

Material demand forecasting with classical and fuzzy time series models

Sergey Zakrytnoy
Sievo Oy, Mikonkatu 15 A,
00100 Helsinki, Finland,
Email: sergey.zakrytnoy@sievo.com

Pasi Luukka
School of Business and
Management, LUT University,
Yliopistonkatu 34,
53851 Lappeenranta, Finland
Email: pasi.luukka@lut.fi

Jan Stoklasa
School of Business and
Management, LUT University,
Yliopistonkatu 34,
53851 Lappeenranta, Finland, and
Palacky University Olomouc,
Faculty of Arts, Department of
Economic and Managerial Studies
Email: jan.stoklasa@lut.fi
Email: jan.stoklasa@upol.cz

Abstract—Direct material budgeting is an essential part of financial planning processes. It often implies the need to predict quantities and prices of hundreds of thousands of materials to be purchased by an enterprise in the upcoming fiscal period. Distortion effects in demand projections and overall uncertainty cause the enterprises to rely on internal data to build their forecasts.

In this paper we are dealing with material demand forecasting and evaluate the feasibility of fuzzy time series forecasting models as compared to classical forecasting models. Relevant methods are shortlisted based on existing practice described in academic research. Three datasets from industry are used to evaluate the predictive performance of the shortlisted methods. Our findings show an improvement in prediction accuracy of up to 47% compared to naïve approach. Fuzzy time series models are reported to be the most reliable forecasting method for the analyzed intermittent time series in all three datasets.

I. INTRODUCTION

MODERN digitalization technologies and computational methods provide new levers for business decision-support impacting financial performance of an enterprise. Many of those levers are to be found (either to originate or to be applied) in supply chain management. Chopra and Meindl [1] state that the objective of a supply chain is the maximization of the overall generated value, where value is defined as the difference between sales revenue and total incurred costs throughout the chain of decision-making units. With shortened delivery timelines, those units are looking to introduce supply chain forecasting (SCF) models in order to meet customer's demand with the highest possible efficiency in terms of accuracy of the forecast and the work effort required for its generation. While this paper analyzes material forecasting from requirements planning perspective, downstream demand forecasting has recently been outlined as a symmetrically important business challenge with major impact on profitability of an enterprise [2]. To find a suitable approach to upstream SCF, the physiology of a supply chain should be considered from three different perspectives: length, depth and time.

When trying to quantify and forecast upstream demand propagation, it is important to recognize the complexity of the supply chain and different factors that may influence or distort the projections. Lee et al. [3] defined Bullwhip Effect as the amplification of demand variance that takes place as the value proceeds through the chain nodes. Main reasons for this are operational inefficiencies and external factors that affect the deviation between expected and realized demand quantities.

It was noted by Chopra and Meindl [1] that one way to handle incomplete information, its distortion effect on demand projections and operational inefficiency of manufacturers, would be the development of collaborative concepts where information is shared between supply chain entities. The main concepts that were proposed are collaborative forecasting and replenishment (CFAR) systems where interchange of decision-support models and strategies to facilitate forecasting processes is suggested [4]. Other concepts that have emerged include Collaborative planning, forecasting and replenishment (CPFR), Vendor Management Inventory (VMI) and other information systems [5]. While unintuitive, it was shown that collaborative supply chain forecasting can yield negative dynamics in the performance, widening the Bullwhip Effect and burdening the procurement function [6], [7].

Since collaborative forecasting mechanisms prove to be ineffective both in terms of accuracy and incurred workload, it is becoming increasingly relevant to explore possibilities for the autonomous forecasting of demand. This research is based on anonymized historical purchasing data from several industry partners operating globally. In terms of the length of a supply chain, the dataset provides full visibility to the first-tier suppliers of different products, while lacking an extended view to adjacent nodes of the supply chain, which represents a typical setup for developing SCF process as a business application. The main objective of this research is to evaluate and compare the performance of different forecasting

This work was supported by Sievo Oy. The authors would also like to acknowledge the support of this publication by LUT research platform AMBI- Analytics-based management for business and manufacturing industry.

methods in the SCF domain and, if possible, to identify methods of choice for material demand forecasting.

