Audit Financiar, XXIV, Nr. 1(181)/**2026**, *186-202*
ISSN on-line: 1844-8801

Paper presented at the IX[th] Congress of the Romanian financial auditor profession

Forecasting Budget Revenues Using the ARIMA Time Series Model

Ionela URSU, Ph. D. Student,
Faculty of Economics and Business Administration,
„Alexandru Ioan Cuza" University, Iasi, Romania,
e-mail: ionela.ursu92@gmail.com

Abstract

In recent decades, information has become one of the most important resources of a company, and modern management, based on flexibility and dynamism, requires complex and operational information. But how relevant, pertinent and in time information can be obtained? What methods and tools can be used to provide the most accurate information when it comes to planning and forecasting? A basic tool of accounting and controlling is represented by the budget, especially its forecasting function. In this context, for this study, the forecast of the budget revenues is done by using the ARIMA time series model. An increasing trend in the forecast for the analyzed year was observed, which represents a favorable sign for the company in terms of revenue growth.

Key words: *budget; forecasting; forecasting models; time series; ARIMA; SPSS;*

JEL Classification: *M41*

To cite this article:
Ursu, I. (2026), Forecasting Budget Revenues Using the ARIMA Time Series Model, *Audit Financiar*, vol. XXIV, no. 1(181)/2026, pp.186-202, DOI: 10.20869/AUDITF/2026/181/005

To link this article:
http://dx.doi.org/10.20869/AUDITF/2026/181/005
Received: 16.07.2025
Revised: 4.08.2025
Accepted: 13.01.2026

1. Introduction

Historically, budgets have always played an important role in business (Libby, Lindsay, 2010). The business budget developed mainly in the period 1895–1920 as a result of the industrial development, production standardization and competition in the market. Budgets are used all over the world and serve for several purposes: planning, performance evaluation, employee motivation, resource allocation, control and strategy implementation. Budget represents the fundamental short-term planning tool (usually 1 year), oriented towards a profitable forecast. The objectives of the budget must correspond to the strategy defined at the entity level translated into the plan (Caraiani C., 2010, p. 37). Forecasting is the starting point for budgeting. Thus, to budget means to foresee and decide, and to forecast means to know before acting (Badea F., Dobrin C., 2006, p. 71 and 81).

Currentl, the use of forecasts is not only made for predictions on economic, political, social aspects but also for the real-time business decisions. Forecasting is a commitment to the future but has also a scientific attitude. Forecasts are based on scientific statistical, economic data so forecasting is based on knowledge (Badea F., Dobrin C., 2006, p. 82).

Based on the structure of the budget system, we can distinguish two types of budgets: determinant (represent the foundation of the budget) and resultant (are induced by determinant budgets). Determinant budgets are represented by the sales budget and the production budget. Resultant budgets are represented by the supply budget, administrative expenses, investments, treasury etc. (Budugan D., 2001, p. 52).

Given the concept that the starting point in budgeting is the revenue forecast, we propose to create a sales prediction using time series model that will then represent the basis for the construction of the other budgets, more specifically, the ARIMA time series approach will be used.

The data used are taken from a company whose business activity is production. For confidentiality reasons, we cannot disclose the name of the company, the processed data are real and were extrapolated from the SAP ERP system. For this study, historical data from budgets regarding the manufactured units during the period 2019-2023 will be used and the prediction will be made for the year 2024.

The paper includes 3 sections as follows: the first section aims to analyze the literature with an emphasis on notions regarding forecasting models. This is followed by the presentation of the research methodology, the variables used, the proposed ARIMA models and the results of the model prediction, and the last part aims to present the conclusions of the study.

2. Literature review

The budget system is a short-term, predictive management tool that includes budget planning and control. The objective of budget systems is to obtain information for the purpose of planning events within an entity (Caraiani C., et. al., 2010, p. 42). In practice, the totality of individual budgets coordinated with each other is called a budget system (Horvath et. al., 2009, p. 145).

The budget is a planning tool and the budget objectives must correspond to the entity's strategy translated into a plan (Caraiani C., et. al., 2010, p. 37). The budget expresses the management's vision of the expectations regarding the company's activity for a period of time. The process of achieving and executing the budget involves complex actions because a multitude of variables and factors can affect and influence the achievement or not of the proposed objectives (Rachlin R., 2007, p. 3).

Budgeting also represents, in the authors' view (Horvath et. al., 2009, p. 140), a planning system of the financial results, that determines a clear development of the action plans. The budget is a numerical forecast of the objectives, which must correspond to the entity's strategy (Caraiani C., 2010, p. 37). The budget predicts the anticipated result of a business strategy (Dimitris C.N., 2006, p. 237).

If budgeting represents the data resulting from forecasts (Budugan D., 2001, p. 43), the following questions that can be asked are: What information do we enter into budgets? How do we obtain this information? The forecast represents the starting point of budgeting. Thus, to budget means to foresee and decide, and to forecast means to know before acting (Badea F., Dobrin C., 2006, p. 71 and 81).

Currently, the use of forecasts is not only made for predictions about economic, political, social aspects but also for the real-time business decisions. Forecasting is a commitment to the future but has also a scientific attitude. Forecasts are based on scientific, statistical, economic data, so forecasting is based on knowledge. To forecast, we need information that refers to the present and past status of the enterprise, so that information becomes the raw material for management decisions (Badea F., Dobrin C., 2006, p. 82). Forecasting or prediction comes from the French *"prevision"* which means to anticipate the evolution and occurrence of events, based on known data from the past (but also the present); as well as the study of objective laws in a temporal and spatial context (Guță A.J, Catrina I., 2020).

How are forecasts obtained? With the help of forecasting models. Forecasting models represent the "heart" of a prediction. To get the best results we need the right data but also the right model. The more compatible the data and the model are, the more accurate the prediction of a phenomenon will be. According to the authors Jain L. D. C, Malehorn J. (2006, p. 46) there are three main types of models:

a. **Time Series Model**: this model works based on extrapolation of past data and assumes that the same trend will continue in the future. Examples of such models include exponential smoothing and ARIMA model (Zhao L., Mbachu J., Zhang H., 2019);

b. **Cause-Effect Model**: it is based on the existence of a cause (called the driver or independent variable) and an effect (called the dependent variable). For example, if sales depend on the total advertising expenditure for a product, then sales are the dependent variable and advertising is the independent variable, the driver, the cause. With the help of this model, we can determine the relationship between the variables and project them into the future. Another example starts from the following relationship between expenses (y) and activities (x): y=F(x), where, the expected expenses Y were established starting from the level of activity that can be predicted (Badea F., Dobrin C., 2006, p. 72). This model is used when there is a close relationship between the variables and it will not change significantly in the future or at least in the forecast period;

c. **The Critical Model**. This model is used especially when we do not have previous data for analysis. In this case we will use specific procedures to make a forecast such as: the Analog, Delphi, Diffusion, PERT (performance evaluation review technique), Survey model. We will briefly describe them further.

   - Analog – within this model a similar variable is used to be used as a basis for the forecast. For example, when televisions first appeared on the market, analysts estimated sales based on the sales of radio products. This technique is used for new products on the market, where we do not have a past record;
   - Delphi - in this case the forecasting is carried out with the help of a group of experts in the respective field;
   - Diffusion - analysts forecast data based on the product life cycle;
   - PERT - through this approach the analyst forecasts with the help of an expert in the field who will provide him with 3 estimates (pessimistic, optimistic, most likely) which will serve as the basis for the forecast;
   - Survey - primary data comes from questionnaires, applied by email, telephone or face to face, which represent the starting point for forecasting in certain specific cases.

