Forecasting Capital Structure Using Time Series Theory

E. M. Bogatov^{1*}, E. G. Demidova²

 PhD in Math., Associate Professor, Gubkin Branch of NITU «MISIS», Gubkin, Russia; Associate Professor, STI NITU MISIS, Stary Oskol, Russia
 PhD in Econ., Associate Professor, STI NITU MISIS, Stary Oskol, Russia,

Abstract: This study aims to develop a mechanism for extrapolating the values of determinants (Return on Equity (ROE), Return on Sales (ROS), Turnover Ratio (Y), Interest Burden (Ib), and Tax Burden (Tb)) for subsequent integration into an author-developed lag model of capital structure. This approach enables the forecasting of corporate capital structure in an impersonal manner, eliminating reliance on expert assessments of future values for ROE, ROS, Y, Ib, and Tb. The identification of the ROE observation series using the Box-Jenkins methodology within the ARIMA model framework as a first-order autoregressive model AR(1) allows for the interpretation of the corollary from the formula for increments of multivariate functions as an autoregressive model concerning the equity multiplier M. Consequently, this facilitates the calculation of the future period value M_1 based on the actual value M_0 from the preceding period and the relationship between known and predicted values of the lag model's determinants (ROE, ROS, Y, Ib, Tb). The proposed algorithm for calculating capital structure is tested using Magnitogorsk Iron and Steel Works (MMK) as a representative agent. Time series identification based on a 25-year dataset (2000–2024) yielded the following results: Return on Equity and Return on Sales are described by first-order autoregressive models; the Interest and Tax Burden coefficients correspond to stationary «white noise» processes; and the behavior of the Turnover Ratio is described by a Seasonal Autoregressive Integrated Moving Average (SARIMA) process. The justification for the applicability of lag capital structure models and the definition of their determinant values for future periods based on time series identification is presented here for the first time.

Keywords: capital structure, forecasting, time series, lag models, ARIMA.

1. Introduction

Forecasting financial and economic indicators is a key challenge in contemporary economics and corporate management. The accuracy of such forecasts directly impacts the quality of strategic decisions, financial planning, and business sustainability. Among the most critical indicators are capital structure and the equity multiplier, which influence a company's financial leverage and risk profile. Existing methods for forecasting economic indicators can be broadly categorized into three groups: classical statistical models, time series models, and modern machine learning techniques.

Classical statistical methods, including regression analysis and models based on the DuPont formula, are characterized by ease of implementation and high interpretability. However, they are limited by assumptions of data stationarity and low adaptability to dynamic changes in the external environment, which reduces forecast accuracy under conditions of uncertainty [4, 7, 10].

Time series models, particularly Autoregressive Integrated Moving Average (ARIMA) models, are widely used for analyzing and forecasting the dynamics of economic indicators. The Box-Jenkins methodology facilitates the identification of seasonal fluctuations, trends, and random variations, thereby enhancing forecast quality [6]. However, ARIMA models require strict verification of stationarity conditions and regular model updates in response to changes in the economic environment [12, 5].

Modern machine learning methods and artificial neural networks offer flexible tools for modeling complex nonlinear relationships and processing large datasets. These methods can adapt to changing patterns but often suffer from limited interpretability and require significant computational resources, complicating their application in practical financial management [9, 11].

Despite significant progress in predictive methodologies, the use of models with strong economic interpretability and transparency remains crucial for financial management tasks. Lag models, based on the classical DuPont formula and its extensions, continue to play a vital role by combining simplicity with the potential for integration with extrapolation methods. Subjective expert assessments often lead to systematic errors, highlighting the need for alternative approaches to forecasting economic indicators. A key aspect of improving forecast quality is the correct identification of time series, which is a necessary condition for forecasting corporate capital structure using the proposed lag model. This is important for two reasons. On one hand, the Box-Jenkins methodology ensures the objectivity of input data through a formalized forecast based on identified statistical patterns. On the other hand, ignoring the mathematical relationship between determinants via the increment formula (a variant of the autoregressive model) could lead to incorrect forecasts of their values, causing a chain error in the calculation of the multiplier M₁ and, consequently, the capital structure. Furthermore, using ARIMA for identifying a series spanning 25 observations helps smooth out the impact of one-time distortions and yields robust, interpretable models for the behavior of each determinant.

