+1)K}(t) = [y_i(t)]_{1\times(L+1)K} \in \mathbb{R}^{(L+1)K} \quad (2)$$

by

$$y_k(t) = x_k(t) \text{ for } k = \overline{1,K} \text{ and } y_k(t-l) = x_k(t-l) \text{ for } k = \overline{1,K} \text{ and } l = \overline{1,L}. \quad (3)$$

The vector of lagged responses is

$$\mathbf{B}_{-N}(t) = [b(t-j)]_{1\times N} \in \mathbb{U}_{0-1}^N \subset \mathbb{R}^N, \quad (4)$$

where $\mathbb{U}_{0-1}^N$ is the unit hypercubic lattice in $\mathbb{R}^N$ whose vertices are of only 0's and 1's.

The dynamic logistic model assumes that

$$\ln\left(\frac{P\left[b(t)=1 \mid \mathbf{Y}_{(L+1)K}(t); \mathbf{B}_{-N}(t)\right]}{1-P\left[b(t)=1 \mid \mathbf{Y}_{(L+1)K}(t); \mathbf{B}_{-N}(t)\right]}\right) = \alpha_0 + \sum_{i=1}^{(L+1)K} \alpha_i y_i(t) + \sum_{j=1}^{N} \mu_j b(t-j) = \alpha_0 + \mathbf{A}_{(L+1)K} \cdot \left[\mathbf{Y}_{(L+1)K}(t)\right]^{\mathrm{T}} + \mathbf{M}_N \cdot [\mathbf{B}_{-N}(t)]^{\mathrm{T}}, \quad (5)$$

where

$$P\left[b(t)=1 \mid \mathbf{Y}_{(L+1)K}(t); \mathbf{B}_{-N}(t)\right]=p_t \qquad (6)$$

is the conditional probability of bankruptcy at time period $t+1$, denoted by p_t,

$$\mathbf{A}_{(L+1)K}=[\alpha_i]_{1\times(L+1)K} \in \mathbb{R}^{(L+1)K} \qquad (7)$$

is a vector of $(L+1)K$ model predictor parameters,

$$\mathbf{M}_N=\left[\mu_j\right]_{1\times N} \in \mathbb{R}^N \qquad (8)$$

is a vector of N model lagged-response parameters, and α_0 is the intercept parameter [9].

So, overall there are K contemporaneous and $LK+N$ lagged predictors in model (5), gathered in vectors (7) and (8), respectively. We estimate them by penalised maximum likelihood with an L2 (ridge) penalty:

$$\max_{\{\alpha_0, \mathbf{A}_{(L+1)K}, \mathbf{M}_N\}} \left(\sum_{t=1}^{T} [b(t)\ln p_t + (1-b(t))\ln(1-p_t)] - \right.$$
$$\left. -\frac{\lambda}{2}\sum_{i=1}^{(L+1)K}\alpha_i^2 - \frac{\lambda}{2}\sum_{j=1}^{N}\mu_j^2 \right), \qquad (9)$$

where conditional probability (6) is deduced from (5) as

$$p_t = \left[1 + \exp(-\alpha_0 - \mathbf{A}_{(L+1)K}\cdot[\mathbf{Y}_{(L+1)K}(t)]^{\mathrm{T}} - \right.$$
$$\left. -\mathbf{M}_N \cdot [\mathbf{B}_{-N}(t)]^{\mathrm{T}})\right]^{-1}, \qquad (10)$$

and where λ is the regularisation hyperparameter, $\lambda \geq 0$. The ridge (L2 regularisation) is chosen because it shrinks coefficients (mitigating variance and multicollinearity) without performing variable selection; this preserves all financial ratios and their lags for interpretation.

Synthetic data design for a building-materials manufacturer in Ukraine

We must generate a realistic synthetic time series of financial ratios for a mid-sized building-materials manufacturer and a corresponding binary outcome indicating whether the firm will declare bankruptcy in the next period. Our time span is five years with monthly observations (quarterly observations might have fitted better, but this would elongate the span to 15 years, which is quite unreliable due to the current situation in Ukraine). Denote the time span length by $T=60$. This gives enough temporal depth for lagged effects while remaining plausible for a single firm.

In our synthetic dataset, the total number of predictors (independent variables) corresponds to the main financial ratio categories presented in Table 1. The liquidity below 1 is considered weak, whereas the liquidity above 3 may mean over-capitalised. For many manufacturing firms the leverage is about 0.4 to 1.0. Capital intensive sectors may accept higher leverage, but beyond 1 it is often seen as riskier. The profitability is very dependent on industry margin norms; the building-materials sector may have lower profitability due to cost-and-price pressures. For many manufacturing firms the profitability is about 0.05 to 0.15, and value 0.2 is considered top strong. The efficiency for manufacturing might be 0.5 to 2.0 times per year, but the efficiency ratios vary hugely: inventory turnover, receivables turnover, asset turnover – all have different benchmark ranges. The solvency above 3 is often considered safe, whereas values below 2 indicate risk.

To approximately cover the typical ranges of the predictors, we generate their values with using

Table 1. Main financial ratio categories used as predictors

Name of the predictor	Denotation	Financial meaning	Example formula (simplified)	Typical benchmark range	Role in bankruptcy prediction
Current ratio	$x_1(t)$	Liquidity	Current assets / Current liabilities	0.4–3.5	Low → risk of insolvency
Debt-to-equity ratio	$x_2(t)$	Leverage	Total debt / Equity	0.1–0.95	High → risk of default
Return on assets	$x_3(t)$	Profitability	Net income / Total assets	−0.15–0.20	Low or negative → poor performance
Asset turnover	$x_4(t)$	Efficiency	Sales / Total assets	0.2–2.0	Low → inefficient resource use
Interest coverage ratio	$x_5(t)$	Solvency	EBIT (earnings before interest and taxes) / Interest expense	0.5–6.5	Low → high debt servicing stress

values ρ of uniformly distributed random variable on interval (0; 1):

$$x_1(t) = 0.8 + 1.2\rho\,, \qquad (11)$$

$$x_2(t) = 0.2 + 1.8\rho\,, \qquad (12)$$

$$x_3(t) = -0.05 + 0.15\rho\,, \qquad (13)$$

$$x_4(t) = 0.3 + 1.2\rho\,, \qquad (14)$$

$$x_5(t) = 0.5 + 6.5\rho\,. \qquad (15)$$

The descriptive statistics of the dataset is presented in Table 2, where we can see that the data generated by (11)–(15) are successfully covered by plausible real-world ratio intervals in Table 1.

