Zeszyty Naukowe Politechniki Częstochowskiej
Zarządzanie Nr 38 (2020) s. 60-66
dostępne na: http://www.wz.pcz.pl/znwz

UTILITY OF FORECASTING ALGORITHMS IN THE CASE OF SELECTED DISTRIBUTION NETWORKS

Mariusz Kmiecik[1]

Silesian University of Technology
Faculty of Organization and Management

Abstract: The research paper focuses on verifying the use of forecasting algorithms which are available in the library(forecast) in R programming environment. Examination of the algorithms takes place in five selected distribution networks. Each of the algorithms is compared via MAE as a measure of model fitting to time series. The main goal of present article is to show the use of particular forecasting methods based on different indicators like the coefficient of variation. The added value of following paper is that it shows the most fitted algorithms in the selected variability of time series. It allows the algorithms to be adjusted to particular time series basing on their variance level.

Keywords: distribution network, forecasting, R software

DOI: 10.17512/znpcz.2020.2.05

Introduction

The aim of cooperation in distribution networks is to provide customers the required goods and services. Enterprises are striving to integrate their activities and are trying to predict the levels of future demand. Demand fluctuation is one of the factors of distribution network development. Demand management is a complex issue. One of the elements of demand management is a properly projected and implemented forecasting algorithm. Nowadays, there are several known methods and ways to make forecasts.

The article focuses on testing the possible ways of forecasting using R environment. The main goal of the present paper is to find the forecasting algorithms best fitted to the time series. The studied forecasting algorithms are available in R programming environment. Fitting of algorithms is based on ex-post errors of forecasting results in examined periods. The study will allow the best forecasting solutions to different structures of time series to be found.

The aim of following paper is to find the algorithms fitted to different SKU (Stock Keeping Units). The main variation feature of the mentioned SKU in this paper is the time series variability indicator. It will allow better planning and better fulfilling of operations connected with knowledge about future SKU release levels. The paper's hypothesis was established as: there are some algorithms connected with automatic forecast calculation in R environment which are fit to time series and create the most appropriate forecasts, mainly in identified variation levels.

[1] Mariusz Kmiecik, MA Eng., mariusz.kmiecik@polsl.pl, ORCID: 0000-0003-2015-1132

Chosen forecasting algorithms from library(forecast) in R environment

The algorithms used in the current paper are the algorithms available in R environment. *Table 1* shows the chosen set of algorithms used in the created forecasting tool, whose results are describing in present paper.

Table 1. Brief description of chosen forecasting functions

Function	Brief function description
ses()	Forecasting stationary time series connected with simple exponential smoothing methods.
holt()	Forecasting time series with trend using Holt Method.
holtWinters()	Forecasting time series with trend and seasonality using Holt-Winters Method. In its default form, the HoltWinters()function is constructed for additive seasonality. To use the multiplicative version of the method the function needs to be modified as follows: HoltWinters(x, seasonal = c("multiplicative"), where x is the time series.
ets()	Forecasting based on three dimensions: error (E), trend (T) and seasonality (S). Function gives possibility to determine mentioned parameters. In whole cases: N – none, A – additive, M – multiplicative and Z – automatically selected. In default function configuration whole parameters are automatically selected and default function is equivalent to ets(x, model = "ZZZ"), where x is the time series.
ar()	Forecasting based on fitting time series autoregression model to input data.
arima()	Forecasting based on fitting ARIMA(p,q,d) model to time series. Focus on connection between autoregression process (AR(p)) to moving average process (MA(q)).
auto.arima()	Forecasting based on ARIMA(p,q,d) model, but additionally taking into account information criteria such as AIC (Akaike Information Criterion), AICc(Corrected Akaike Information Criterion) and BIC(Bayesian Information Criterion).
arfima()	Forecasting based on autoregressive fractionally integrated moving average model (ARFIMA(p,d,q)) with two-step procedure where parameters p, d and q are determined separately. Parameters of autoregression (p) and moving average (q) are determined by Hyndman-Khandakara algorithm. Integration level (d) is determined by Haslett and Raftery algorithm.
tbats()	Forecasting based on exponential smoothing with Box-Cox transformation (tb), ARMA errors (a) as well as trend (t) and seasonal (s) components.
splinef()	Forecasting based on cubic smoothing splines equivalent of ARIMA(0,2,2) model but with some parameter restrictions.
stlf()	Forecasting based on time series decomposition using local regression. Function uses Seasonal Decomposition of Time Series by LOESS (developed by Helsel and Hirsch in 1997).
meanf()	Forecasting based on assumption that random component is independent and identically distributed to whole time series.
rwf()	Forecasting based on random walk with drift model.
snaive()	Forecasting based on middle of range which contains most observations. It refers to one of direct modal estimators called Chernoff's estimator.
nnetar()	Forecasting based on neural network (neural network time series forecasting). In function default settings forecasting demand is based on simple, feedforwarded neural network with one hidden layer.

