Prytomanova O.,
Doctor of Physical and Mathematical Sciences, Professor,
Neshchasnyi O.,
Bachelor Student,
Kyiv National Economic University named after Vadym Hetman

MODELING THE MACROECONOMIC FACTORS' IMPACT ON THE SALES DYNAMICS OF A DISTRIBUTION COMPANY

Ukrainian`s economy is developing under the pressure of full-scale war, which causes unstable social-economic and political conditions for business development. Distribution companies, especially those who has partnership with international brands, are highly vulnerable to macroeconomic disruptions such as currency volatility, inflation, reduced consumer purchasing power and so on. These companies must adapt their strategies to remain competitive and financially stable.

Developing economic and mathematical models to analyze and predict the impact of macroeconomic factors on sales volume is crucially important to enable timely decision-making and risk mitigation, supporting the sustainability of distribution operations in a turbulent environment.

The aims of this research are to model the influence of macroeconomic factors on the sales dynamics of a distribution company. The analysis includes an identification of key macroeconomic drivers, forecasting using time series models with STL-decomposition and Box-Cox transformation, and GAM data modeling to understand the impact of macroeconomic factors.

Sales policy can be described in both narrow and broad context. In terms of complex understanding, distribution represents a system of delivering goods and services to the end customer and involves sales, logistics, promotional activities, and channel management . While companies can influence their microenvironment through internal decisions and strategies, the macroeconomic environment remains beyond their control [1].

The macroenvironment can be categorized into six key factors: demographic, economic, political-legal, cultural, natural, and technological. Each has a distinct influence on consumer behavior, distribution efficiency, and purchasing [2]. Higher state revenues typically allow for increased public investment, social transfers, and subsidies, which can boost household income and aggregate demand. This often leads to higher consumption levels, benefiting distributors, especially those in consumer goods and retail sectors.

The research employs ARIMA and SARIMA time series models combined with STL-decomposition and Box-Cox transformation to forecast sales allow to adapt to time series with complex trend and seasonal components. These models are applied to different product categories and sales channels [3].

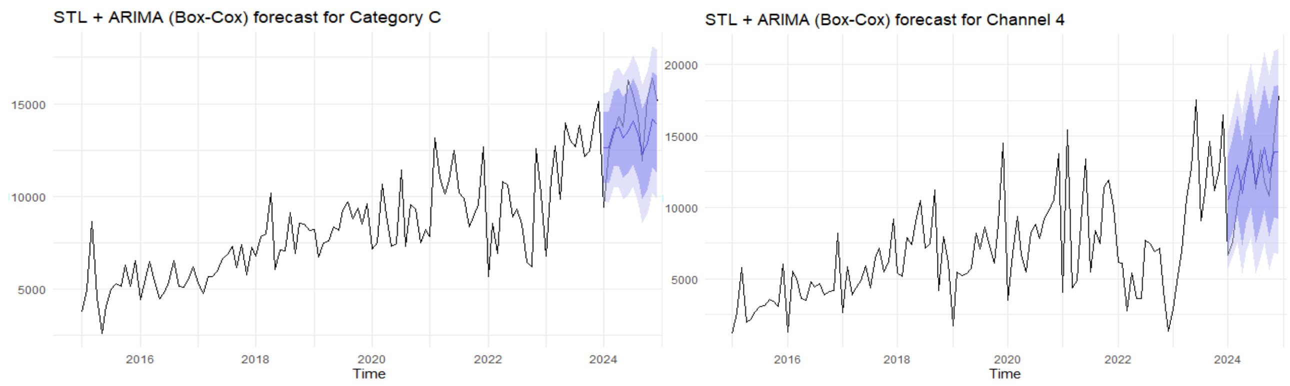
$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t, \quad (1)$$

- *where Y_t — the value of the time series at time;*
- *c — constant, free term;*
- *ϕ_1, ϕ_1 — autoregression (AR) coefficients;*
- *θ_1, θ_2 — moving average (MA) coefficients;*
- *ε_t — residual errors at time t.* [4]

Table 1. The statistical evaluation of the best models among categorical groups

Group	Model	AIC	RMSE	MAE	MAPE
Category C	STL+ARIMA(0,1,1)(0,0,0)[12]	633,28	2145,41	1603,86	10,69
Category D	STL+ARIMA(2,1,0)(0,0,0)[12]	1683,31	1254,57	985,68	10,96
Channel 2	STL+ARIMA(0,1,1)(0,0,0)[12]	-327,21	3005,1	2454,44	18,84
Channel 4	STL+ARIMA(0,1,1)(0,0,0)[12]	826,19	2163,96	1814,79	16,6

Table 1 presents a comparison of STL+ARIMA models for the most representative groups, according to their volume in sales, based on AIC, RMSE, MAE, and MAPE. Category C achieves a strong balance between model fit and accuracy, with a low AIC of 633.28 and a MAPE of 10.69. Category D, while having a higher AIC, shows the most accurate forecasts with the lowest RMSE and MAE. Channel 2 stands out with the best AIC (-327.21) but suffers from the highest forecast errors, indicating a poor fit to the actual data. Channel 4 performs better than Channel 2 but remains less accurate than the category-based models. Overall, Categories C and D demonstrate superior predictive performance.