Table 2. Descriptive statistics of the synthetic time series for a building-materials manufacturer in Ukraine

Denotation	Predictor	Minimum	Mean	Maximum	Standard deviation
$x_1(t)$	Liquidity	0.8218	1.4009	1.9944	0.3326
$x_2(t)$	Leverage	0.2085	1.2088	1.9927	0.5201
$x_3(t)$	Profitability	−0.0496	0.0318	0.0972	0.0405
$x_4(t)$	Efficiency	0.3036	0.9002	1.4901	0.3777
$x_5(t)$	Solvency	0.5068	3.3272	6.9835	1.9405

The total number of events $b(t) = 1$ is 22, which is 36.67 % (Fig. 1, where higher and lower bars correspond to 1 and 0, respectively). This number is obtained by following interpretable if-then rules built from the ratios in Table 1. This is due to a stakeholder (owner of the firm), for example, sees only the values of those five financial indicators and decides whether $b(t) = 0$ or $b(t) = 1$ based on them. The stakeholder (owner) does not know anything about those parameters (7) and (8) or model (10). So, a realistic way of assigning the value of $b(t)$ (binary risk of bankruptcy) must be based on only values of liquidity, leverage, profitability, efficiency, solvency at a time t and, plausibly, at a few previous time periods $t - 1, \ldots, t - L$. For instance, $b(t) = 1$ if $x_2(t) > 1.025$, but $b(t) = 0$ if $b(t - 1) = 1$. Above that, $b(t) = 0$ if the values of liquidity, profitability, efficiency, solvency are simultaneously above their 90 % means.

The suggested way of generating a synthetic dataset for a building-materials manufacturer is very interpretable. Although it is brittle, not smooth, and can lead to either many or very few events unless tuned, it reflects the current economic instability and riskiness of Ukrainian building-materials manufacturers. The bankruptcy binary risk presented in Fig. 1 openly exposes the situation for the past five years, where the longest span without bankruptcy risk is just three months (there are three such quarters). This piano-like picture is quite realistic for a mid-sized building-materials manufacturer in Ukraine.

Estimation and software

In the case of five predictors and one lag, including one-lagged-response predictor, model (10) is simplified to:

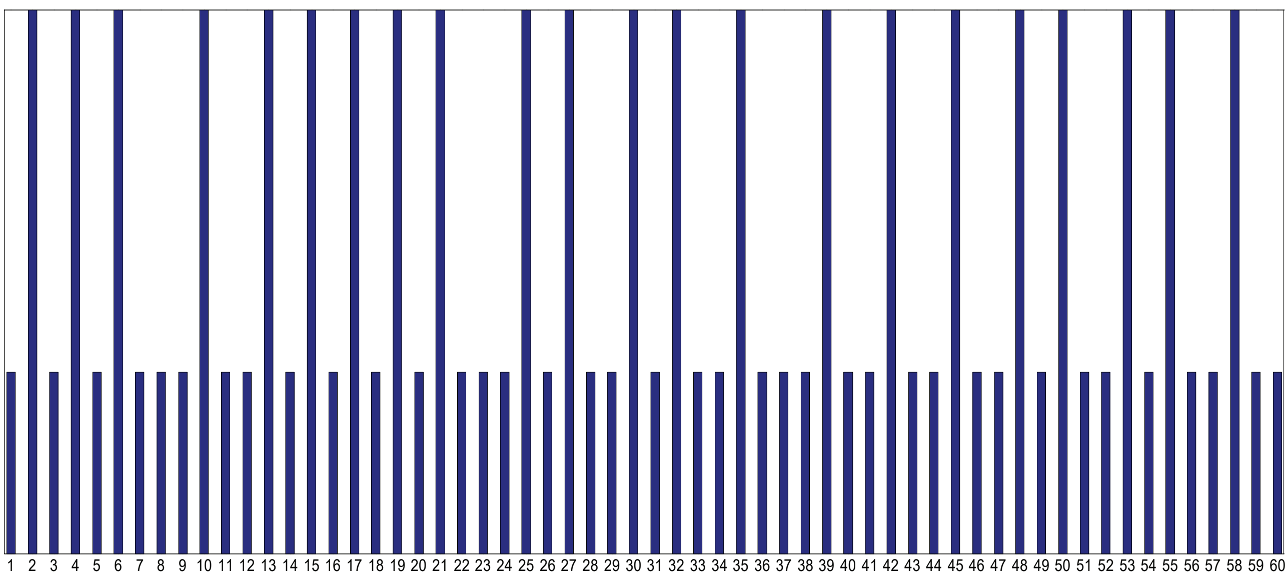

Fig. 1. Binary risk of bankruptcy (higher bars) throughout 60 periods

$$P[b(t)=1 \mid \mathbf{Y}_{10}(t); \mathbf{B}_{-1}(t)] = p_t =$$
$$= \left[1+\exp\left(-\alpha_0 - \mathbf{A}_{10}\cdot[\mathbf{Y}_{10}(t)]^{\mathrm{T}} - \mathbf{M}_1\cdot[\mathbf{B}_{-1}(t)]^{\mathrm{T}}\right)\right]^{-1} =$$
$$= \left[1+\exp\left(-\alpha_0 - \mathbf{A}_{10}\cdot[\mathbf{Y}_{10}(t)]^{\mathrm{T}} - \mu_1\cdot b(t-1)\right)\right]^{-1}. \quad (16)$$

Dynamic logistic regression model (16) is built by solving problem (9) as

$$\max_{\{\alpha_0, \mathbf{A}_{10}, \mu_1\}} \left(\sum_{t_1\in \mathrm{T}_1} \left(\alpha_0 + \mathbf{A}_{10}\cdot[\mathbf{Y}_{10}(t_1)]^{\mathrm{T}} + \mu_1\cdot b(t_1-1)\right) + \right.$$
$$+\sum_{t=1}^{T} \ln\left(\alpha_0 + \mathbf{A}_{10}\cdot[\mathbf{Y}_{10}(t)]^{\mathrm{T}} + \mu_1\cdot b(t-1)\right) -$$
$$\left. -\frac{\lambda}{2}\sum_{i=1}^{10}\alpha_i^2 - \frac{\lambda}{2}\cdot\mu_1^2 \right) \quad (17)$$

by

$$b(t_1) = 1 \ \ \forall t_1 \in \mathrm{T}_1 \subset \left\{\overline{1,T}\right\} \text{ and}$$
$$b(t) = 0 \ \ \forall t \in \left\{\left\{\overline{1,T}\right\} \setminus \mathrm{T}_1\right\} \quad (18)$$