Source: (Burnham, Anderson 2004; Chernoff 1964; Haslett, Raftery 1989; Hyndman, Akram, Archibald 2008; Hyndman, Khandakar 2008; Hyndman et al. 2002; Rogalska 2016; Zagdański, Suchwałko 2016)

Using the described functions and obtaining a demand forecast requires employing a function forecast().

Nowadays distribution networks are susceptible to disruptions. To minimize this threat, resilient logistics systems should be built. These kinds of actions could include, for example, improving supply and warehousing areas such as eliminating the lack of SKU in distribution networks (Kramarz, Kmiecik 2017). Some of the disruptions could be reduced, for instance, by introducing a proper forecasting system. Forecasts are considered as one of the risk categories in material flows (Kramarz 2013). The risk factors connected with forecasting are: imprecision, seasonality, product differentiation, short product life cycle, a too small customer data base and information deviation. An entity could make plans for the required fleet capacity, human resources and also create a schedule adjusted to transport of the forecasted goods.

Forecasting has special importance as one of the means used in enterprise management processes because incorrect prediction of future trends could imply devastating consequences for whole company. The request for forecasts in an enterprise is formed by 2 basic reasons (Dittmann 2000): the lack of certainty regarding the future, and the delay between the moment of the decision and its effects. uncertainty regarding the future, and the delay between the moment of the decision is made and its effect.

Research methodology

The research methodology focuses on analysing the created forecasting tool in five distribution networks. A common issue in whole networks is the presence of a logistics operator. A logistic operator provides logistics services to the manufacturer. The author proposed a demand forecasting tool which can be implemented in the activities of a logistics operator. According to the main goal, the algorithms are analysed in two directions. First, analysis of the best fitting algorithms to time series, and second, analysis of the best fitting algorithms in different levels of time series variability.

First of all, the author chose five distribution networks. The brief characteristics of the networks are shown in *Table 2*.

Table 2. Brief characteristics of distribution networks

Number and characteristics of distribution networks	Number of SKU (Stock Keeping Units)	Number of assortment groups	Number of variability groups
1. Distribution network of household chemicals and cosmetics	1 362	19	15
2. Distribution network of food products	1 152	15	15
3. Distribution network of construction--related products	415	12	13
4. Distribution network of sweets	60	5	12
5. Distribution network of household chemicals	272	8	8

Source: Own elaboration

The number of variability groups was calculated using distributive series. The number of sections was calculated as follows: $k \leq 5 \times \text{logn}(n)$, where n – number of SKU. A particular SKU was fitted to a particular section based on variability indicator values. The variability indicator describes the variability of the daily release level according to two years of time series data.