Picture 1. The results of modeling the dynamic of sales for category C and channel 4

The left plot shows the time series for Category C, which displays a steady upward trend and clear seasonal patterns. The 2024 forecast indicates continued increase, with wide 80% and 95% confidence intervals reflecting uncertainty. Overall, the model performs well, capturing stable and predictable dynamics. On the other side, time series for the 4^{th} channel is more irregular, with sharp fluctuations likely driven by external factors. While the forecast suggests partial stabilization, the broad prediction intervals point to high uncertainty and lower reliability. However, all models require testing their residuals for statistical significance, which will allow us to be confident in their reliability.

Table 2. The p-value of tests for residuals research

Group	Model	test Jargue-Bera	test ARCH	test Ljung-Box
Category C	STL+ARIMA(0,1,1)(0,0,0)[12]	0,4636	0,0407	0,6959
Category D	STL+ARIMA(2,1,0)(0,0,0)[12]	0,5587	0,9863	0,7569
Channel 2	STL+ARIMA(0,1,1)(0,0,0)[12]	0,2136	0,4048	0,3738
Channel 4	STL+ARIMA(0,1,1)(0,0,0)[12]	0,00001	0,9997	0,8981

Overall, all residuals of models satisfy conditions for statistical significance: there is no heteroskedasticity or autocorrelation, which is indicated by high p-values. Moreover, all residuals have normal distribution except for model of Channel 4 with low p-value $< 0,02$.

A Generalized Additive Model (GAM) is a flexible statistical model that extends the Generalized Linear Model (GLM) by allowing non-linear relationships between the predictors and the response variable through smooth functions. This approach is particularly useful when the effect of predictor variables on the response is not strictly linear, enabling more accurate modeling of complex data patterns [4].

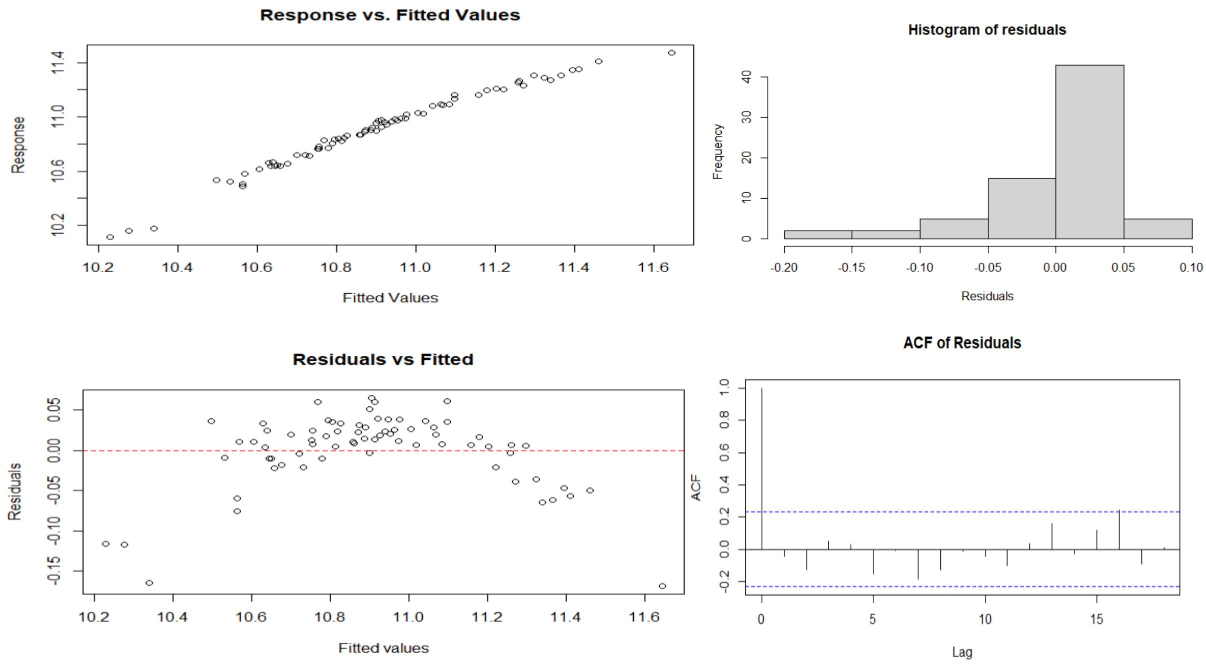
$$g(E[Y_i]) = \beta_0 + \sum_{j=1}^{p} f_i(x_{ij}) \tag{2}$$