2. Research objective and methodology

This work employs a five-factor lag model of capital structure, where the equity multiplier for the next period is expressed through key economic indicators (ROS, Y, Ib, Tb, and ROE) from previous and subsequent periods. To enhance predictive stability and the objectivity of factor assessment, the Box-Jenkins methodology is applied to identify and forecast time series of observations for the capital structure determinants.

This integrated approach eliminates reliance on subjective assessments, yielding more accurate and economically substantiated forecasts of capital structure. This is particularly important for enterprises with high volatility in financial indicators, such as those in the metallurgical industry.

3. Methods

In recent works by the authors [2, 3], capital structure models were examined, enabling the determination of the equity multiplier for the subsequent period (M_{i+1}) based on its value in the preceding period (M_i) combined with information on the values of determinants (ROS, Y, Ib, Tb, ROE) from both the preceding and subsequent periods. The foundation is the formula for the increment of a multivariate function $F = F(x_1, ..., x_5)$:

$$\Delta F = F(x_1 + \Delta x_1, ..., x_5 + \Delta x_5) - F(x_1, ..., x_5),$$

where x_1 represents the Return on Sales (ROS) coefficient, x_2 the Turnover Ratio (Y), x_3 the Multiplier (M), and x_4 and x_5 the Interest Burden (Ib) and Tax Burden (Tb) coefficients, respectively. The dependent variable is Return on Equity (ROE).

Since for the five-factor DuPont model the equality

$$F = x_1 x_2 x_3 x_4 x_5, (1),$$

must hold, the increment formula provides a specific rule for calculating subsequent F values:

$$F_{i+1} = F_i + (x_1^i + \Delta x_1) \cdot \dots \cdot (x_5^i + \Delta x_5) - x_1^i \cdot \dots \cdot x_5^i, \tag{2}$$

where the subscript i(i+1) denotes the preceding (subsequent) values of the variables or function.

Equation (2) can be rewritten as:

$$F_{i+1} = F_i + \alpha(x_j^i) - \beta(x_j^{i+1}),$$
 $j = 1, ..., 5$ (3)

where
$$x_j^{i+1} = x_j^i + \Delta x_j$$
, $\alpha(x_j^i) = \prod_{j=1}^5 x_j^{i+1}$; $\beta(x_j^{i+1}) = \prod_{j=1}^5 x_j^i$.

Assuming the existence of time series of observations for $F(F_t)$ and for $x_i(x_i^t)$, formula (3) can be expressed as:

$$F_{t+1} = F_t + (\alpha_t - \beta_{t+1}), \tag{4}$$

 $F_{t+1} = F_t + (\alpha_t - \beta_{t+1}), \tag{4}$ In the simplest case, where $(\alpha_t - \beta_{t+1})$ is white noise, model (4) represents a first-order autoregressive model.

Let the last observation period be denoted as τ_0 , and the subsequent period as τ_1 . Corresponding observed values of determinants x_i and function F will be superscripted with 0 and 1, respectively. Then, from formula (3) for the last period, we obtain:

$$F^{1} - F^{0} = \prod_{j=1}^{5} x_{j}^{j} - \prod_{j=1}^{5} x_{j}^{0} . \tag{5}$$

Returning to the economic meaning of variables x_i and function F, and expressing the multiplier M value for period τ_1 from formula (5), we get:

$$M_1 = \frac{\Delta ROE}{ROS_1 Y_1 T_b^1 I_b^1} + M_0 \frac{ROS_0 Y_0 T_b^0 I_b^0}{ROS_1 Y_1 T_b^1 I_b^1}.$$
 (6)

where $\Delta ROE = ROE_1 - ROE_0$;

 M_1 = new, target equity multiplier = Assets/Equity;

 M_0 = equity multiplier of the base period;

ROS₀ and ROS₁ = Return on Sales coefficients in base and analysis periods = (Profit Before Tax + Interest Expense) / Revenue;

 Y_0 and Y_1 = Asset Turnover ratios in base and analysis periods = Revenue / Assets;

 Ib_0 and Ib_1 = Interest Burden coefficients in base and analysis periods = Profit Before Tax / (Profit Before Tax + Interest Expense) [3].