If all the lags are ignored, dynamic logistic regression model (16) is further simplified to a static logistic regression model:

$$P[b(t)=1 \mid \mathbf{X}_5(t)] = p_t =$$
$$\left[1+\exp\left(-\alpha_0 - \mathbf{A}_5\cdot[\mathbf{X}_5(t)]^{\mathrm{T}}\right)\right]^{-1}. \quad (19)$$

Static logistic regression model (19) is built by solving problem (9) as

$$\max_{\{\alpha_0, \mathbf{A}_5\}} \left(\sum_{t_1\in \mathrm{T}_1} \left(\alpha_0 + \mathbf{A}_5\cdot[\mathbf{X}_5(t_1)]^{\mathrm{T}}\right) + \right.$$
$$\left. +\sum_{t=1}^{T} \ln\left(\alpha_0 + \mathbf{A}_5\cdot[\mathbf{X}_5(t)]^{\mathrm{T}}\right) - \frac{\lambda}{2}\sum_{i=1}^{5}\alpha_i^2 \right) \quad (20)$$

by (18). To estimate performance of both models (16) and (19), we use the initial 70 % of chronological data for training and last 30 % for testing (temporal split) to emulate forecasting performance. So, the size of the training set is 42. For the case of static model (19) the size of the test set is 18, and for the case of dynamic model (16) the size of the test set is 17 due to the lag. It is worth noting that the number of nonzero risk indicators in the test set is 6 (it is well seen in Fig. 1), which is about one third of the test set size.

We use MATLAB-function “lassoglm” with “binomial” family and the parameter of the pure ridge [15, 16] to perform penalised logistic regression by sweeping λ between 0 and 0.1:

$$\lambda \in \left\{0, \left\{w\cdot 10^{-5}\right\}_{w=1}^{9}, \left\{w\cdot 10^{-4}\right\}_{w=1}^{9}, \left\{w\cdot 10^{-3}\right\}_{w=1}^{100}\right\}. \quad (21)$$

Then a λ^* is selected such that optimises the performance metrics, including:

1) accuracy (fraction of correctly predicted days), which is to be maximised;

2) precision, recall, F1-score for the bankruptcy class, which all are to be maximised;

3) confusion matrix whose integer entries of True Positive (TP) and True Negative (TN) are to be maximised, while False Positive (FP) and False Negative (FN) are to be minimized.

Before running the optimisation problem within MATLAB-function “lassoglm”, all the predictors are standardised by subtracting the mean value over the training set and dividing by the standard deviation over the training set [12, 17, 18]. That is, both the training and test sets are standardised. This allows applying the penalty evenly [15].

We establish the simplest way to predict the bankruptcy risk. If $p_t > 0.5$ then the bankruptcy in the next period is predicted; if $p_t < 0.5$ then no bankruptcy in the next period is predicted. The case $p_t = 0.5$ is not excluded, though; if it occurs then the next period (which is $t + 1$ here) is declared the 0.5-uncertainty-bankruptcy period. In fact, this case is almost as bad as the case $p_t > 0.5$.

Results and interpretation

First we try static model (19) over set (21). The best value of λ is $\lambda^* = 0.07$, where static model (19) is

$$P[b(t)=1 \mid \mathbf{X}_5(t)] = p_t =$$
$$= [1+\exp(0.6111 + 0.3614x_1(t) - 0.8290x_2(t) +$$
$$+0.4264x_3(t) + 0.1576x_4(t) - 0.2507x_5(t))]^{-1} \quad (22)$$

at which the FN number is 2, while there are no FP predictions, Owing to that, the precision is 100 %. The accuracy is 88.89 %, which is usually said that it could be better. Due to the two bankruptcy-risky periods unseen by the static model (Fig. 2), the recall is 66.67 % being unsatisfactory. The F1-score is 80 %, which is not considered acceptable as well.

However, as we look closely at Fig. 2, the weak inaccuracy of static model (22) appears even worse. The matter is that the predicted probability polyline is squeezed and thus many non-risky periods have not really low probabilities, whereas bankruptcy-risky periods have not really high probabilities. In fact, the predicted probability varies between 0.0719 (at $t = 52$) and 0.7993 (at $t = 55$). Moreover, Fig. 2 reveals that there are two periods ($t = 51$ and $t = 54$), where the predicted probability is very close to 0.5 (p_{51} = 0.4995 and p_{54} = 0.4905), that is the 0.5-uncertainty-bankruptcy periods factually ensue.

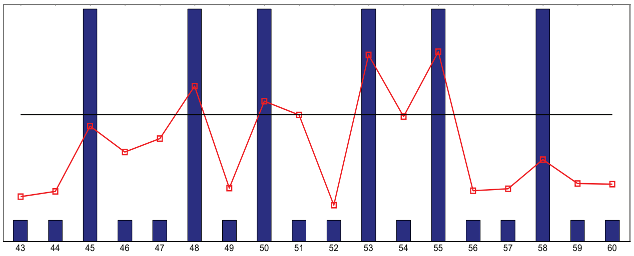

Fig. 2. The test set of 18 periods for static model (22) and the square-marked predicted probability, where the black horizontal line is at 0.5 level (the two bankruptcy-risky periods at t = 45 and t = 58 are left unseen by the model)

Besides, there is another pretty strange moment in static model (22): this is the positivity of the solvency coefficient α_5 = 0.2507, which means that the increasing solvency must raise the probability of bankruptcy. By its absolute value this coefficient is far less than the leverage coefficient α_2 = 0.8290 (whose positivity is quite natural and understandable), but still it is more influential than the efficiency coefficient α_4 = −0.1576. This and the other inconsistencies mentioned above prompt to try including lagged predictors into consideration, as their influence may rectify predictability and improve performance of regression models.