A large number of SKU occurred in a particular distribution network and the operational requirements indicate the necessity to create a demand forecasting tool which is able to automatically create a daily forecast. The created tool divides time series into two parts – train and test parts. The train part is a part of time series to find the best fitted algorithm and the test part is to test the algorithm during the established period. The train part consists of time series of daily release data from two years. The duration of the test part is established at the level of 14 days. To create a forecast tool, data is imported from the WMS (Warehouse Management System) of the logistics operator to the R script. Next, the time series are analysed in a loop. Analysis in the loop focuses on reducing outliers (using tsclean()), dividing the time series into the train and test part and creating a forecast using the 15 described algorithms as well as calculating MAE (Mean absolute error) for all the algorithms and choosing the best of them.

Results

According to the smallest MAE, the best fitted algorithms are the algorithms shown in *Figure 1*.

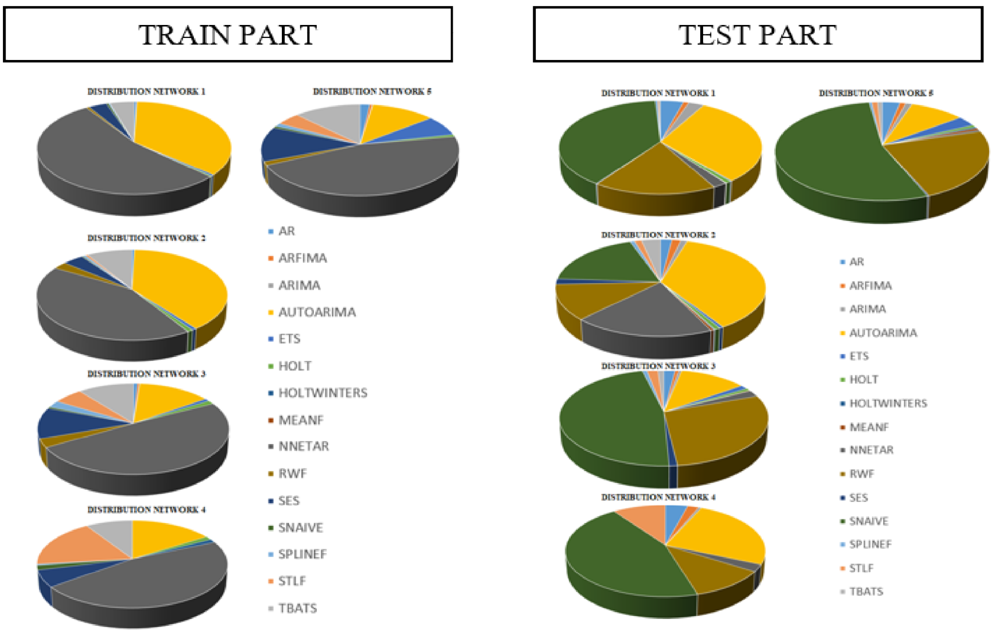

Figure 1. Best fitted algorithms to particular time series parts

Source: Own elaboration

The forecasting tool was tested during two months of activity with weekly data and forecast updates. In each of the distribution networks the results of the best fitted algorithms was similar. In the training set the best fitting algorithm was nnetar(), but the best fitting algorithm in the test sets was snaive(). It could lead to the conclusion that the considered time series are hardly predictable using only the best fitted model to time series structure. Although the algorithm with the minimal value of MAE in the training set in the majority was the algorithm based on a neural network, the author recommends using mainly the results from the algorithm based on Chernoff's estimator (snaive()), which showed the smallest forecast errors in the testing part.

Analysis of the time series was also characterized by a relatively high level of variability. The level of variability was examined using the indicator for the whole time series. The adjustments to particular variability groups is shown in *Figure 2*.