Equation (6) corresponds to a five-factor lag DuPont model concerning multiplier M. Its correct application requires identifying the time series corresponding to the ROE indicator and variables x_i and forecasting their values for period τ_1 using the Box-Jenkins methodology [6].

The essence of the developed extrapolation mechanism lies in combining two methodologies:

- forecasting determinant values (ROS, Y, Ib, Tb) for period t+1 using classical time series models (ARIMA/SARIMA) based on their historical values calculated according to Russian Accounting Standards (RAS).
- utilizing forecasted values in the author's five-factor lag model (formula (6)), which links the change in ROE with changes in determinants and allows for calculating the required equity multiplier value M₁ for the future period. The core of the identification and extrapolation mechanism for the aforementioned determinants precisely involves applying this two-step approach to forecasting the capital structure of Russian and foreign companies and demonstrating its efficacy.

4. Results

Using Magnitogorsk Iron and Steel Works ("MMK") as a representative agent, time series for its indicators ROS, Y, M, Ib, Tb, ROE were constructed based on annual financial statements according MMK from 2000 to 2024 (see Table 1).

To analyze ROE behavior, a graph was plotted using MS Excel (see Fig. 1).

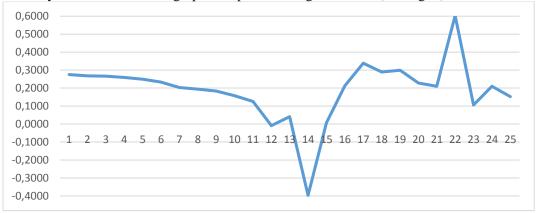


Figure 1. Dynamics of ROE changes for "MMK" over 25 years

Table 1. Annual data on "MMK" indicators for 25 years

Year	ROE	Y	ROS	Ib	Tb	M
2 000	0,2752	1,1112	0,1805	0,9091	0,7500	2,0123
2 001	0,2683	1,0937	0,1803	0,9091	0,7500	1,9948
2 002	0,2659	1,1428	0,1805	0,9091	0,7500	1,8904
2 003	0,2591	1,1650	0,1810	0,9091	0,7501	1,8023
2 004	0,2496	1,1648	0,1806	0,9091	0,7500	1,7407
2 005	0,2338	1,1451	0,1740	0,9091	0,7500	1,7207
2 006	0,2035	1,1376	0,1637	0,9091	0,7500	1,6029
2 007	0,1941	1,1406	0,1647	0,9094	0,7528	1,5093
2 008	0,1832	1,1825	0,1515	0,8878	0,7881	1,4613
2 009	0,1576	0,5642	0,2588	0,9560	0,8066	1,3994
2 010	0,1259	0,6647	0,1582	0,9292	0,8236	1,5648
2 011	-0,0090	0,7432	-0,0066	3,1375	0,3280	1,7653
2 012	0,0410	0,7396	0,0650	0,7212	0,6952	1,6996
2 013	-0,3968	0,8404	-0,2352	1,0736	0,9683	1,9313
2 014	0,0072	0,9127	0,0114	0,2284	1,2907	2,3629
2 015	0,2125	1,0199	0,1390	0,8811	0,7976	2,1329
2 016	0,3384	1,1488	0,2554	0,9603	0,8173	1,4696
2 017	0,2898	1,1172	0,2179	0,9845	0,7987	1,5141
2 018	0,2988	1,1840	0,2130	0,9897	0,7628	1,5695
2 019	0,2279	1,1108	0,1616	0,9780	0,8082	1,6061
2 020	0,2102	0,9349	0,1657	0,9728	0,7984	1,7472
2 021	0,6034	1,3598	0,3454	0,9931	0,8196	1,5784
2 022	0,1061	0,8861	0,1083	0,9539	0,8088	1,4322
2 023	0,2103	0,8141	0,2253	0,9745	0,7828	1,5029
2 024	0,1523	0,8149	0,1885	0,9722	0,7592	1,3434