Hence, subsequently, we try dynamic model (16) over set (21). The best value of λ is λ^* = 0.007, where dynamic model (16) is

$$P[b(t)=1 \mid \mathbf{Y}_{10}(t); \mathbf{B}_{-1}(t)] = p_t =$$
$$=[1+\exp(1.8066+0.0868x_1(t)-2.6287x_2(t)+$$
$$+1.0353x_3(t)+0.2122x_4(t)+0.1474x_5(t)+$$
$$+0.0762x_1(t-1)+0.7351x_2(t-1)-0.6606x_3(t-1)-$$
$$-0.7530x_4(t-1)+0.1537x_5(t-1)+$$
$$+2.5959b(t-1))]^{-1}. \qquad (23)$$

Dynamic model (23) performs over the test set of 17 periods far better than static model (22). Indeed, including the lags helped not only increase the accuracy up to 94.44 % and F1-score to 92.31 %, but also to stretch predicted probability polyline (Fig. 3), whereas there are no FN predictions and only one FP prediction (at t = 47). Now the predicted probability varies between 0 and 0.9971, and there are no 0.5-uncertainty-bankruptcy periods (despite p_{58} = 0.5183, it is farther from the real-practice uncertainty, and this risky period has been correctly predicted). The performance metrics of static model (22) and dynamic model (23) are presented in Table 3 for comparability.

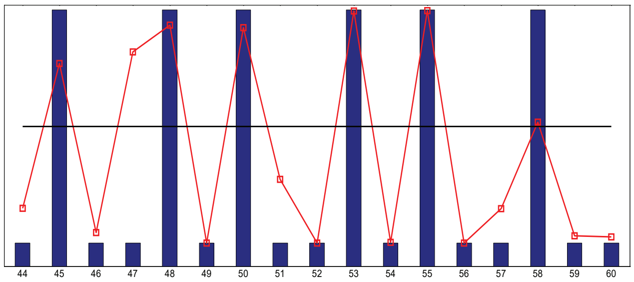

Fig. 3. The test set of 17 periods for dynamic model (23) and the square-marked predicted probability, where the black horizontal line is at 0.5 level

Table 3. Performance metrics of the static and dynamic models for the synthetic 60-period time series in Fig. 1

Performance metric	Static model (22)	Dynamic model (23)
Accuracy	0.888 889	0.944 444
Precision	1	0.857 143
Recall	0.666 667	1
F1-score	0.8	0.923 077
TP	4	6
TN	12	11
FP	0	1
FN	2	0

Although dynamic model (23) drops the precision down to 85.71 %, it is caused by the single FP prediction (which, obviously, is a way better than having an FN prediction). The remaining performance metrics indicate the clear advantage of the lagged regression. Another optimistic peculiarity of the one-lagged regression model is that its coefficients at the five non-lagged predictors all have interpretable signs — negative ones at liquidity, profitability, efficiency, and solvency, while the leverage coefficient sign is positive. The lagged predictors do not completely follow this pattern as the lagged leverage coefficient is negative, while the lagged profitability and efficiency coefficients are positive. The lagged response has a negative coefficient (μ_1 = −2.5959) as well. However, this is commonly normal due to the following reasons:

1. Lagged predictors may flip signs because the current ratios already explain most of the variation in bankruptcy, so the lagged ones only capture left-

over corrections, often producing opposite-direction effects due to multicollinearity.

2. A lagged variable enters the model after its current version, so its coefficient reflects changes rather than levels; this naturally yields inverted signs.

3. Lagged response gets a negative sign because, in our synthetic setup, distress is reversible: if the last period was distressed but today's ratios look healthy, the model learns a mean-reversion effect.

4. When distress does not persist automatically but depends on today's fundamentals, the lagged response serves as a "temporary shock indicator", leading to a negative coefficient.

Hence, the regularised dynamic logistic regression model (16) by (17) proves to be interpretable and quantitatively consistent for bankruptcy risk prediction of a building-materials manufacturer in Ukraine. Despite not very large training set (42 periods are used to estimate 12 coefficients of the model), the lagged regression successfully captures bankruptcy-risky periods during 17 test set points for synthetic datasets generated by (11)–(15) with descriptive statistics similar to that in Table 2. Nevertheless, the question is whether the lagged regression model is enough robust and sustainable. To answer this question, the synthetic dataset should be extended and the respective model should be tested on a much longer test span.

So, using the similar if-then rules built from the ratios in Table 1, by which the bankruptcy binary risk is generated for 60 periods (Fig. 1), we generate