In an informational context, a certain clarification is required: what is the difference between data and information? Information represents a message, brings an increase in knowledge and reduces uncertainty, and data represent symbols for potential information but requires processing to become useful information. Thus, data become the raw material for obtaining information (Nica P., Iftimescu A., 2008, p. 363). The basis of a forecast is the data we have available. Depending on the data we have available, we can also select the correct model. When preparing data, we must take into account the following aspects (Jain L.D.C., Malehorn J., 2006, p. 52):

- data consistency: we ensure that we have data for the entire analyzed period;
- the appearance of extreme values (outliers): by outlier we understand extreme, unusual, out of the ordinary values. These are random occurrences that are not part of the model, so they are either adjusted or a model is chosen that keeps track of such an occurrence, for example using a dummy variable in regressions that adjust the outlier;
- data structure: these can change with the appearance of new products, mergers of companies, entry into a new market, etc., in such cases comparison with past data is no longer possible;

- missing data: if we are missing data from a certain period (e.g. a month, 1 year), we must insert the missing values (through estimates); if we want to apply the time series model for example, we could encounter problems;
- seasonality: refers to fluctuations that occur periodically. For example, sales of a store with certain specific products can be very high in December, when Christmas takes place, or certain sports equipment can be sold more in winter than in summer and vice versa. If we observe such a trend, we will take it into account during the forecasting;
- the existence of a cause-effect relationship between the data: if our sales are directly affected by the practice of advertising campaigns, then the cause-effect model is more appropriate to use because Time Series will not capture such a relationship between the data;
- errors in the forecasting process: these are the most difficult to measure. We can calculate the errors, as appropriate, using one of the formulas (Jain L.D.C., Malehorn J., 2006, p. 60):
    a. Forecast error (%) = (Actual-Forecast)/Actual*100 - shows us how much the forecast deviates from the current data;
    b. Forecast error (%) = (Forecast-Actual)/Forecast*100 - shows us how much the actual data deviates from what we predicted.

The methodology applied depends on the company's business. For example, for products such as cement, oil, electricity, the forecast can be made for each product separately. For consumer products, the forecast is made in terms of production capacity or sales. There may be restrictions at the forecast level, for example restrictions generated by the market where the company will sell its products, the appearance of competition, etc., which is why we resort to methods that reduce this uncertainty (Badea F., Dobrin C., 2006, p. 126).

3. Research methodology. ARIMA time series model

Among the functions of the budget, the forecasting function stands out (in addition to the financial control and balance function), which represents a financial estimate of the resources and expenses for the company's activities (Budugan D., 2001, p. 46). Given that the starting point in the development of budgets is the revenue forecast, we further propose to create a sales prediction model using time series that will then represent the basis for the construction of the other budgets. The time series data are ordered according to the ***time variable*** (Cărbunaru B., Băcescu C., 2013). The general expression of a time series is (Anghelache C., Manole A., 2012):

$$Y_i=f(t_i), \text{ where}$$

Y_i= values of the studied variable

t_i= numerical values of the time variable

It is often preferred **to replace the effective independent variables**, which act on sales (such as: consumer income level, inflation level, population structure, product price, advertising budget, etc.) with the ***time variable***; this is due to the lack of information about the causal variables but also because of the high costs of obtaining such information (Zamfir M., 2017, p 65).

Budget planning always begins with the sales budget, which indicates the physical quantity of units to be produced and sold in a given period (Chorafas Dimitris 2006, p. 237). This budget then represents the basis for the production cost budgets, which will be taken into account for the quantities of raw materials and materials required per unit of product, from which the materials cost budget will result. Taking this information into account, the procurement budget is then developed. The budget for direct and indirect salary expenses is then planned, as well as the general administrative expenses budget. The investment budget takes into account the production budget for the purchase or modernization of equipment. The transition from numerical to cash-based budgeting represents the budgeting of cash flow as well as the development of a planned balance sheet and the forecasted income statement (Horvath et. al., 2009, p. 154).

Sales forecasting can be done: for the short term (for the development of production programs, including the forecast of finished product stocks or supply programs with raw materials and materials, etc.) and for the long and medium term (for the development of investment programs and financing plans), (Badea F., Dobrin C., 2006, p. 121). The authors Ikenna A., et. al. (2017) argue that the estimation, forecast of sales represents the foundation, the starting point for the construction of budgets. The forecast of budget revenues was carried out for this study using chronological series, more specifically the ARIMA time series model. We show in **Figure no. 1** the stages of this process, which will then be presented in detail.

Figure. no. 1. Steps used for the ARIMA model

Research framework for the ARIMA model

1. Graphical analysis of the historical data;

↓

2. Stationarity testing (graphical and numerical method);

↓

3. Transformation of the series from non-stationary to stationary;

↓

4. Parameters setting for the ARIMA model (p, d, q);

↓

5. Parameters setting for the SARIMA model (p, d, q) (P, D, Q);

↓

6. Testing and selecting the models by using the accuracy indicators, information criteria, model coefficients, the Ljung-Box test for the errors' stationarity and the correlograms of the model errors;

↓

7. Parameters estimation and model validation;

↓

8. Making the predictions for the chosen model.

Source: own processing

The first concept to clarify when talking about ARIMA (Auto-Regressive Integrated Moving Average) time series is whether a series is stationary or not. Determining this concept is important because the stationarity of a series can strongly influence its behavior and properties (Brooks C., 2008, p. 233). A series that is stationary is characterized by the fact that the mean, variance and covariance do not depend on time, they are constant. A non-stationary series, on the other hand, does not respect one of the conditions listed above, means the value of the variable depends on the previous value. If a series is non-stationary, the phenomenon of artificial correlation between variables may occur, means, although the value of the coefficient is high, there is no concrete connection between them. The phenomenon of false regression may also occur, there is a logical connection, the model is significant but we do not have a logical connection between the variables (Jemna D., 2009, p. 209-210).

In practice, of course, it is very difficult to find stationary series, which is why we will resort to series transformation procedures. For testing stationarity, graphical methods (based on autocorrelation coefficients) and numerical methods (based on tests) can be used. To achieve these, we used the SPSS v30 statistical program, and the data used are represented by historical data from the revenue budget (in physical volumes).

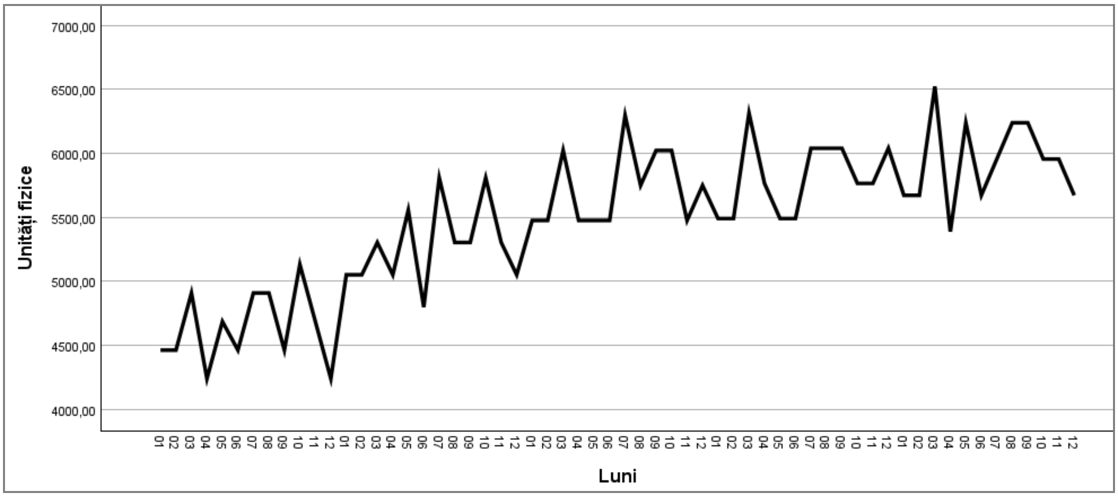
First of all, we present the data series in **Figure no. 2**, where at a first analysis we already realize that we have a stationary series because we observe an increasing trend and not a constant one.