The ROE observation series was identified using the Mathematica package [13] (see Fig. 2).

```
data1 = {0.2752, 0.2683, 0.2659, 0.2591, 0.2496, 0.2338, 0.2035, 0.1941, 0.1832, 0.1576,
   0.1259, -0.009, 0.041, -0.3968, 0.0072, 0.2125, 0.3384, 0.2898, 0.2988, 0.2279, 0.2102,
   0.6034, 0.1061, 0.2103, 0.1523};
tsm = TimeSeriesModelFit[data1]
Normal[tsm]
ARProcess[0.116939, {0.379079}, 0.0239841]
```

Figure 2. Identification of the ROE observation series in Mathematica

Despite sharp fluctuations in the graph, the process is confirmed to be an AR(1) process, i.e., described by a first-order autoregressive model:

$$F_{t+1} = 0.117 + 0.34F_t + \varepsilon_t$$

 $F_{t+1}=0.117+0.34F_t+\varepsilon_t,$ where 0.117 is the process constant, ε_t is white noise. The residual variance is approximately 0.024.

If the identification of this process within the ARMA model family had been different from AR(1), the extension of the increment formula to a non-stationary situation (and thus the application of formula (6)) would have been questionable.

As mentioned, determining the subsequent value of indicator M requires forecasting the values of determinants ROS, Y, Ib, Tb for the next period. Graphs of their changes were plotted based on Table 1 data, and corresponding processes were identified and subsequently predicted (Fig. 3).



Figure 3. Dynamics of ROS changes for "MMK" over 25 years

Applying the Mathematica package to analyze the ROS change process yields the following result (see Fig. 4).

Figure 4. Identification and forecasting of the ROS observation series in Mathematica

Similar to the ROE, the process is described by a first-order autoregressive model: $x_1^{t+1} = 0.093 + 0.393x_1^t + \varepsilon_1^t$,

where 0.093 is the process constant, ε_1^t is white noise. The residual variance is approximately 0.01. The forecasted ROS value for 2025, as seen, is approximately 0.167.

Proceeding to the analysis of the next indicator's dynamics – Y (see Fig. 5).



Figure 5. Dynamics of Y changes for "MMK" over 25 years

Applying the Mathematica package to analyze the Y change process yields a different result from the previous two determinants (see Fig. 6).

Figure 6. Identification and forecasting of the Y observation series in Mathematica

The identification result (SARIMA) indicates that the model uses single non-seasonal differencing ($\nabla y_t = y_t - y_{t-1}$), seasonal autoregression SAR (5 lags), and seasonal moving average SMA (4 lags). The mean level of the series after accounting for differencing and seasonal effects is 0.8454. The forecasted Y value for 2025 is 0.7352.

It should be clarified that in the context of the SARIMA model, the term «seasonality» indicates the presence of internal cyclicity (long-term lags) in the data series, and in this case, it refers specifically to a mathematical property of the series (cyclicity with a period >1), not quarterly or monthly seasonality.

3,2000

2,7000

1,7000

1,2000

0,7000

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

The next indicator in sequence is Ib; its dynamics are shown in Fig. 7.

Figure 7. Dynamics of Ib changes for "MMK" over 25 Years

Conducting process identification as before reveals (see Fig. 8) that this process is a stationary white noise process with a mean approximately equal to 0.8761.

Figure 8. Identification and forecasting of the Ib observation series in Mathematica

In this study, the Interest Burden (Ib) according to RAS was calculated using the formula: Ib = Profit Before Tax / (Profit Before Tax + Interest Expense)

For «MMK» over the analyzed period 2000–2024, the Ib indicator, calculated per RAS, demonstrates statistical properties close to a stationary process. The result of identifying the Ib observation series as stationary "white noise" is an empirical finding for this company over the analyzed period, indicating relative stability in the ratio of operating profit to financial expenses under RAS accounting. The forecasted Ib value for 2025 is 0.8761.

A similar situation arises in the analysis of the Tb process (see Figs. 9-10).

1,2500

1,0500

0,8500

0,4500

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 9. Dynamics of Tb changes for "MMK" over 25 years

Figure 10. Identification of the Tb observation series in Mathematica

The only difference lies in the mean value of this latter process, which (as is the forecasted value) is approximately equal to 0.7862.