II. METHODOLOGICAL BACKGROUND

Time series forecasting, i.e. prediction of future or missing entries in a series of numerical values indexed in time order [8], is a broad research domain which is historically relevant for multiple application areas, incl. natural sciences, industrial engineering, economy, business and many others. Time series forecasting can be divided into three main methodological types [9]. These are 1) Explanatory models where the dependent variable is represented as a function of external factors (regressors or independent variables) and a causal relationship is assumed (or at least the ability to compute the values of the regressand from the values of the regressors) for modelling by fitting the function to existing data 2) Autoregressive models where the forecast is generated based on historical values of time series without external variables 3) Mixed models, which contain explanatory and dynamic components, that include dynamic regressions, transfer function models, linear systems, vector alternatives of the above mentioned models, machine learning models etc. In this paper, autoregressive time series models are used due to a lack of numerical data points available in an independent enterprise SCF concerning additional explanatory variables. We selected three different model types to be fitted to the data and also considered a naïve benchmark model to be able to compare the performance of the selected models with a reference model. Given the type of the time series being forecasted, only models capable of reflecting seasonality in the time series were considered.

A. Naïve benchmark

The naïve forecasting method is the basic estimation technique in which time series value from the last period is taken as the forecast for the next one, without attempting to adjust it or establish causal factors. It is represented as

$$y_{t+1} = y_t \quad (1)$$

where y_t is time series in question and it is assumed that at time t we need to make a forecast of the value of the time series for times $t + m$, $m \in \mathbb{N}$, $m > 0$. In other words we assume that the historical values of the time series being forecasted are known including the current value of the series, but no information is available after time t . Predicting the last known value, that is $y_{t+m} = y_t$ for all $m > 0$, is one of the most commonly used benchmark methods due to its simplicity.

B. Holt-Winters exponential smoothing

The exponential smoothing models were proposed as forecast generators through weighted average of previous observations while weights decrease exponentially over time periods (more historical values influence the forecast less than more recent ones).

In Holt-Winters (HW) seasonal method [10]–[12] the time series are decomposed, and the series estimation formula is split into three equations: level, trend and seasonality. All of them consider different smoothing coefficients and comprise a system of simultaneous equations as follows:

$$\begin{cases} S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \\ b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \\ I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L} \\ y_{t+m} = (S_t + mb_t)I_{t-L+m} \end{cases} \quad (2)$$

where y_t is observation of the series, S_t is the smoothed observation, b_t is the trend factor, I_t is the seasonal index, y_{t+m} is the forecast at m periods ahead; α, β and γ are smoothing parameters that are estimated so as to minimize the fitting error. The baseline value for trend can be computed as

$$b_0 = \frac{1}{L} \left(\frac{y_{L+1} - y_1}{L} + \frac{y_{L+2} - y_2}{L} + \dots + \frac{y_{L+L} - y_L}{L} \right) \quad (3)$$

where L is the length of the season, y_t are observation series, while the initial season factor is calculated as

$$I_0 = \frac{\sum_{p=\epsilon}^N \frac{y_{t+pL}}{A_p}}{N} \quad (4)$$

where t is the time period, N is the number of complete seasons we have the data for, y_t are observation series and $A_p = \frac{\sum_{i=1}^L y_i}{L}$, $p = 1, 2, \dots, N$.

C. Seasonal Autoregressive Moving Average

Autoregressive Moving Average (ARMA) model family consist of autoregressive (AR) and stochastic (MA) components [13]. Autoregressive components reflect the dynamic structure of the series describing its linear relation to order p while the moving average component is a linear combination of q lags of the error term. Alongside with exponential moving average models, they are commonly used in SCF for benchmarking purposes [2].

ARMA models are formulated as follows and they require the time series to be weakly stationary.

$$y_t = C + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (5)$$

where y_t is the estimated series, C is the constant term, φ_i is the coefficient for the autoregressive component of order i , θ_j is the coefficient for the moving average component of order j , and ε_t is the error term.

Seasonal autoregressive integrated moving average (SARIMA) model is an extension of the traditional integrated ARMA activating the pattern recognition potential through a set of new parameters: seasonal autoregressive component (P), seasonal integration (D) and seasonal moving average (Q). These parameters are combined (the order of seasonal integration being set to $D = 0$) in the following equation:

$$y_t = C + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{k=1}^p \gamma_k y_{t-kL} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{r=1}^Q \mu_r \varepsilon_{t-rL} \quad (6)$$

where, in addition to the terms from (5), we introduce γ_k and μ_r as seasonal parameters to be estimated with the length of seasonal period L .