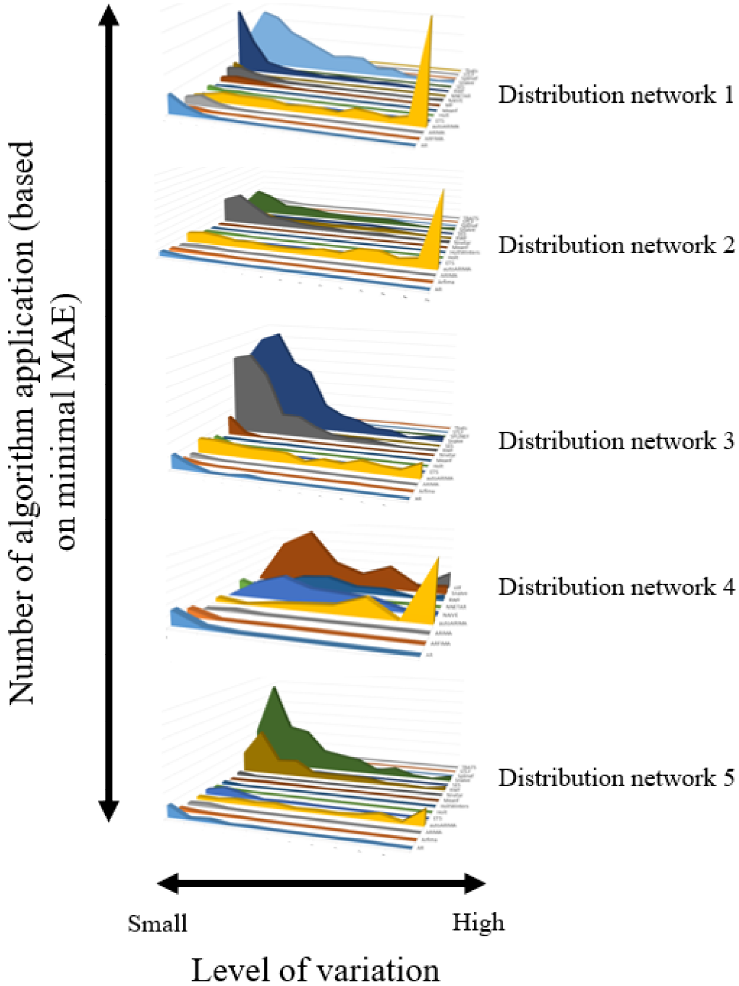

Figure 2. Best fitted algorithms to particular time series parts

Source: Own elaboration

Figure 2 shows the use of particular forecasting algorithms based on the test set of time series in the discussed distribution networks. The figure takes into account only the results for the test part of particular time series. The best fitted algorithms in extreme cases (small and high level of variation) are as follows:

- Distribution network 1 – household chemicals and cosmetics. For a small level of variation the best forecasting algorithms are rwf() and snaive(). For a high level of variation – auto.arima().
- Distribution network 2 – food products. The best forecasting methods for a small level of variation are nnetar() and snaive() and for a high level of variation auto.arima().
- Distribution network 3 – construction-related products. The forecasting methods the best fitted to a small level of variation are snaive() and rwf() and to a high level are auto.arima() and snaive().
- Distribution network 4 – sweets. For a small level of variation, the best forecasting algorithms are ar() and nnetar(). Calculation based on snaive() is the most common in the rest of the time series. For a high level of variation – auto.arima().
- Distribution network 5 – household chemicals. The best forecasting methods for a small level of variation are snaive() and rwf() and for a high level of variation auto.arima().

It could lead to the conclusion that to construct a proper forecast for time series characterized by a small level of variation, using algorithms based on Chernoff's estimator (snaive()) or on random walk with the drift model (rwf()) is recommended. The best forecasting method for the time series with a high level of variation is the method based on the ARIMA (p,q,d) model with information criteria (auto.arima()). It confirms the hypothesis set out in the introduction of current paper. In the future, this will allow the use of personalized algorithms based on the coefficient of variation for each time series.