Before proceeding to the testing with the two models, we want to specify some concepts that we will use. The first is the autocorrelation function (ACF), which measures the intensity of the link between variables and takes the value between -1 and +1. The partial autocorrelation function (PACF) measures the intensity of the link between variables, but also controlling the influence of the variables for a lag k. The lag represents an operator that establishes the correspondence between the value of the series at a time t and a previous one. The difference operator is used to transform a non-stationary series into a stationary series and establishes the correspondence between the value of the series at time t and the difference between the value of the series at time t and the value of the series at a previous time (Jemna D., 2009, p. 211).

Using the graphical method to test stationarity, we will use the correlogram of the autocorrelation coefficients. If the values of the autocorrelation function remain within the limits of the confidence interval, it means that the series is stationary. On the other hand, if the values decrease exponentially, it means that the series is non-stationary (Jemna D., 2009, p. 212).

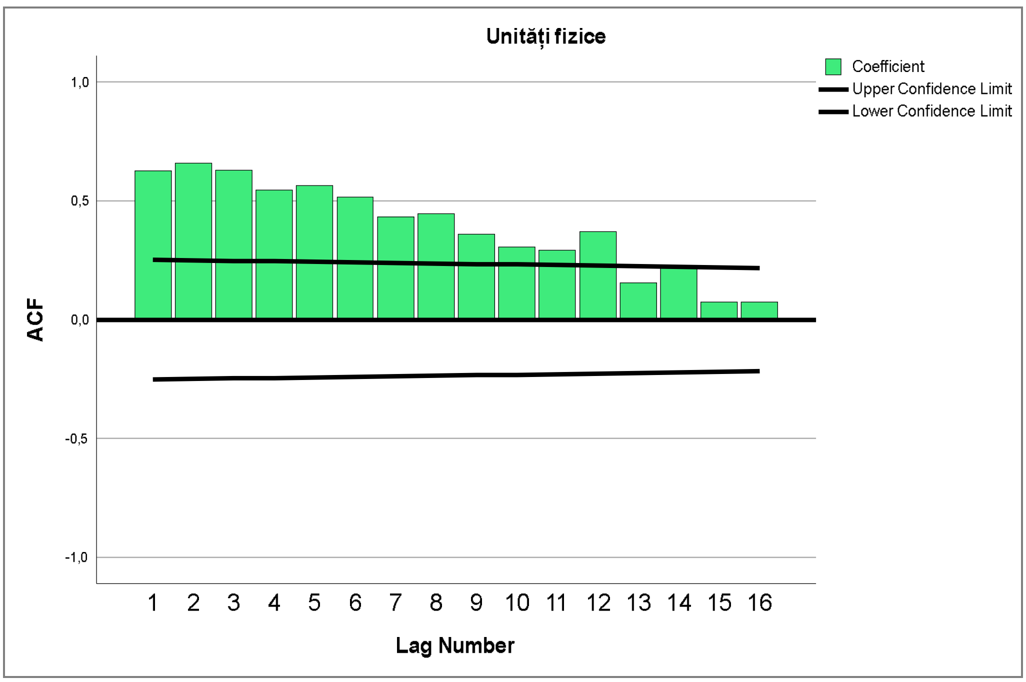
We can easily see from the correlogram represented in **Figure no. 3** that the series is non-stationary, because the values decrease and exceed the limits of the interval up to a lag equal to 12.

Figure. no. 2. Data series



Source: own processing in SPSS

Figure. no. 3. Correlogram of the non-stationary series



Source: own processing in SPSS

The numerical method is carried out using the Ljung-Box test (**Table no. 1**), starting from the formulation of the hypotheses (Jemna D., 2009, p. 213):

H0: ρk= 0, all coefficients are simultaneously zero – the series is stationary

H1: ρk all coefficients are not simultaneously zero – the series is not stationary

Result that Sig =<0.001<0.05, means for each lag the significance of the test is <0.001 → the hypothesis H0 is rejected and the hypothesis H1 is accepted, means the series is non-stationary.

Table no.1. Numerical testing method

Autocorrelations

Series: Unități fizice (Physical units)

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.626	.126	24.712	1	<.001
2	.659	.125	52.588	2	<.001
3	.630	.124	78.464	3	<.001
4	.546	.123	98.247	4	<.001
5	.563	.122	119.716	5	<.001
6	.517	.120	138.115	6	<.001
7	.431	.119	151.183	7	<.001
8	.445	.118	165.366	8	<.001
9	.360	.117	174.826	9	<.001
10	.306	.116	181.801	10	<.001
11	.292	.115	188.288	11	<.001
12	.371	.114	198.965	12	<.001
13	.156	.112	200.883	13	<.001
14	.224	.111	204.945	14	<.001
15	.075	.110	205.404	15	<.001
16	.073	.109	205.856	16	<.001

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Source: own processing in SPSS

The non-stationary time series can be transformed into a stationary series using the difference operator of a certain order. This transformation can be performed using SPSS (menu forecasting, sequence chart, transform, difference 1). In **Figure no. 4** we can see the data series transformed into a stationary series of the first order (means we applied the operator only once to obtain the stationary series).

The ARIMA process is determined by 3 parameters: p, d and q and the model is written in the form:

ARIMA (p, d, q),

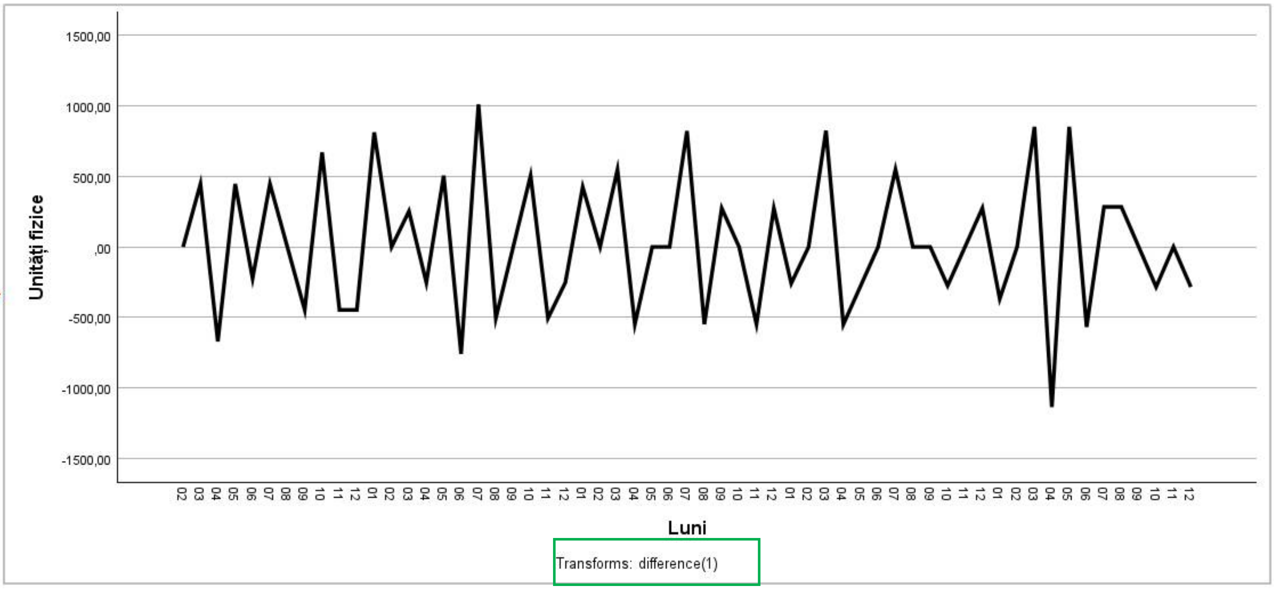
where p represents the autoregressive component (AR), d the order of integration (I) and q the moving average component (MA).