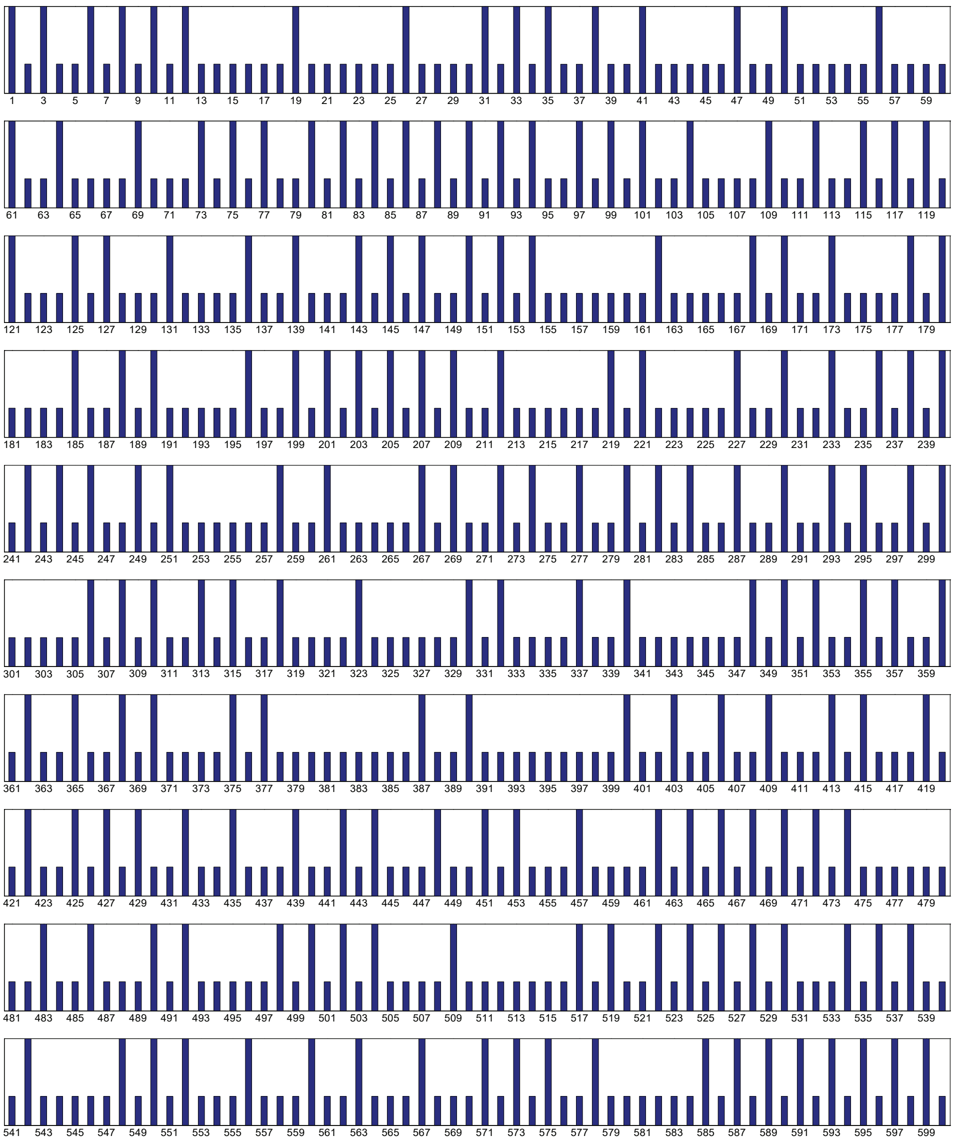

Fig. 4. Binary risk of bankruptcy throughout 600 periods, divided into 10 equal spans

a synthetic 600-period time series of the bankruptcy binary risk (Fig. 4). Herein, the total number of events $b(t) = 1$ is 188, which is 31.33 %, and so we keep roughly a one third of risky periods. The size of the training set is 420. For the case of static model (19) the size of the test set is 180, and for the case of dynamic model (16) the size of the test set is 179 due to the lag. It is worth noting that the number of nonzero risk indicators in the test set is 59 (it is well seen and can be calculated in the three bottom spans in Fig. 4), which is about one third of the test set size. The 59 risky periods are almost equally scattered throughout the test span (20, 19, and 20 risky periods in the three bottom spans in Fig. 4).

When we apply static model (19) for the 600-period time series, the regularisation hyperparameter is set to $\lambda^* = 0.07$ for keeping congruence with static model (22). In this way, our static model (19) becomes one with interpretably signed coefficients (negative coefficients at liquidity, profitability, efficiency, and solvency, while the leverage coefficient sign is positive):

$$P[b(t) = 1|\mathbf{X}_5(t)] = p_t =$$
$$= [1+\exp(1.0236+0.0308x_1(t)-1.0503x_2(t)+$$
$$+0.2479x_3(t)+0.1247x_4(t)+0.2086x_5(t))]^{-1}. \quad (24)$$

However, static model (24) performs poorly on the 180-period test span: its accuracy is 72.22 %, the precision is 60.47 %, while the recall and F1-score drop down to 44.07 % and 50.98 %, respectively. The number of FN predictions is 33, which is huge (55.93 %) with respect to the 59 bankruptcy-risky periods in the test span. The number of FP is 17, so altogether the 50 false predictions constitute 27.78 % of the test set, which is quite unacceptable. In addition, just like static model (19) for the 60-period time series, the predicted probability polyline is squeezed and again many non-risky periods have not really low probabilities, whereas bankruptcy-risky periods have not really high probabilities (Fig. 5). In fact, the predicted probability varies between 0.0433 (at $t = 437$) and 0.7696 (at $t = 502$). Fig. 5 also reveals that there are four periods ($t = 442$, $t = 522$, $t = 527$, $t = 578$), where the predicted probability is no farther from 0.5 than by 0.004, that is the 0.5-uncertainty-bankruptcy periods factually ensue.

For keeping congruence with dynamic model (23), the regularisation hyperparameter is set to $\lambda^* = 0.007$ in applying dynamic model (16) for the 600-period time series. Then the respective dynamic model (16) also has interpretably signed coefficients at the non-lagged predictors (negative coefficients at liquidity, profitability, efficiency, and solvency, while the leverage coefficient sign is positive), although the lagged predictors and response have coefficients that all are sign-inverted:

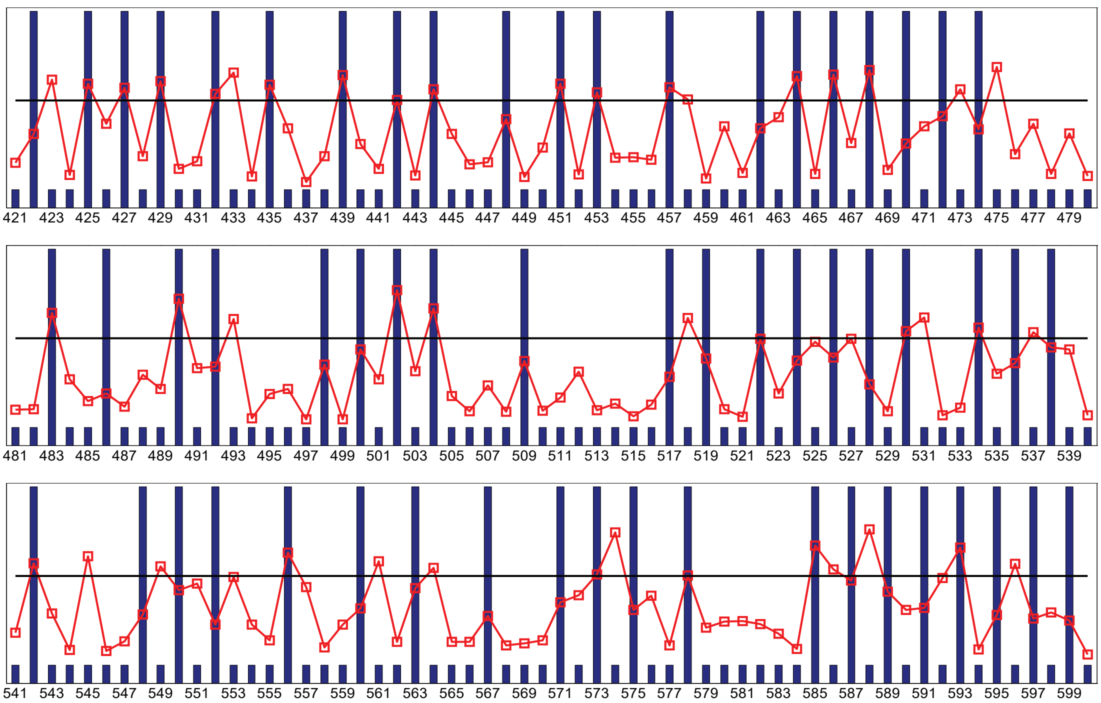

Fig. 5. The three spans of the test set of 180 periods for static model (24) and the square-marked predicted probability