- *where Y_i — response variable for observation i;*
- *$E[Y_i]$ — expected value of response;*
- *β_0— intercept;*
- *$f_i\ (x_{ij})$ — smooth function of the j-th predictor for observation i;*
- *p— number of predictors.*

```
Family: gaussian
Link function: identity

Formula:
log(prodazi_tis_grn) ~ s(misac, bs = "cc", k = 12) + s(rik, bs = "cr",
    k = 5) + lag_sales + dohodi_derzavnogo_budzetu

Parametric coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                9.857e+00  2.286e-02 431.109  < 2e-16 ***
lag_sales                  1.912e-05  4.355e-07  43.910  < 2e-16 ***
dohodi_derzavnogo_budzetu -4.654e-08  1.328e-08  -3.504 0.000824 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
             edf Ref.df     F p-value
s(misac) 0.07478 10.000 0.008  0.3563
s(rik)   2.31181  2.786 1.963  0.0836 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.972   Deviance explained = 97.3%
-REML = -78.271  Scale est. = 0.0023174  n = 72
```

Picture 2. Results of non-linear GAM

Both parameters of the model on the picture 2, namely lagged sales and state budget revenues, have very low p-values 0.05. Although the seasonal component was not statistically significant, the smooth trend over years was marginally significant ($p \approx 0.08$). Additionally, the model achieves an excellent result of R^2 being 0.972.



Picture 3. Research of residuals of GAM

Residuals are not symmetric with a slight right skew however it is acceptable for log-transformed data and distribution is close to normal. There is no strong autocorrelation, with all lags within confidence bounds. Response vs fitted is tightly aligned, showing strong predictive alignment. It is suggested some heteroskedasticity, as variance decreases slightly at higher fitted values.

In the research, the GAM model was applied to analyze the impact of macroeconomic indicators on monthly sales. The model showed very strong performance with an adjusted R^2 of 0.972, meaning it can explain 97% of variance of the sales. The results revealed that previous sales and government budget revenues have a significant influence on future sales, suggesting the presence of demand inertia and possible sensitivity of consumer markets to fiscal flows.

In contrast, other factors like the number of registered unemployed, consumer price index, and average monthly wage did not show a statistically significant effect, which may indicate a lagged or indirect transmission of these indicators to actual spending behavior. Additionally, models for time series like STL+ARIMA did not show effective forecast usage for real complex time series due to wide confidence intervals, which limits their practical applicability in volatile economic conditions. The findings support the use of flexible nonlinear models like GAM for understanding and forecasting economic activity when structural complexity is present.

References

1. Trihatmoko, R. A., & Mulyani, R. (2018). Distribution strategy for new product marketing success: Fast moving consumer goods (FMCG) business. *Management and Human Resource Research Journal*, *7*(12), 19-32.

2. Lamb, C. W., Hair, J. F., & McDaniel, C. (2011). Principles of marketing (Chapter 1.3). OpenStax, Rice University.

3. Lem, K. H. (2024). The STL-ARIMA approach for seasonal time series forecast: A preliminary study. *ITM Web of Conferences*, *67*, 01008.

4. Greene, William H. (2012) Econometric analysis. 7th ed. Prentice Hall. Pearson. 1189 p.

5. Miller, D. L. (2025). Bayesian views of generalized additive modelling. *Methods in Ecology and Evolution*, *16*(4), 512-526.

Soloviev V.N.,
Doctor of Physics and Mathematics, Professor
Kryvyi Rih State Pedagogical University
Matviychuk A.V.,
Doctor of Economics, Professor
Kyiv National Economic University named after Vadym Hetman
Bielinskyi A.O.,
PhD Student
Kyiv National Economic University named after Vadym Hetman
Myronenko T.I.,
Bachelor Student
State University of Economics and Technology

ADVANCED PERMUTATION ENTROPY METRICS FOR BITCOIN: TOWARDS ROBUST EARLY WARNING INDICATORS OF MARKET INSTABILITY

Permutation entropy (PEn) is a widely recognized nonlinear statistical measure used to quantify the complexity inherent in time series data [1]. Despite its widespread