How do we decide the parameters for the ARIMA model? With the help of the autocorrelation and partial autocorrelation function for a stationary series (our series has already been transformed as described above), as follows:

- the parameter **p** refers to the number of past values (lags) that the model uses to make predictions (Gupta A., Mehta V., 2023) and can be determined using the partial autocorrelation function, more specifically the partial autocorrelation correlogram (**Figure no. 6**), which we obtain using SPSS. Based on this graph, we observe the terms that are significantly different from zero;
- the parameter **d** refers to the number of differencings performed to eliminate non-stationarity (Gupta A., Mehta V., 2023) and in our case it is 1, because the difference operator used above is of order 1;
- the parameter **q** refers to the number of past error terms used by an ARIMA model when making

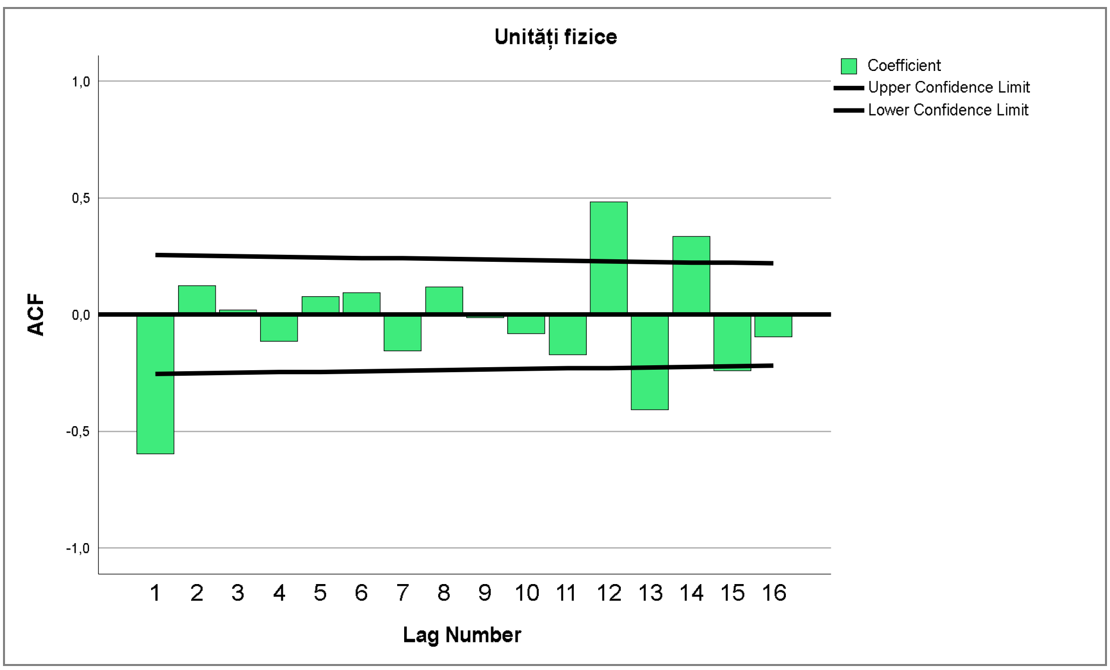
predictions (Gupta A., Mehta V., 2023) and can be determined using the autocorrelation correlogram (**Figure no. 5**), where we observe the terms that are significantly different from zero;

Figure. no. 4. Transforming non-stationary series into first-order stationary series



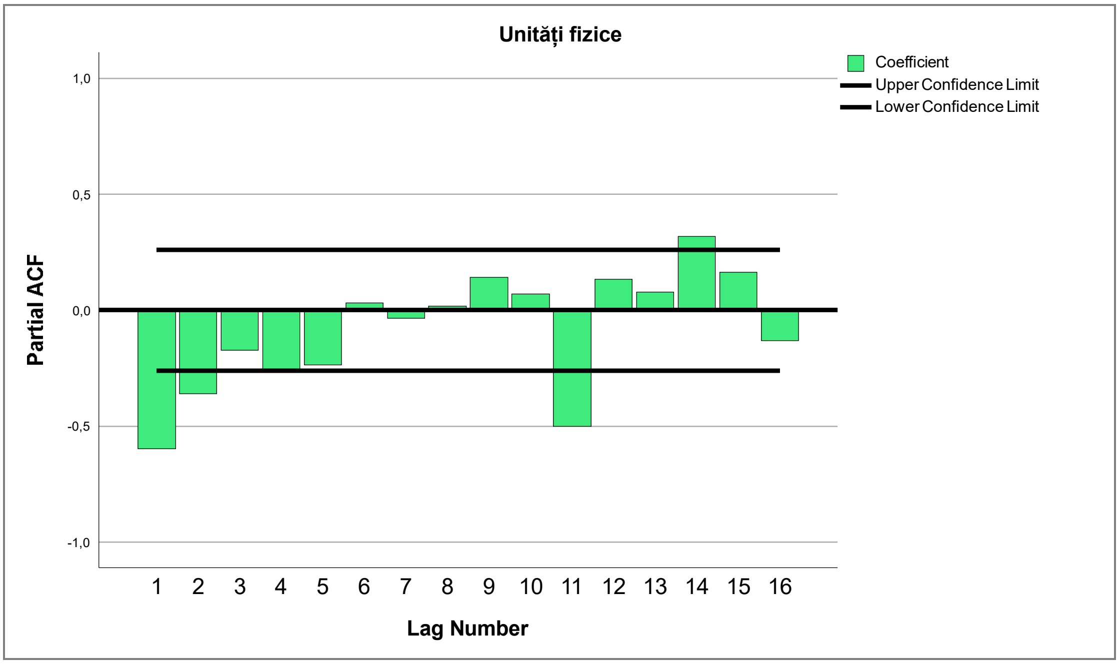
Source: own processing in SPSS

Figure. no. 5. Autocorrelation correlogram of the transformed series



Source: own processing in SPSS

Figure. no. 6. Partial autocorrelation correlogram of the transformed series



Source: own processing in SPSS

When the data in the model shows elements of seasonality, we refer to a seasonal ARIMA model, also called SARIMA. The data in the model shows seasonality in certain periods when they are influenced by year-end or religious holidays (Devianto D., et. al., 2024), by holiday periods where, for example, the airline industry records higher demands (Firat M., Kaplan Y.D., Sampli R., 2021). Retail sales can also show seasonality, such as the food industry in retail trade (Nari Sivanandam N. A., Ahrens D., 2015) or certain equipment that tends to have high or low values during the winter and summer (Noureen S., Atique S., Roy V., Bayne S., 2019).