D. Fuzzy time series model

Fuzzy time series (FTS) is a concept from the fuzzy data analysis domain, which is based on the fundamental concept of a fuzzy set, introduced by Zadeh [14]. A fuzzy set is a flexible way to model uncertainty through assigning a gradual membership value $\mu_A(x) \in [0,1], x \in U$ to a specified set A for every element x of a universe of discourse U , instead of quantifying phenomena with a single crisp value from the set $\{0,1\}$.

In 1993 Song and Chissom [15] introduced fuzzy time series $F(t)$ on the subset of real numbers $Y(t)$ ($t = 0,1,2,\dots$). A fuzzy time series $F(t)$ is a collection of fuzzy sets A_t ($t = 1,2,\dots$) with membership functions $\mu_{A_t}(x)$ ($t = 1,2,\dots,n \in Y_t$). The real time series can be transformed into their fuzzy representation with the appropriate membership function, universe of discourse, and assigning membership degree values for real numbers in question.

The fuzzy time series forecasting models rely on the notion of fuzzy logical relationships (FLR). If A_i and A_j denote the fuzzy sets that form part of fuzzy time series $F(t)$, the logical relationship can be expressed with notation $A_i \rightarrow A_j$ (FTS model of order 1) or $[A_i, A_k] \rightarrow A_j$ (high-order FTS model with 2 lags). In the examples above, A_i and $[A_i, A_k]$ are called left-hand side (LHS) of an FLR, while A_j is its right-hand side (RHS).

The FLRs observed from historical data can be organized into fuzzy logical relationship groups (FLRGs). They comprise the knowledge- or rule base that is further inferred to generate forecast for future or missing values.

A simple FTS model generates forecast based on the following algorithm; let $F(t) = A_i$, then

- if $A_i \rightarrow \emptyset$, that is if there is no rule in the FLRG with A_i as LHS, then $F(t+1) = A_i$ and the defuzzified forecast $Y(t+1)$ is the midpoint of A_i , if defuzzification is needed;
- if $A_i \rightarrow A_j \in \text{FLRG}$, then $F(t+1) = A_j$, $Y(t+1)$ being the midpoint of A_j ;
- if $A_i \rightarrow A_{j_1}, A_{j_2}, \dots, A_{j_k} \in \text{FLRG}$, there is no single fuzzy representation of $F(t+1)$, there are more possible fuzzy-set outputs, and the defuzzified value, if needed, is derived directly as the arithmetic average of the midpoints of $A_{j_1}, A_{j_2}, \dots, A_{j_k}$.

Weighted FTS (WFTS) is a model type that handles the scenario of $A_i \rightarrow A_{j_1}, A_{j_2}, \dots, A_{j_k}$ in a different way. The defuzzification of the forecast is then calculated as

$$Y(t+1) = \sum_{j \in \text{RHS}} w_j c_j \quad (7)$$

with

$$w_j = \frac{\#A_j}{\#RHS} \quad \forall A_j \in \text{RHS} \quad (8)$$

where $\#A_j$ is the number of occurrences of A_j in FLRs with the same LHS and $\#RHS$ is the total number of temporal patterns within that FLRG and c_j is j^{th} midpoint [16].

Probabilistic Weighted FTS (PWFTS) incorporate information about membership degrees of the LHSs of the FLRs. The knowledge base for PWFTS is given as

$$\begin{aligned} \pi_1 A_1 \rightarrow w_{11} A_1, \dots, w_{1i} A_i \\ \pi_i A_i \rightarrow w_{i1} A_1, \dots, w_{ii} A_i \end{aligned} \quad (9)$$

where each weight π_i is the normalized sum of all LHS values of membership functions where the LHS is fuzzy set A_i [17]. Thus, π_i can be interpreted as the empirical a priori probability of having A_i as an LHS. The weight w_{ij} is the normalized sum of all RHS memberships where LHS is A_i and RHS is A_j , which can be understood as a conditional probability $P(F(t+1) = A_j | F(t) = A_i)$.