$$P[b(t)=1\,|\,\mathbf{Y}_{10}(t);\mathbf{B}_{-1}(t)]=p_t=[1+\exp(2.2584+$$
$$+0.1765x_1(t)-2.4866x_2(t)+0.3840x_3(t)+$$
$$+0.3259x_4(t)+0.4966x_5(t)-0.0521x_1(t-1)+$$
$$+0.2099x_2(t-1)-0.2184x_3(t-1)-0.1234x_4(t-1)-$$
$$-0.1195x_5(t-1)+2.3423b(t-1))]^{-1}. \qquad (25)$$

Just like in the case of the static models, whose respective six coefficients differ significantly by their absolute values, the 12 coefficients of lagged regression (25) significantly differ from those of lagged regression (23). Dynamic model (25) nonetheless performs far better than static model (24): its accuracy is 91.11 %, the precision is 92.16 %, although the recall and F1-score are not that good (Table 4). The number of FN predictions is 12, which is still pretty poor (20.34 %) with respect to the 59 bankruptcy-risky periods in the test span. The number of FP is 4, so altogether the 16 false predictions constitute 8.89 % of the test set, which could be acceptable, though. All the more, the predicted probability polyline is not squeezed (Fig. 6), where the predicted probability varies between 0.0000235 (at $t = 499$) and 0.9896 (at $t = 502$, where the static model performed its best prediction as well). The predicted probability is no closer to 0.5 than by 0.0017, but there is one period ($t = 548$), where it is no farther from 0.5 than by 0.004, ensuing factually a 0.5-uncertainty-bankruptcy period. If to widen the 0.5-uncertainty interval to 0.01, another risky-uncertain period ($t = 567$) emerges. By the way, the model labels both the periods non-risky, making its contribution to the FN number.

Although the lagged regression performs not poorly over the tenfold data, especially compared to the static regression (Table 4), the FN number is still badly high. This implies that the logistic model cannot perform satisfactorily on too large datasets like that one in Fig. 4. Therefore, the model should be re-estimated through shorter or not very long time spans like that one in Fig. 1.

According to dynamic model (23), where L2 regularisation helps manage multicollinearity between $\mathbf{X}(t)$ and $\mathbf{X}(t-1)$, the lagged financial indicators help anticipate bankruptcy earlier, particularly for firms showing deteriorating performance over several periods. The dynamic logistic regression model, which incorporates both contemporaneous and one-period-lagged financial indicators, reveals several important insights into bankruptcy risk. As expected, the current-period predictors display intuitive signs: higher liquidity, profitability, efficiency, and solvency reduce the probability of distress, while higher leverage increases it. These results are fully consistent with standard financial theory and

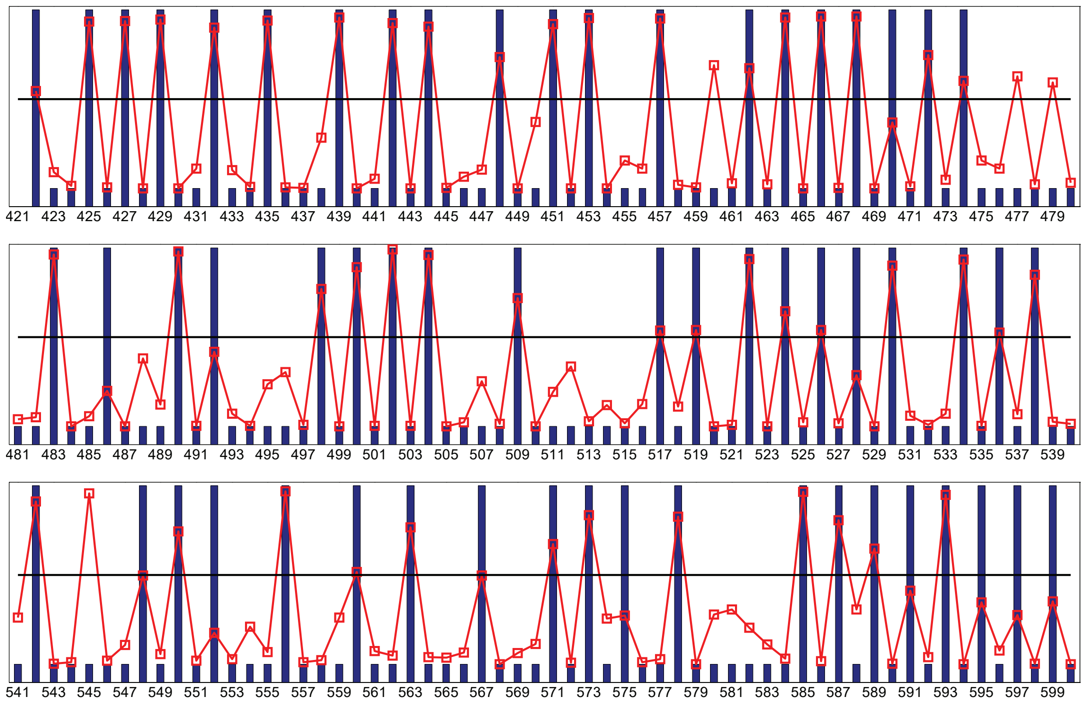

Fig. 6. The three spans of the test set of 179 periods for dynamic model (25) and the square-marked predicted probability

Table 4. Performance metrics of the static and dynamic models for the synthetic 600-period time series in Fig. 4

Performance metric	Static model (24)	Dynamic model (25)
Accuracy	0.722 222	0.911 111
Precision	0.604 651	0.921 569
Recall	0.440 678	0.79 661
F1-score	0.509 804	0.854 545
TP	26	47
TN	104	117
FP	17	4
FN	33	12

empirical evidence. The more interesting behaviour emerges in the lagged predictors, where we unexpectedly obtain negative leverage along with positive profitability and efficiency. Nevertheless, these inverted or non-intuitive signs do not indicate estimation errors; instead, they reflect the structure of the synthetic data and the dynamic relationships embedded in the system.