When ARIMA incorporates, in addition to seasonality, external variables into the model (exogenous factors), means external information such as weather, exchange rate etc. (Gupta A., Mehta V., 2023), we speak of a model called SARIMAX (Nari Sivanandam N. A., Ahrens D., 2015).

Like non-seasonal time series, data can be modeled and forecasted as an ARIMA process. The parameters and the SARIMA model are written as follows (Noureen S., Atique S., Roy V., Bayne S., 2019):

ARIMA(p, d, q) (P, D, Q),

where the terms p, d, q represent the non-seasonal component and the terms P, D, Q represent the seasonal component of the model.

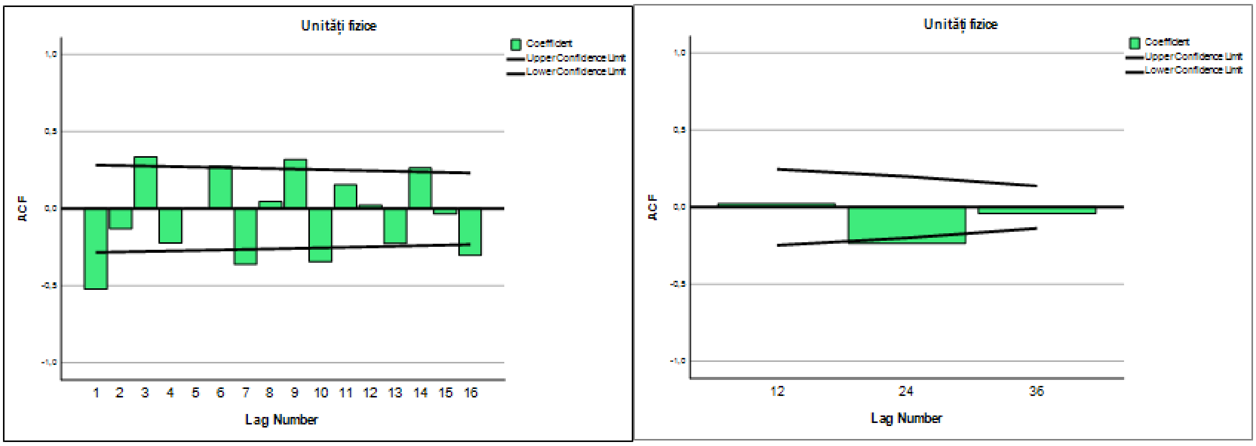
Next, we can show the autocorrelation correlogram and the partial autocorrelation correlogram of the transformed (order 1) SARIMA series (Figures no. 7 and no. 8) for our data, which were obtained using SPSS v30. The lags of the correlograms with the seasonal component are presented with a seasonality of 12 months (12, 24, 36). These graphs will help us determine the model parameters.

Moving average modeling (MA parameter) can pose problems in practice; in this case, the author Jemna D. (2009, p. 226) presents the solution of testing and using only the AR – p parameter in the model. We will also take this information into account when testing the models.

We now have all the necessary information to identify and test ARIMA and SARIMA models in SPSS.

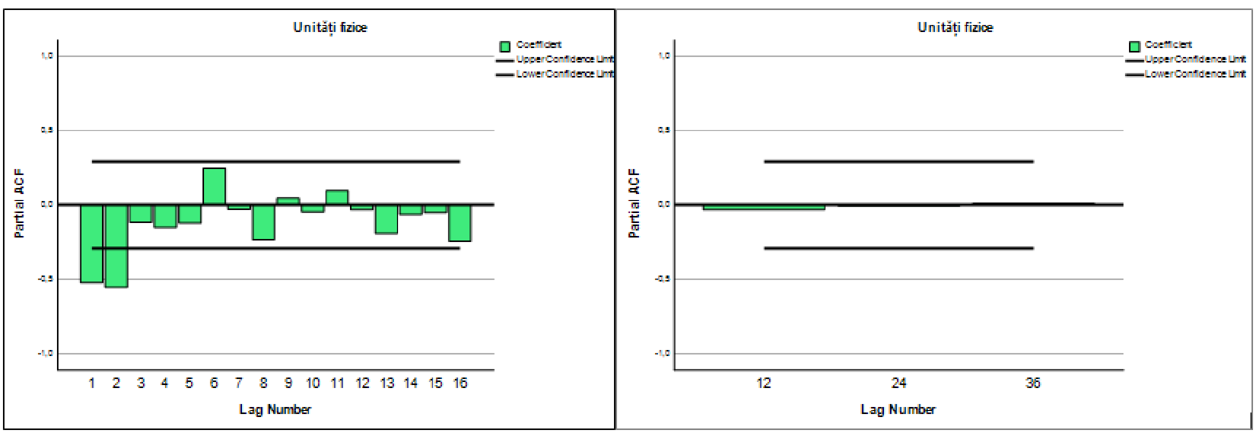
The literature urges us, for model selection, to use not only correlograms, but also other techniques that remove some of the subjectivity in the interpretation of ACF and PACF, called ***information criteria***. The three most popular criteria are: Akaike's (1974) information criterion (AIC), Schwarz's (1978) Bayesian information criterion (BIC) and Hannan-Quinn criterion (HQIC), (Brooks C., 2008, p. 232-233). The lowest BIC/AIC criterion (value) is preferable (Tripti D., Shamshad A., Mohammad S., 2020). SPSS provides us with the BIC criterion that we will also use in our analysis.

Figure. no. 7. Autocorrelation correlograms of the transformed series (non-seasonal and seasonal component)



Source: own processing in SPSS

Figure. no. 8. Partial autocorrelation correlogram of the transformed series (non-seasonal and seasonal component)



Source: own processing in SPSS

In addition to information criteria, ***accuracy indicators*** (which measure actual vs. predicted values) can be used, such as MAE (mean absolute error), RMSE (root mean square error) and MAPE (mean absolute percentage error); of course, the indicator with the lowest values may represent the most appropriate model (Ho J.S., Zhang Y., 2022). We will choose to use MAPE, which expresses the results in relative values.

The calculation formula is given below (Sagaert Y. R., et. al, 2017):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{Y_{t+h} - Y'_{t+h}}{Yt+h}, \text{ where}$$

Yt+h represents the real data (original data),

Y't+h are the forecasted values.

Table no. 2. Proposed ARIMA and SARIMA models (ordered by MAPE indicator)