Thus, extrapolating one period forward yields the following values for the Return on Sales coefficient, Turnover ratio, Interest Burden, and Tax Burden, respectively will be the next: $ROS_1 = 0.167$; $Y_1 = 0.7352$; $Ib_1 = 0.8761$; $Tb_1 = 0.7862$.

Substituting these into formula (6) gives the predicted value of the multiplier $M_1 = 1.8187$, which differs from the multiplier $M_1 = 1.4563$ predicted based on identifying its observation series via a second-order moving average model (see Fig. 11):

$$M_t = 1.694 + \varepsilon_t + 0.806\varepsilon_{t-1} + 0.15\varepsilon_{t-2},$$

where ε_{t-i} (i=0,1,2) is white noise. This discrepancy arises because the forecasting did not account for the relationship between indicators according to formula (1).

Figure 11. Identification and forecasting of the multiplier M observation series in Mathematica

5. Discussion

To model the capital structure using formula (6) with a given (desired) ROE and considering the relationship between preceding and forecasted determinant values (ROS, Y, Ib, and Tb), several scenarios for potential capital structuring are considered under the conditions presented in Table 2.

Table 2. Capital structuring options depending on the forecasting result of model determinants and ROE

No	Indicator	Previous value,2024, (fact)	Forecast (ARIMA) without taking into account the relationship between determinants,2025	Modeling capital structure with a given △ROE (lag model) △ ROE→ forecast by ARIMA, 2025
1	ROS	0,188	0,167	0,167
2	Y	0,815	0,735	0,735
3	Ib	0,972	0,995	0,995
4	Tb	0,759	0,787	0,787
5	ROE	0,152	0,175	0,023
6	M	1,34	1,46	1,82
7	Equity,%	74,7	68,5	54,9
8	Debt capital,%	25,3	31,5	45,1

Clarifications regarding the determination of values in row No. 6: Substituting «pure» forecasted determinants into the classical DuPont model does not ensure the equality of the resulting ROE indicator to the product of ROS, Y, M, Tb, Ib, as ARIMA extrapolation does not account for the mutual influence of each factor. The lag model represented by formula (6) formalizes the interrelationship of factors and shows how the combined change in ROS, Y, Ib, and Tb (forecasted via ARIMA), given a target Δ ROE value, determines the necessary change in multiplier M_1 , i.e., the capital structure. Applying the lag model in conjunction with the Box-Jenkins time series identification and forecasting methodology enables a shift from a retrospective approach to analyzing factor influence on ROE to a proactive one, allowing for modeling capital structure with a specified change in ROE based on a statistically grounded forecast of the model's operational and financial determinants.

This is demonstrated with a specific example. According to formula (6), calculating M_1 requires knowing $\Delta ROE = ROE_1$ - ROE_0 , where ROE_0 is the actual 2024 profitability of 15.2%. For ROE_1 , either the forecasted or a desired (target) value can be used. For this example we will utilize the forecasted value obtained after identifying this process within the ARIMA model family. Substituting into formula (6) the actual 2024 values for ROS_0 , Y_0 , Tb_0 , Ib_0 and their extrapolated values ROS_1 , Y_1 , Tb_1 , Ib_1 , along with the obtained $\Delta ROE = 0.175 - 0.152 = 0.023$, allows calculation of the new multiplier value M_1 , which in our case equals 1.82. Converting this multiplier value into equity and debt proportions is done using standard formulas:

M = Assets / Equity

Eq = Assets / M

D = Assets - Eq.

The result of modeling the representative agent's capital structure with the expected return on equity is presented in Fig. 12.