The forecasting procedure in PWFTS starts with the computation of probability distribution

$$\begin{aligned} P(Y(t) | Y(t-1)) = \\ \sum_{A_j \in \tilde{A}} \frac{P(Y(t) | A_j) * \sum_{i=1}^k P(Y(t+1) | A_i, A_j)}{\sum_{i=1}^k P(Y(t) | A_i)} = \\ \sum_{A_j \in \tilde{A}} \frac{\pi_j \frac{\mu_{A_j}(Y(t))}{Z_{A_j}} * \sum_{i=1}^k w_{ij} \frac{\mu_{A_i}(Y(t+1))}{Z_{A_i}}}{\sum_{i=1}^k \pi_i \frac{\mu_{A_i}(Y(t))}{Z_{A_i}}} \end{aligned} \quad (10)$$

where, in addition to previous notations, $\mu_A(Y)$ is degree of membership of continuous value Y to a fuzzy set A , Z_A is the total area under membership function of A , and \tilde{A} is the set of all fuzzy sets considered on the given universe, for example \tilde{A} can be a set of the fuzzy-set meanings of the linguistic values of a linguistic variable used to describe the values of the time series to be forecasted. The point forecast is then produced by

$$Y(t+1) = \sum_{A_j \in \tilde{A}} \frac{P(Y(t) | A_j) * E[A_j]}{\sum_{A_j \in \tilde{A}} P(Y(t) | A_j)} \quad (11)$$

with $E[A_j] = \sum_{i \in A_j^{\text{RHS}}} w_{ij} * mp_i$, mp denoting a midpoint of a fuzzy set A_j .

FTS represent a real alternative to the traditional econometrics methods since fuzzification of original time series makes the stationarity requirement redundant and reduces the allowed value domain to a finite number of fuzzy sets (or linguistic values the meanings of which are

represented as fuzzy sets) which works as an embedded normalization technique.

III. DATA EXTRACTION AND PREPROCESSING

Extraction and preprocessing of the data included such subtasks as 1) selection of appropriate time period, 2) temporal aggregation, 3) scoping (reducing dimensionality – due to computational reasons) the list of time series included in the analysis and 4) handling outliers.

For the time period selection, three main criteria were considered: availability of data on codified direct purchases, potential to reveal annual seasonality and relevance for the business. Based on those criteria, the timeframe for the transactional dataset was set to January 2016 – November 2020, the latter being the most recent reported period in the source data. The time step was one calendar month. Cross-sectional aggregation was performed on a product-location level, which means that each time series represents monthly values of purchased quantities of a product by a given operating unit.

Pareto principle also known as the “80-20” rule was considered to narrow down the focus of the quantitative research. The dominating share of spend originated from a relatively low number of biggest purchase items, hence the scope of research could be limited to comply with limitations of available computing resource. Depending on the industry partner, from 1.34% to 3.70% of available time series were such that they added up to 90% of cumulative spend, and thus were included to the research scope. Further filtering of the data is described assuming that 100% of original time series represent the reduced number.

The underlying products of purchased quantity time series need to remain relevant to the business. We therefore only included the time series that contained non-zero values of quantity and spend in the last 12 months of the recorded period. Across the three partner datasets, 77.31-90.41% of time series fulfilled the requirements.

In order to ensure availability of sufficient training data, we removed the series where the period between the earliest and the most recent observation was under 3 years. In the study, 18.30-53.96% of the time series have enough observations. Combined with the previous filtering criterion, there is an overall acceptance rate of 17.99-50.99% of the initial number of time series for the subsequent analysis and experiments.

IV. DESIGN OF EXPERIMENTS

In this part, we describe the experiments conducted to evaluate the quality of the selected time series forecasting methods.

Performance measurement

For performance measure Root Mean Squared Error (RMSE) was selected. RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{pred_i} - y_i)^2}{n}} \quad (12)$$

where y_{pred_i} is the forecasted value of the series for time i , and y_i is the corresponding original value for all the investigated values of i .

Training and testing data

An appropriate representation of data is essential in a quantitative study. Time series data are commonly split over temporal indices to ensure original order.

First, we specify the forecast horizon i.e. the number of future observations that we want to generate as a model output. The business needs dictate that budgetary revisions are performed on a quarterly basis; thus the forecast horizon is specified to 3 months.

In order to avoid potential bias related to seasonality or coincidence in externalities, multiple testing windows are included in the analysis as per availability and volume of original data. An expanded rolling window approach is adopted, meaning a gradual increase in the number of observations in the training dataset, shifting the index of the testing period in a way that provides additional dimension to the analysis of results by revealing the sensitivity of algorithms to the amount of training data. The resulting cuts of the original time series range between 36 and 58 observations in length; we characterize the amount of training data as “low” if it represents a time period of less than 4 full years, and “high” otherwise.