No.	MODEL	MAPE (%)	BIC	R-squared	Model coefficients (sig<0.05 for AR and MA)	Ljung-Box test – errors stationarity (sig<0.05)	Observations – residuals correlograms	Total predicted units
0	-	*(3,2% - 5,6%)*	*(11,43 - 12,78)*	*(51% - 79%)*	-	-	-	*(69.758 - 74.775 pcs.)*
1	SARIMA(2,1,3)(0,1,2)	3,202	11,571	0,675	not significant statistically	0,174	Errors are within range	70.768
2	SARIMA(2,1,1)(0,1,2)	3,352	11,501	0,624	not significant statistically	0,089	Errors are within range	71.811
3	SARIMA(1,1,3)(0,1,2)	3,451	11,656	0,606	not significant statistically	0,024	Errors are within range	72.076
4	SARIMA(2,1,3)(0,1,0)	3,477	11,498	0,626	not significant statistically	0,34	Errors are within range	70.127
5	SARIMA(2,1,1)(0,1,0)	3,579	11,442	0,593	**statistically significant**	0,098	Errors are within range	70.843
6	SARIMA(2,1,0)(0,1,2)	3,581	11,47	0,595	not significant statistically	0,006	Errors are within range	70.907
7	SARIMA(1,1,3)(0,1,0)	3,631	11,572	0,552	not significant statistically	0,022	Errors exceed the range	70.665
8	SARIMA(1,1,1)(0,1,2)	3,635	11,527	0,572	not significant statistically	0,005	Errors are within range	71.750
9	ARIMA(11,1,0)	3,71	12,134	0,784	not significant statistically	0,016	Errors are within range	74.424
10	ARIMA(11,1,1)	3,741	12,718	0,793	not significant statistically	<,001	Errors are within range	74.049
11	**SARIMA(2,1,0)(0,1,0)**	**3,793**	**11,445**	**0,513**	**statistically significant**	**0,01**	**Errors are within range**	**69.758**
12	ARIMA(2,1,12)	3,854	12,4	0,785	not significant statistically	<,001	Errors are within range	71.926
13	SARIMA(1,1,1)(0,1,0)	3,872	11,433	0,519	not significant statistically	0,03	Errors exceed the range	70.341
14	ARIMA(2,1,14)	3,937	12,602	0,781	not significant statistically	<,001	Errors exceed the range	73.983
15	ARIMA(2,1,13)	4,218	12,702	0,735	not significant statistically	<,001	Errors exceed the range	73.119
16	ARIMA(2,1,1)	4,935	11,891	0,655	not significant statistically	<,001	Errors exceed the range	74.521
17	ARIMA(1,1,1)	4,946	11,807	0,654	not significant statistically	<,001	Errors exceed the range	74.520
18	**ARIMA(2,1,0)**	**5,363**	**11,962**	**0,596**	**statistically significant**	**<,001**	**Errors are within range**	**73.879**
19	ARIMA(1,1,0)	5,64	12,021	0,532	**statistically significant**	<,001	Errors exceed the range	71.247

Source: own processing

In **Table no. 2** we present the proposed models which are ordered by the MAPE (Mean Absolute Percentage Error) indicator.

We observe from **Table no. 2** that for the analyzed models, MAPE has values between 3.2% and 5.6%. This represents a high accuracy of the models' forecast (Seval E., Nursel O., 2017), as we can see from **Table no. 3**. Likewise, the BIC values are very close to each other (11.43 - 12.78) and R Squared (the determination ratio that shows us what percentage of the variation of physical units is influenced by the variation of the time variable) has values between 51% - 79%.

After analyzing the indicators, to choose the optimal model for forecasting, several criteria were taken into account: the model coefficients (for a sig.<0.05 for AR and MA), the results of the Ljung-Box test on the stationarity of the model errors (for a sig.<0.05) as well as the analysis of the residuals (white noise) correlograms; this information helps us to validate the final model.

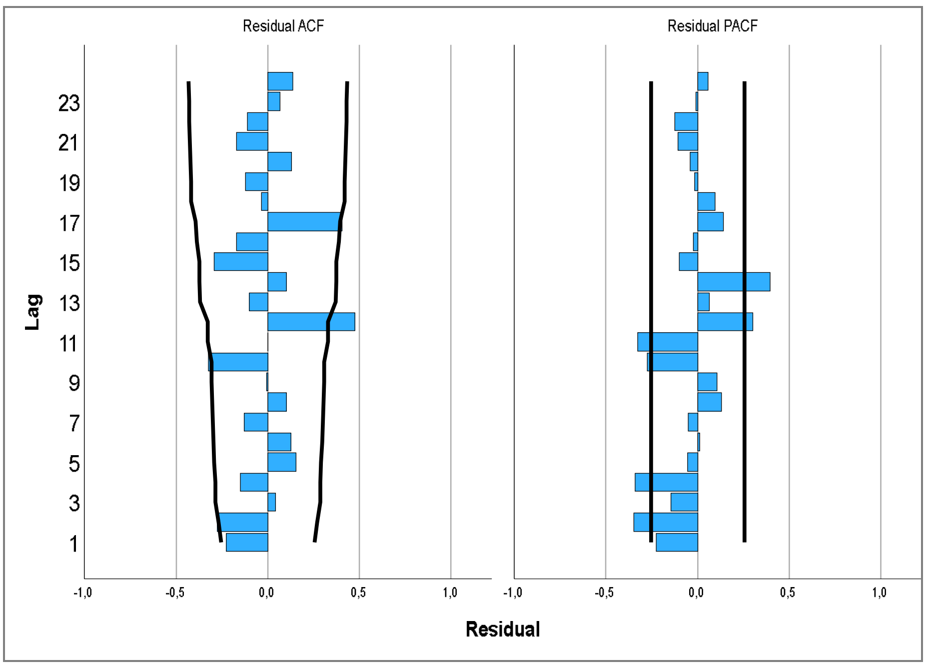
Table no. 3. Ranges for MAPE and forecast accuracy

No.	Reference values - MAPE	Forecast
1	≤ 10%	High accuracy
2	10% - 20%	Good accuracy
3	20% - 50%	Feasible accuracy
4	≥50%	Low accuracy

Source: Seval E., Nursel O., 2017

From the **Table no. 2** we can deduce that only models from the 5, 11, 18 and 19 rows have statistically significant coefficients (sig.<0.05) for AR (p) or MA (q). The rest of the models that have statistically insignificant coefficients (sig.>0.05) can no longer be considered for prediction.

Figure. no. 9. Residual values for the ARIMA(1,1,0) model



Source: own processing in SPSS

We proceed with the analysis of the remaining models, SARIMA(2,1,1)(0,1,0), SARIMA(2,1,0)(0,1,0), ARIMA(2,1,0) and ARIMA(1,1,0).

The SARIMA(2,1,1)(0,1,0) model in row 5 does not meet the condition regarding the stationarity of the model errors (sig.= 0.098 > 0.05), means the residuals are non-stationary.

SARIMA(2,1,0)(0,1,0) represents a model that can be used (row 11), sig.= 0.01, means the estimated model errors are stationary and according to the correlograms residuals, the errors are within the confidence interval.

ARIMA(2,1,0) also represents a potential model (row 18) with a sig.<.001, meaning that the estimated model errors are stationary and the correlograms residuals are within the confidence interval.

The ARIMA(1,1,0) model on row 19, although records a sig. <0.05, meaning that the estimated model errors are stationary, the correlogram errors exceed the confidence interval, as can be seen from **Figure no. 9**.

Finally, the SARIMA(2,1,0)(0,1,0) and ARIMA(2,1,0) models meet the criteria used, but which one can we choose in the end? We observe from table no. 2 that although the stationarity hypothesis of the estimated errors is accepted for both models, the sig. value is more significant for the ARIMA(2,1,0) model, means sig.=<.001, compared to SARIMA(2,1,0)(0,1,0) for which sig. is 0.01.

Based on these arguments, we choose the ARIMA(2,1,0) model and proceed with the forecasting process.

We can now proceed with the estimation of the ARIMA(2,1,0) model, using the SPSS v30 program, from the Analyze, Forecasting, Create Traditional Models menu.

The estimate of the autoregressive parameter is shown in **Table no. 4**. We observe that lag 1 (means the lagged variable is used as a predictor) is significant for sig.= <.001 and lag 2 is significant for sig.= .005.

We proceed with the validation of the model and verify the hypotheses regarding the stationarity of the model errors for a sig. <0.05.

H0: ε≠0, the estimated model errors are non-stationary

H1: ε=0, the estimated model errors are stationary.