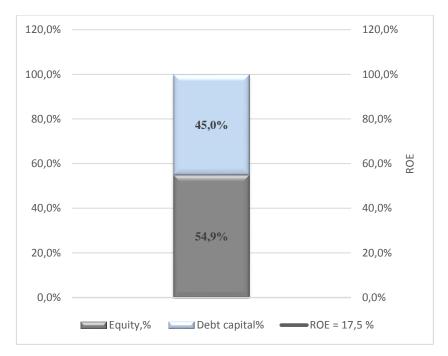


Figure 12. Capital structure of "MMK", generated based on the lag model and forecasting, 2025

The choice of capital structure is undoubtedly a multi-criteria task. Depending on management goals, the optimality criterion could be financial stability, profit maximization through minimizing the cost of capital (WACC), or ensuring a specific return for owners. In the context of this study, ROE serves not as an optimality criterion but as a target modeling parameter, allowing verification of the capital structure modeling algorithm using a combination of the lag model and determinant extrapolation. The proposed mechanism is modular and can be further integrated with other criteria (e.g., minimizing WACC) within a game-theoretic problem formulation (see, for example, [14]).

6. Conclusion

Forecasting capital structure based on lag models of the form (6) should be preceded by studying the time series structure, confirming the possibility of expressing its subsequent value through the preceding one via a first-order autoregressive model. Correct model identification requires a sufficient dataset (annual, semi-annual, quarterly). The Box-Jenkins methodology applied to the time series integrates the aggregate influence of all past external and internal factors into the structure of the models itself (autoregressive components, moving average, etc.), which enables the application of justified extrapolation for proactive capital structure management. Further use of the lag model can follow several scenarios regarding forecasting possible values of ROS, Y, Ib, and Tb. Beyond the time series theory methods discussed here, neural network and other machine learning models could be employed [8]. It would also be interesting to compare the behavior of the key determinant series of model (6) with corresponding series for other enterprises and other countries in the metallurgical industry over a similar period, raising questions about the stability and resilience of industry-specific economic processes [1]. However, these are topics for future consideration.

References

- [1] Bogatov E. M., Demidova E. G. "On One Statistical Method for Analyzing the Balance of Economic Processes". Economics and Mathematical Methods, vol. 60, no. 1, (2024), pp. 121–132.
- [2] Demidova E. G., Bogatov E. M. "Modeling the Structure of Corporate Capital Based on the DuPont Five-Factor Model". The Fifth Russian Economic Congress "REC-2023". Vol. XIV. Firm Economics, Industry Markets and Industrial Policy. Moscow, (2023), pp. 26–30.
- [3] Demidova E. G., Bogatov E. M. "A Five-Factor Lag Model for Optimizing the Capital Structure of a Corporation According to the Criterion of Ensuring a Given Return to Shareholders". Financial Management, no. 5, (2025). pp. 74–79.
- [4] Kolesnikova S. N. "Uncertainty of the External Environment as a Factor of Prognostic Instability". Fundamental Research, no. 5, (2023), pp. 142–147.
- [5] Moody, J. Forecasting the Economy with Neural Nets: A Survey of Challenges and Solutions. In: Montavon, G., Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. (2012).
- [6] Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. "Time Series Analysis: Forecasting and Control". 5th ed. Wiley, (2015).
- [7] Brooks C. "Introductory Econometrics for Finance". 4th ed. Cambridge University Press, (2019). DOI: 10.1017/9781108524872
- [8] Gaytan J. C., Ateeq K., Rafiuddin A., Alzoubi H. M., Ghazal T. M., Tariq Ahamed Ahanger T. A., Chaudhary S., Viju G. K. "AI-based prediction of capital structure: Performance comparison of ANN SVM and LR models". Computational Intelligence and Neuroscience, (2022). DOI: 10.1155/2022/8334927.
- [9] Goodfellow I., Bengio Y., Courville A. "Deep Learning". MIT Press, (2016).
- [10] Gujarati D. N., Porter D. C. "Basic Econometrics". 5th ed. McGraw-Hill, (2009).
- [11] Hastie T., Tibshirani R., Friedman J. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". 2nd ed. Springer, (2009). DOI: 10.1007/978-0-387-84858-7.
- [12] Hyndman R. J., Athanasopoulos G. "Forecasting: Principles and Practice". 3rd ed. OTexts, (2021).
- [13] Stine R. A. "Time Series Models and Mathematica. In: Economic and Financial Modeling with Mathematica". Ed. by H. R. Varian. Springer, New York, (1993).
- [14] Tudose B. M. "Game theory: minimising the cost of capital vs. Maximising the return of investors". Studies and Scientific Researches. Economics Edition, no. 20, (2014), pp. 103–108.