First, the lagged leverage coefficient becomes negative because the current leverage already absorbs nearly all the explanatory power related to financial pressure. The lagged value therefore captures changes rather than levels: a firm whose leverage was very high in the previous period but has decreased today is less likely to be distressed. In this sense, the lagged coefficient reflects a deleveraging recovery effect.

Similarly, the positive coefficients on lagged profitability and efficiency arise because the model interprets them relative to the current values. When profitability or efficiency drop sharply from one period to the next, bankruptcy risk increases – so a high previous-period value combined with a low current value signals deterioration. The lagged coefficients thus take positive signs to encode this downward momentum effect.

Finally, the negative coefficient on the lagged response confirms that distress in this synthetic setup is not a persistent absorbing state. Firms marked as distressed in one period can recover in the next if their financial ratios improve. Therefore, when the model sees $b(t-1)=1$ but today's fundamentals look healthy, it interprets past distress as a temporary shock that is likely to reverse, producing a negative effect.

Altogether, these patterns confirm that the dynamic model does not merely reproduce static financial relationships but captures binary transitions between sustainable and critical states, momentum effects, and reversions, all driven by the interplay between current and lagged financial indicators [2, 6, 19]. Despite some signs appearing counterintuitive when viewed in isolation, the combined structure reflects precisely the data-generation mechanism and yields strong predictive accuracy, especially for out-of-sample periods.

Conclusions and implications

This study demonstrates that logistic regression modelling can be successfully applied to the prediction of potential bankruptcy of industrial firms. Using a synthetic dataset representing monthly financial indicators of a building-materials manufacturer in Ukraine, both static and dynamic versions of the logistic regression model were estimated with L2 regularisation to ensure parameter stability and full inclusion of correlated predictors.

The static logistic model, based on current-period financial ratios, provided a baseline prediction accuracy of approximately 89 %, correctly identifying most solvent and insolvent states of the firm, but losing two bankruptcy-risky months. The dynamic model, incorporating one-period lagged financial ratios, achieved around 94 % accuracy, showing a consistent improvement across recall and F1-score metrics, while no one out of six bankruptcy-risky months was lost. This gain reflects the dynamic model's ability to capture temporal dependencies in firm financial health and to anticipate deterioration before formal bankruptcy occurs.

From a methodological standpoint, the results confirm that:

1. L2-regularised logistic regression remains a reliable and interpretable tool for financial distress modeling even with small or correlated datasets.

2. Introducing lagged predictors effectively transforms the static framework into a dynamic, autoregressive-like model that captures financial inertia – the persistence of past conditions affecting present solvency. In the dynamic model, current ratios dominate and lagged variables correct, whereas these corrections often appear with opposite signs.

3. Logistic regression's probabilistic interpretation makes it particularly well-suited for risk-based early-warning systems, where firms can be monitored by thresholding predicted bankruptcy probabilities.

In the Ukrainian building and construction materials industry, where firms often operate under volatile demand, credit constraints, and high capital intensity, predicting financial distress is vital. The model developed here provides a quantitative framework for:

1. Early detection of bankruptcy risk several quarters in advance, enabling proactive interventions.

2. Scenario testing, such as evaluating the impact of rising debt or declining profitability.

3. Integrating predictive analytics into credit scoring systems used by banks, suppliers, and regulators.

The synthetic firm analysed in this paper mimics the structure and financial behaviour of a typical mid-sized Ukrainian manufacturer, but the model can easily be recalibrated using real company data, once available. Although the presented model provides valuable insight, it also has three important limitations. First, the model relies on a synthetic dataset rather than empirical data. Second, it assumes constant coefficients over time, while real-world relationships may evolve. Third, only one lag was used in the dynamic model, but longer financial memory might improve forecasting accuracy. Future research may address these limitations by using panel data across multiple firms and estimating firm-specific random effects, by applying time-varying or adaptive regularisation to allow coefficients to change with macroeconomic conditions [20, 21], and by exploring nonlinear extensions such as kernel logistic regression or neural-network-based survival models for richer temporal dependencies.

The presented work confirms that dynamic logistic regression with regularisation offers a theoretically sound and computationally tractable tool for predicting bankruptcy risk in industrial firms. Its interpretability, modest data requirements, and stable performance make it particularly attractive for early-warning analytics in transitional economies like Ukraine, where financial data are often limited but timely risk identification is crucial.

References

[1] P. du Jardin, “Bankruptcy prediction using terminal failure processes”, *European Journal of Operational Research*, 2015, Vol. 242, Iss. 1, pp. 286–303. Available: https://doi.org/10.1016/j.ejor.2014.09.059

[2] D.S. Nugroho and T. Dewayanto, “Application of statistics and artificial intelligence for corporate financial distress prediction models: a systematic literature review”, *Journal of Modelling in Management*, 2025, Vol. 20, Iss. 6, pp. 1999–2023. Available: https://doi.org/10.1108/JM2-12-2024-0412

[3] F. Fasano et al., “The dilemma of accuracy in bankruptcy prediction: a new approach using explainable AI techniques to predict corporate crises”, *European Journal of Innovation Management*, 2025, Vol. 28, Iss. 11, pp. 1–22. Available: https://doi.org/10.1108/EJIM-06-2024-0633

[4] E. Lyandres and A. Zhdanov, “Investment opportunities and bankruptcy prediction”, *Journal of Financial Markets*, 2013, Vol. 16, Iss. 3, pp. 439–476. Available: https://doi.org/10.1016/j.finmar.2012.10.003

[5] E.I. Altman, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”, *Journal of Finance*, 1968, Vol. 23, No. 4, pp. 589–609. Available: https://doi.org/10.2307/2978933

[6] J.A. Ohlson, “Financial Ratios and the Probabilistic Prediction of Bankruptcy”, *Journal of Accounting Research*, 1980, Vol. 18, No. 1, pp. 109–131. Available: https://doi.org/10.2307/2490395

[7] M. Hesse and T. Loy, “Unlocking bankruptcy clues: A novel sentence-based machine learning approach”, *International Journal of Accounting Information Systems*, 2025, Vol. 56, Art. no. 100751. Available: https://doi.org/10.1016/j.accinf.2025.100751