According to the Ljung-Box test (from **Table no. 5**) it results that Sig is <.001, means <0.05 → the hypothesis H0 is rejected and the hypothesis H1 is accepted, means the hypothesis regarding the stationarity of the errors is accepted.

Table no. 4. ARIMA model parameters (2,1,0)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Physical units - Model_1	Physical units	No transformation	Constant		22.375	21.521	1.040	.303
			AR	Lag 1	-.815	.125	-6.538	<.001
				Lag 2	-.366	.125	-2.937	.005
			Difference		1			

(Source: own processing in SPSS)

Table no. 5. ARIMA statistical model (2,1,0)

Model Description

			Model Type
Model ID	Physical units	Model_1	ARIMA(2,1,0)

Model Statistics

Model	Model Fit statistics					Ljung-Box Q(18)			
	R-squared	RMSE	MAPE	MAE	Normalized BIC	Statistics	DF	Sig.	Number of Outliers
Unități fizice-Model_1	.596	356.799	5.363	293.430	11.962	64.230	16	<.001	<.001

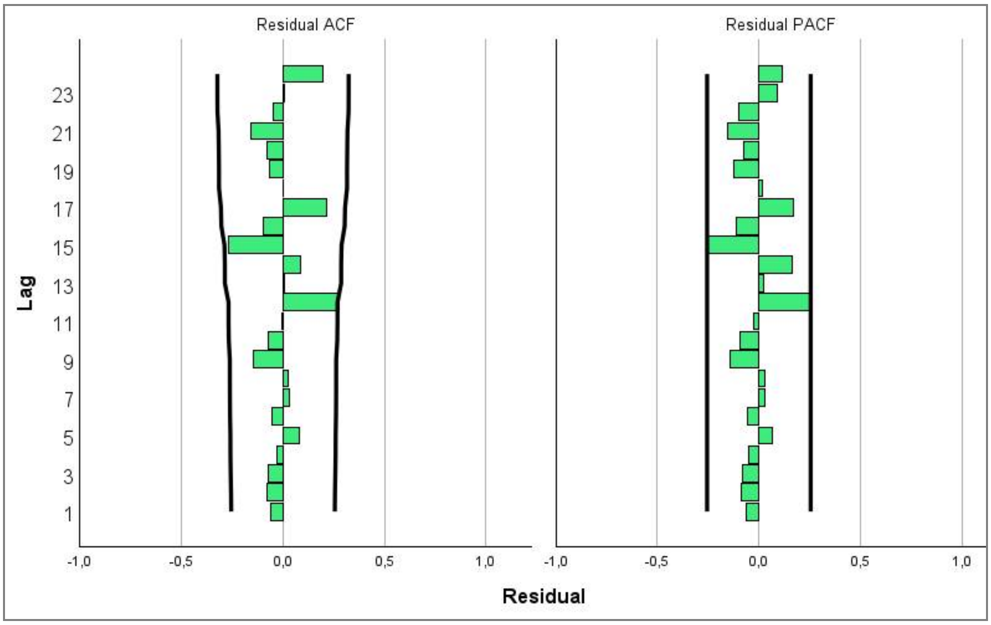
Source: own processing in SPSS

This is also confirmed by the correlograms from **Figure no. 10**, where the errors are within the range and do not differ significantly from zero. Basically, what remains in this range (residual values) represents the “white noise”, with mean 0 and variance 0, it’s that part that cannot be predicted and represents unknown factors.

According to the results presented in **Table no. 5**, we can also state that 59.6% of the variation in sales volumes is influenced by the variation in the time variable.

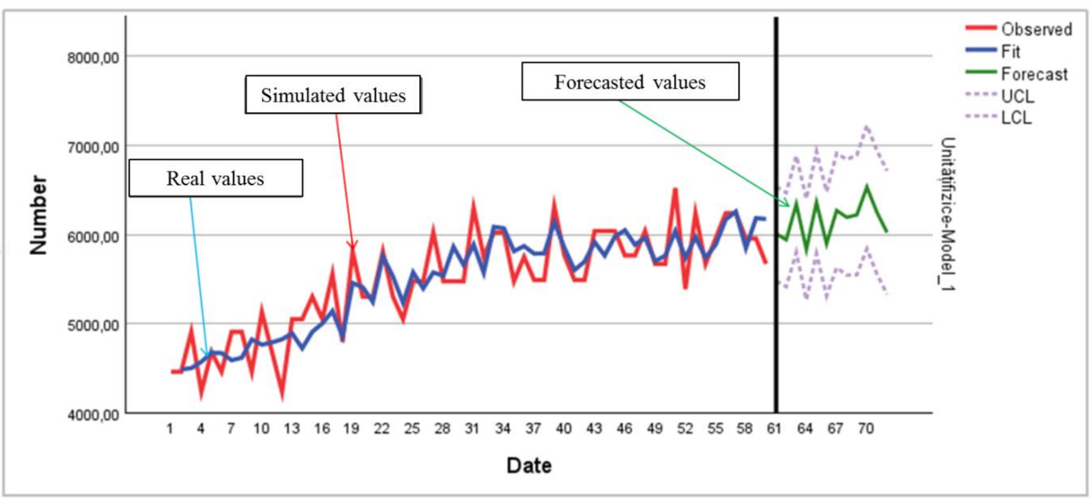
In **Figure no. 11** we observe the real values of the data series, the simulated values by the model as well as the predicted values which maintain an increasing trend.

Figure. no.10. The residual values of the model



Source: own processing in SPSS

Figure. no.11. The real, simulated and predicted values of the model



Source: own processing based on the graph obtained with SPSS

Table no. 6. Forecasted values for 2024 and the lower and upper limit values

YEAR_	MONTH_	DATE_	Predicted_Unitățifizice_Model_1	LCL_Unitățifizice_Model_1	UCL_Unitățifizice_Model_1
2024	1	JAN 2024	6010	5483	6538
2024	2	FEB 2024	5943	5416	6469
2024	3	MAR 2024	6342	5800	6883
2024	4	APR 2024	5835	5268	6401
2024	5	MAY 2024	6361	5788	6933
2024	6	JUN 2024	5898	5315	6481
2024	7	JUL 2024	6271	5631	6910
2024	8	AUG 2024	6194	5546	6842
2024	9	SEP 2024	6221	5550	6891
2024	10	OCT 2024	6532	5841	7224
2024	11	NOV 2024	6248	5557	6940
2024	12	DEC 2024	6024	5332	6715
TOTAL			**73.879**	**66.528**	**81.229**

Source: data obtained in SPSS

Table no. 7. Forecasted sales budget for 2024 (volumes and revenues)

2024 Budget	01	02	03	04	05	06	07	08	09	10	11	12	Total 2024
Estimated units	6.010	5.943	6.342	5.835	6.361	5.898	6.271	6.194	6.221	6.532	6.248	6.024	**73.879**
Forecasted income	2.939.812 €	2.907.039 €	3.102.211 €	2.854.211 €	3.111.505 €	2.885.027 €	3.067.481 €	3.029.817 €	3.043.024 €	3.195.150 €	3.056.231 €	2.946.661 €	**36.138.170 €**

Source: own processing in SPSS

In **Table no. 6** we can observe the values obtained for each month of the 2024 year, as well as the values within the confidence interval.

The volumes are then multiplied by the sales price and then the planned budget revenues were obtained (**Table no. 7**).

4. Conclusions

The budget is the fundamental short-term planning tool (usually 1 year), oriented towards a profitable management. The objectives in the budget must correspond to the strategy defined at the entity level translated into the plan (Caraiani C., 2010, p. 37).