All things considered, the experiment for each series is carried out in the following steps:

1. Identify the first and last period with non-zero normalized quantity values and remove leading and lagging null observations;

TABLE I.
COMPARISON OF PERFORMANCE OF DIFFERENT METHODS

	HW RMSE		SARIMA RMSE		FTS RMSE		Naive RMSE	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Company A	0.163	0.075	0.178	0.076	0.121	0.055	0.199	0.101
Company B	0.165	0.084	0.179	0.087	0.125	0.069	0.219	0.123
Company C	0.210	0.107	0.226	0.109	0.156	0.088	0.295	0.153

2. Split the resulting series into $n_{windows}$ expanding windows, starting with the first $33+3=36$ months of data (33 observations for training and 3 for testing purposes) and incrementing the index of last observation included in the sample by $\lceil (i_{max} - 36)/n_{windows} \rceil$ where i_{max} is the largest integer index of the series (starting with 1, equal to number of observations) and $n_{windows}$ is the target number of windows per series;
3. Run all configurations of each model family (Exponential Smoothing, SARIMA or FTS) on each of the windows and store the results in such a format that it would include full information regarding the tested series, values of hyperparameters and observed RMSE. Apply the naïve forecast for benchmarking purposes.

V. RESULTS

In Table I, the methods are evaluated based on average RMSE error term and its standard deviation across time series. There is a visible improvement in prediction accuracy of all three methods compared to the naïve solution.

The improvement is further validated with a visualization of RMSE in form of histograms (Figure 1), one per each dataset.

If we compare individual performance of the models on time series level, we see that FTS would be the best choice in 7947 cases representing almost 60% of the total count,

followed by HW and Naïve benchmark with 2642 and 1624 respectively while SARIMA would only be optimal in 1099 cases. Looking at detailed specifications of the respective models, it is notable that in majority of cases SARIMA becomes a simple arithmetic average, serving as additional benchmark solution.

FTS shows a higher and more stable prediction accuracy which may be explained by its ability to handle the intermittency by fuzzification of original series whereas other methods operate on a continuous scale. Zero values alternating with non-zero ones are translated into a discrete number of fuzzy sets, which reduces the noise in identifying sequential patterns.

VI. CONCLUSION

In scope of this research, we have tested such time series forecasting methods as Holt-Winters exponential smoothing, SARIMA and Fuzzy Time Series forecasting models, on three independent datasets containing historical direct material purchasing data of industry partners. The results reveal that using the Fuzzy Time Series approach, there is a potential to reveal hidden intrinsic and seasonal patterns and achieve a substantial improvement in accuracy compared to simple statistical forecast.

Fuzzy Time Series models showed the best performance in all datasets because of their ability to reduce the noise caused by intermittency of the original series. Holt-Winters is a viable alternative showing stable improvement to the error