Conclusions

The conducted research allowed the best fitted algorithms to time series in particular ranges of variability to be found. I In the future, this will allow the use of personalized algorithms based on the variability indicator for each time series. It could help with assortment management in distribution networks and to meet customer requirements. The main goal of the research was fulfilled, therefore the present paper, based on research in selected distribution networks, could help in future decisions connected with which kind of algorithm should be applied to different SKU. It could help managers to find the fastest way to create proper forecasts and create proper operational plans based on forecasts. The current paper also confirms the hypothesis. There are some algorithms which exhibit very large adjustment to individual ranges of variation (different types of SKU). Of course, the presented research was conducted only in five distribution networks with data from the same points in the distribution network (from the material decoupling point). The presented results should also be examined in different nodes. In future, research could be extended by examination of the possibility of modifying algorithms . It could verify the influence of changing the default parameters on forecast accuracy and search for the best solution.

References

1. Burnham K.P., Anderson D.R. (2004), *Multimodel Inference – Understanding AIC and BIC in Model Selection*, “Sociological Methods & Research”, Vol. 33.
2. Chernoff H. (1964), *Estimation of the Mode*, “Annals of the Institute of Statistical Mathematics”, Vol. 16.
3. Dittmann P. (2000), *Metody prognozowania sprzedaży w przedsiębiorstwie*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
4. Haslett J., Raftery A.E. (1989), *Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource*, “Applied Statistics”, Vol. 38.
5. Hyndman R.J., Akram M., Archibald B.C. (2008), *The Admissible Parameter Space for Exponential Smoothing Models*, “Annals of the Institute of Statistical Mathematics”, Vol. 60.
6. Hyndman R., Bergmeir Ch., Caceres G., Chhay L., O'Hara-Wild M., Petropoulos F., Razbash S., Wang E., Yasmeen F. (2018), *Forecasting Functions for Time Series and Linear Models – Package 'Forecast'*, http://pkg.robjhyndman.com/forecast (accessed: 26.08.2019).
7. Hyndman R.J., Khandakar Y. (2008), *Automatic Time Series Forecasting: The Forecast Package for R*, “Journal of Statistical Software”, Vol. 27.
8. Hyndman R.J., King M.L., Pitrun I., Billah B. (2002), *Local Linear Forecasting Using Cubic Smoothing Splines*, “Australian & New Zealand Journal of Statistics”, Vol. 10.
9. Kramarz M. (2013), *Modelowanie przepływów materiałowych w sieciowych łańcuchach dostaw. Odporność sieciowego łańcucha dostaw wyrobów hutniczych*, Difin, Warszawa.
10. Kramarz M., Kmiecik M. (2017), *Pomiar zakłóceń w wybranym węźle sieci dystrybucji*, “Zeszyty Naukowe Politechniki Częstochowskiej. Zarządzanie”, nr 26.
11. Rogalska M. (2016), *Wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych*, Politechnika Lubelska, Lublin.
12. Zagdański A., Suchwałko A. (2016), *Analiza i prognozowanie szeregów czasowych. Praktyczne wprowadzenie na podstawie środowiska R*, Wydawnictwo Naukowe PWN, Warszawa.

UŻYTECZNOŚĆ ALGORYTMÓW PROGNOSTYCZNYCH NA PRZYKŁADZIE WYBRANYCH SIECI DYSTRYBUCJI

Streszczenie: Artykuł skupia się na sprawdzeniu użyteczności algorytmów prognostycznych, które są dostępne w jednej z bibliotek środowiska programistycznego R library(forecast). Sprawdzenie algorytmów odbywało się w pięciu wyselekcjonowanych sieciach dystrybucji. Każdy z algorytmów został porównany na podstawie generowanego przez niego błędu MAE do analizowanych szeregów czasowych. Głównym celem artykułu jest ukazanie użyteczności poszczególnych algorytmów, bazując na poziomach współczynnika zmienności poszczególnych szeregów. Wartość dodaną publikacji stanowi wskazanie na te algorytmy, które mogą być używane dla różnych produktów o różnych poziomach zmienności w poziomie wydań.

Słowa kluczowe: sieć dystrybucji, prognozowanie, oprogramowanie R