The forecast is the starting point for budgeting. Thus, to budget means to foresee and decide, and to forecast means to know before acting (Badea F., Dobrin C., 2006, p. 71 and 81). As a result, in this study we set out to project the revenue budget based on the historical data from a company whose business activity is production for the time frame 2019-2023 and the predictions are being made for the year 2024.

How can we obtain the forecasts? With the help of the forecast models. The most well-known and used models are: the time series model, such as exponential smoothing and the ARIMA model, the cause-effect model based on the relationship between independent (driver) and dependent variables and finally the critical model used when we do not have previous data to work with.

To make the predictions, the ARIMA time series model was used and the first step consisted of transforming the data series from non-stationary to stationary. If we work with a non-stationary series, there is a risk of the phenomenon of artificial correlation between variables, means there is no concrete connection between them (Jemna D., 2009, p. 209-210).

Based on the autocorrelation and partial autocorrelation correlograms, potential ARIMA and SARIMA (seasonal ARIMA) models were identified, which were then tested (using SPSS v30) and validated according to several criteria, such as accuracy indicators (MAPE - mean absolute error), information criteria (BIC - Bayesian information criterion), Ljung-Box test on error stationarity (for sig<0.05) and analysis of the residuals correlogram (white noise). Finally, the ARIMA(2,1,0) model was validated, which predicted a total of 73,879 physical units for 2024.

The volumes were then multiplied by the sell price and after the planned revenues were obtained. This model shows us an increase in units compared to the period 2019-2023, a favorable sign for the company in terms of revenue growth. Of course, this also leads to higher production costs and the management must consider a control and administration of the direct costs.

The limitations of the study can be represented by the limited access to data and information regarding the budgeting practices and short-term objectives of the company. The restricted access to everything related to the strategic plan and strategies of the company for the following years could have contributed even better to the design of the revenue budget for 2024. Another limitation refers to the method, more specifically the independent variable is represented only by the time factor and other variables that may influence sales in the future were not included in the model.

In this paper, time series based on the ARIMA model were used to predict the revenue budget. However, forecasts can also be made using other statistical approaches, such as time series using a model called exponential smoothing (which aims to reduce the gap between forecast and demand in the event of unexpected changes), RNN (recurrent neural networks) or trend series modeling (predictions are made by using the regression analysis). If aspects regarding seasonality and the influence of exogenous factors can be incorporated into an ARIMA model, a model called SARIMAX can be developed. Moreover, the time variable can be replaced with other independent variables (both internal and external to the company) that can influence sales, such as consumer income levels, inflation levels, population structure, product prices on the market, advertising budget, information about the competition, etc.

References

1. Anghelache C., Manole A. (2012) Seriile dinamice / cronologice (de timp) - prezentare teoretică, structură, relațiile dintre indici, *Romanian Statistical Review* nr. 10
2. Badea F., Dobrin C. (2006) Gestiunea Bugetară a sistemelor de producție, *Ed. Economică*, București
3. Brooks C. (2008) Introductory Econometrics for Finance, second ediction, *Cambridge University Press*, New York
4. Budugan D. (2001) Contabilitate și Control de Gestiune, *Ed. Sedcom Libris*, Iași
5. Caraiani C., Dumitrana M., Dascălu C., Dumitru M., Glăvan M., Lungu C., Calu D., Jinga G., Gușe R., Jianu I., Vulpoi M. (2010) Control de Gestiune, *Ed. Universitară*, București
6. Cărbunaru B.A. (2013) Băcescu Condruz M., Metode utilizate în analiza variaţiilor sezoniere a seriilor cronologice, *Revista Română de Statistică*, nr. 3
7. Chorafas D. N. (2006) IFRS, Fair Value and Corporate Governance, *Ed. Elsevier*, Oxford
8. Devianto D., Permana D., Arif E., Afrimayani A., Yanuar Y., Maiyastri M., Yollanda *M.* (2024) An innovative model for capturing seasonal patterns of train passenger movement using exogenous variables and fuzzy time series, *Journal of Open Innovation: Technology, Market, and Complexity* , no.10
9. Firat M., Kaplan Y.D., Sampli R. (2021) Forecasting Air Travel Demand for Selected Destinations Using Machine Learning Methods, *Journal of Universal Computer Science*, vol. 27, no. 6 (2021), 564-581
10. Gupta A., Mehta V. (2023) Time Series based Predictive Model for Analysing the Health of Power transformer, *International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, Uttarakhand, India, 2023, pp. 1006-1013, doi: 10.1109/ICIDCA56705.2023.10099709
11. Guță A.J, Catrina I. (2020) Forecast activity in the context of the dynamics of economic and social processes and phenomena, *Annals of the University of Petroşani, Economics*, 20(1), 115-124
12. Ho J.S., Zhang Y. (2022) Forecasting Road Traffic Fatalities in Malaysia Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Model, *Journal of Science and Technology*, DOI: 10.47836/pjst.30.2.03

13. Horvath P., Binder B.C.K, Currle M., Esser J., Gerdes S., Graf J., Greiner O., Grunebaum D., Heinzelmann M., Hofmann., Hohner M.A., Kammler-Burrak A., Leyk J., Link M., Muller M., Sasse A., Scheffner J., Schmidt H. (2009) Controlling - Sisteme eficiente de creștere a performanței firmei, Ed. a II-a, *Ed. C.H. Beck*, București
14. Jain L.D.C, Malehorn J. (2006) Benchmarking: Forecasting Practices: A Guide to Improving Forecasting Performance*, Ed. Graceway Publishing Company, NW*
15. Jemna D. (2009) Econometrie*, Ed. Sedcom Libris*, Iași
16. Libby T., Lindsay R.M. (2010) Beyond Budgeting or Budgeting Reconsidered? A Survey of North-American Budgeting Practice*, Management Accounting Research*, 21, 56-75. http://dx.doi.org/10.1016/j.mar.2009.10.003
17. Nica P., Iftimescu A. (2008) Management. Concepte și Aplicații, *Ed. Sedcom Libris*, Iași
18. Noureen S., Atique S., Roy V., Bayne S. (2019) Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load, *IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, Dallas, TX, USA, pp. 521-524, doi: 10.1109/MWSCAS.2019.8885349
19. Rachlin R. (2007) Sistemul complet de bugete ale firmei*, Ed. BMT Publishing House*, București
20. Sagaert Y.R., Aghezzaf EL-H., Kourentzes N., Desmet B. (2017) Tactical sales forecasting using a very large set of macroeconomic indicators*, European Journal of Operational Research*
21. Seval E., Nursel O. (2017) Grey modelling based forecasting system for return flow of end-of-life vehicles*, Technological Forecasting & Social Change* 115, pp. 155–166
22. Sivanandam N. A., Ahrens D. (2015) A Hybrid Seasonal Autoregressive Integrated Moving Average and Quantile Regression for Daily Food Sales Forecasting, *Intern. Journal of Production Economics*, http://dx.doi.org/10.1016/j.ijpe.2015.09.039
23. Tripti D., Shamshad A., Mohammad S. (2020) Time series analysis of climate variables using seasonal ARIMA approach, *J Earth Syst Sci* 129, 149, https://doi.org/10.1007/s12040-020-01408-x
24. Zamfir M. (2017) Controlul de gestiune prin sistemul de bugete*, Ed. Pro Universitaria*, București
25. Zhao L., Mbachu J., Zhang H. (2019) Forecasting residential building costs in New Zealand using a univariate approach, *International Journal of Engineering Business Management*, Volume 11: 1–13, https://doi.org/10.1177/184797901988006