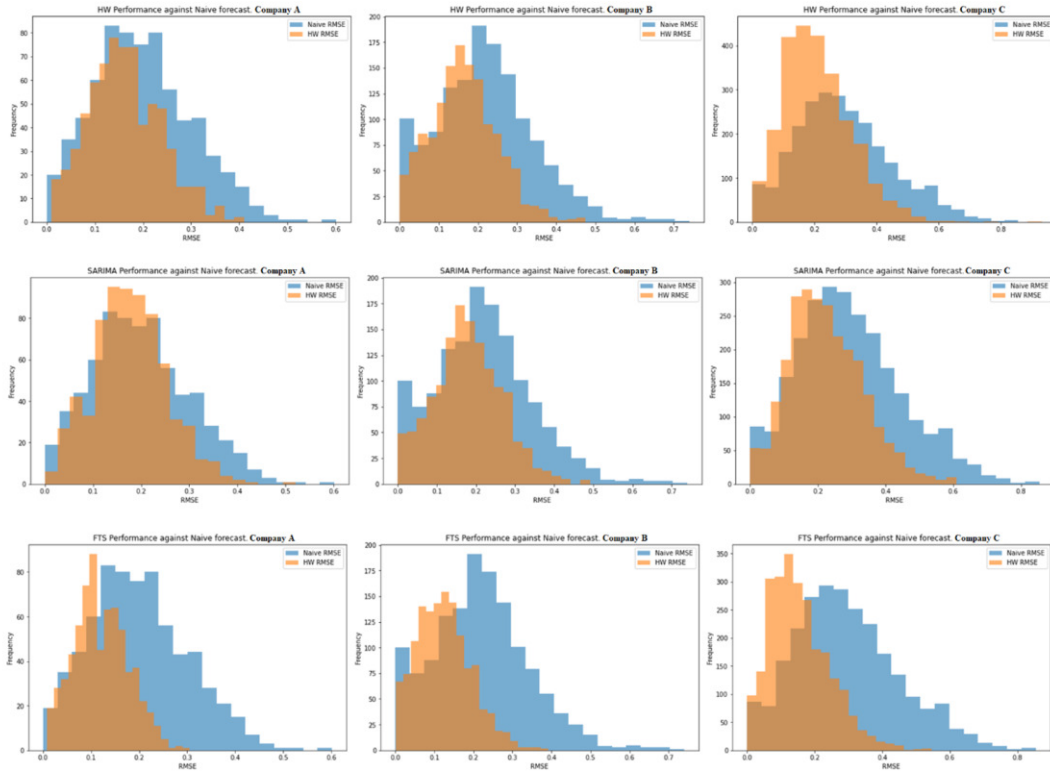


Figure 1. RMSE histogram of HW (row 1), SARIMA (row 2) and FTS (row 3) against benchmark

metrics whereas SARIMA is not recommended in this case due to insufficient amount of training data and its intermittent nature which results in a notable underperformance. Fuzzy Time Series and Holt-Winters can be used to generate automatic forecasts of direct material purchases when the amount of historical data is sufficient.

ACKNOWLEDGMENT

The paper represents a processed summary of the research performed in scope of a Master thesis [18].

REFERENCES

- [1] G Chopra, S., Meindl, P. Supply chain management: strategy, planning and operation. 5th edition. USA New Jersey: Pearson, 2012
- [2] Ganzha, M., Maciaszek, L., Paprzycki, M. & Ślęzak, D. Impact of time series clustering on fuel sales prediction results. Position and Communication Papers of the 16th Conference on Computer Science and Intelligence Systems, ACSIS, Vol. 26, pages 13–21. 2021
- [3] Lee, H. L., Padmanabhan, V., & Whang, S. Information distortion in a supply chain: the bullwhip effect. *Management Science*, 43, 546–558, 1997
- [4] Raghunathan, Srinivasan. Interorganizational Collaborative Forecasting and Replenishment Systems and Supply Chain Implications. *Decision Sciences*. 30. 1053 - 1071, 2007
- [5] Synthetos, A. A., Kholidasari, I., & Naim, M. The effects of integrating management judgement into OUT levels: in or out of context? *European Journal of Operational Research*, 2015
- [6] Thonemann, U.W. Improving supply-chain performance by sharing advance demand information. *European Journal of Operational Research* 142-1, 81–107, 2002
- [7] Heikkilä, J., From supply to demand chain management: Efficiency and customer satisfaction. *Journal of Operations Management* 20 (6), 747–767, 2002
- [8] Box, G. & Jenkins, G., *Time Series Analysis: forecasting and control*, Oakland, California: Holden-Day, 1976
- [9] Hyndman, R.J. & Athanasopoulos, G. *Forecasting: principles and practice*, OTexts: Melbourne, Australia, 2nd Edition, 2018
- [10] Brown, R. G. *Statistical forecasting for inventory control*. McGraw/Hill, 1959
- [11] Holt, C. E. Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA, 1957
- [12] Winters, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342, 1960
- [13] Mills, T. *Time Series Techniques for Economists*. Cambridge University Press, 1990
- [14] Zadeh, L. Fuzzy sets. *Information and Control*, 8, 338–353, 1965
- [15] Song, O., Chissom, B. Fuzzy time series and its model. *Fuzzy Sets and Systems*. 54. 269–277. 1993
- [16] Ortiz-Arroyo, D., Poulsen, J. R. A Weighted Fuzzy Time Series Forecasting Model. *Indian Journal of Science and Technology*, 11(27), 1-11. 2018
- [17] Silva, P. Scalable Models for Probabilistic Forecasting with Fuzzy Time Series, Thesis for: Ph.D. 2019
- [18] Zakrytnoy, S. Comparative study of classic and fuzzy time series models for direct materials demand forecasting. Thesis for: MSc. 2021