[8] X. Chen *et al.*, “Multi-class financial distress prediction based on hybrid feature selection and improved stacking ensemble model”, *Expert Systems with Applications*, 2025, Vol. 282, Art. no. 127832. Available: https://doi.org/10.1016/j.eswa.2025.127832

[9] D.W. Hosmer Jr. *et al.*, Applied Logistic Regression, 3rd Edition, Wiley, 2013. Available: https://doi.org/10.1002/9781118548387

[10] D.G. Kleinbaum and M. Klein, Logistic Regression. A Self-Learning Text, Springer New York, NY, 2010. Available: https://doi.org/10.1007/978-1-4419-1742-3

[11] A. Magrini, “Bankruptcy risk prediction: A new approach based on compositional analysis of financial statements”, Big Data Research, 2025, Vol. 41, Art. no. 100537. Available: https://doi.org/10.1016/j.bdr.2025.100537

[12] T. Hastie *et al.*, The Elements of Statistical Learning, Springer Series in Statistics, Springer, New York, NY, 2009. Available: https://doi.org/10.1007/978-0-387-84858-7

[13] C.M. Bishop, Pattern Recognition and Machine Learning, Springer New York, NY, 2006. Available: https://link.springer.com/book/9780387310732

[14] K.P. Murphy, Machine Learning: A Probabilistic Perspective, *The MIT Press*, 2012. Available: https://mitpress.mit.edu/9780262018029/machine-learning/

[15] J.H. Friedman *et al.*, “Regularization paths for generalized linear models via coordinate descent”, *Journal of Statistical Software*, 2010, Vol. 33, Iss. 1, pp. 1–22. Available: https://doi.org/10.18637/jss.v033.i01

[16] J.K. Tay *et al.*, "Elastic net regularization paths for all generalized linear models", *Journal of Statistical Software*, 2023, Vol. 106, Iss. 1, pp. 1–31. Available: https://doi.org/10.18637/jss.v106.i01

[17] M. Kuhn and K. Johnson, Applied Predictive Modeling, Springer New York, NY, 2013. Available: https://doi.org/10.1007/978-1-4614-6849-3

[18] G. James *et al.*, An Introduction to Statistical Learning with Applications in Python, Springer Texts in Statistics, Springer Cham. Available: https://doi.org/10.1007/978-3-031-38747-0

[19] J.-P. Danthine *et al.*, "Chapter 22 – Financial Equilibrium with Differential Information", In: J.-P. Danthine, J.B. Donaldson, and S. Danthine (Eds.), Intermediate Financial Theory (Fourth Edition), Academic Press, 2025, pp. 641–657. Available: https://doi.org/10.1016/B978-0-443-28902-6.00022-1

[20] S.V. Merinova and V.V. Romanuke, "Perspectives of blockchain technology in business and management: advantages and challenges", *Systems and Technologies*, 2025, Vol. 69, No. 1, pp. 138–144. Available: https://doi.org/10.32782/2521-6643-2025-1-69.17

[21] A. Sutrisno and R.A. Hamka, "Development and stabilization in small open economies", *Journal of Enterprising Communities: People and Places in the Global Economy*, 2025, Vol. 19, Iss. 3, pp. 700–701. Available: https://doi.org/10.1108/JEC-06-2025-272

В.В. Романюк

МОДЕЛІ ДИНАМІЧНОЇ ЛОГІСТИЧНОЇ РЕГРЕСІЇ ДЛЯ ПРОГНОЗУВАННЯ РИЗИКУ БАНКРУТСТВА У БУДІВЕЛЬНІЙ ГАЛУЗІ УКРАЇНИ

Проблематика. Прогнозування фінансової неспроможності та банкрутства набуває дедалі більшої актуальності в умовах післявоєнного економічного відновлення та реструктуризації українських промислових галузей. Підприємства сектору виробництва будівельних матеріалів працюють в умовах високої невизначеності, де раннє виявлення ризику неплатоспроможності є критично важливим для підтримання фінансової стабільності. Логістичні регресійні моделі, широко застосовувані в екологічній та ризик-аналітиці, можуть бути адаптовані для відображення нелінійного переходу від платоспроможності до банкрутства як ймовірнісного процесу.

Мета дослідження. Метою є розробити та оцінити статичні й динамічні логістичні регресійні моделі для прогнозування потенційного банкрутства репрезентативного українського виробника будівельних матеріалів. Динамічне розширення моделі спрямоване на врахування часової інерційності фінансових показників шляхом включення лагових предикторів.

Методика реалізації. Згенеровано синтетичний помісячний набір даних (5 років, 60 спостережень), що імітує реалістичні фінансові коефіцієнти, зокрема ліквідність, левередж, рентабельність, ефективність та коефіцієнт покриття відсотків (платоспроможність). Оцінювання моделей виконано в MATLAB методом логістичної регресії максимальної правдоподібності з L2-регуляризацією (ридж-штрафом) для збереження корельованих предикторів. До динамічної моделі включено одноперіодні лаги всіх фінансових коефіцієнтів і одноперіодний лаг реакції. Прогнозну якість оцінено за точністю, прецизійністю, повнотою, F1-мірою та матрицею сплутувань.

Результати дослідження. Статична логістична модель досягла середньої точності приблизно 89 %, але пропустила два ризикові щодо банкрутства місяці з шести. Динамічна модель підвищила точність до 94 %, не пропустивши жодного ризикового місяця, хоча й помилково класифікувала один неризиковий місяць як ризиковий. Знаки оцінених коефіцієнтів узгоджуються з економічною логікою: більший левередж підвищує ймовірність банкрутства, тоді як зростання ліквідності, рентабельності, ефективності та платоспроможності її знижує.

Висновки. Динамічна логістична регресія з L2-регуляризацією забезпечує інтерпретовану й обчислювально ефективну основу для раннього прогнозування банкрутства українських промислових підприємств. Включення лагових фінансових індикаторів підвищує стабільність і своєчасність прогнозів, що робить модель придатною для практичних систем раннього попередження.

Ключові слова: прогнозування банкрутства; логістична регресія; динамічне моделювання; фінансові коефіцієнти; L2-регуляризація; система раннього попередження; будівельний сектор України.

Рекомендована Радою
факультету прикладної математики
КПІ ім. Ігоря Сікорського

Надійшла до редакції
25 листопада 2025 року

Прийнята до публікації
09 березня 2026 року

Опублікована
30 березня 2026 року