Współczesna Gospodarka

Contemporary Economy
Electronic Scientific Journal
www.wspolczesnagospodarka.pl

Vol. 9 Issue 2 (2018) 41-55
ISSN 2082-677X
DOI 10.26881/wg.2018.2.04

# ANALYSIS OF THE EFFECTIVENESS OF SELECTED DEMAND FORECASTING MODELS

**Anna Borucka**

## Abstract

**Objective**– Twomethods of predictionwereproposed in the article, usingsales data. Modelswereidentified and estimated, forecastsweredetermined, theirreliability was verified, and thenvaluesobtained for eachmethodwerecompared.
**Methodology** – The article presents models belonging to two different categories. They are regression function, which is a classic example of cause-and-effect model, and ARIMA model for time-series analysis.
**Results**– The results obtained for both models were satisfactorily described by empirical data, but the regression model is much easier to estimate and does not require complex transformations orcalculations, nor the use of specialized software. In the analyzed case, demand forecasting based on the linear regression model is sufficient and reflects the nature of studied phenomenon.

**Keywords:**forecasting, ARIMA model, linearregression model, demand.

**JEL classification**: C2, C22

## Introduction

Demandforecasting in the enterpriseisusuallyan importantissue, affectingeveryarea of itsfunctioning. It not onlybalances the demand for goods with supply, but alsofacilitatesdecision-making in manyaspects of the supplychain, supportingproducers, suppliers and sellers. Therearemanytypes of forecasts. For the purpose of thisarticle, the classiccause-and-effectmethod was used, i.e. linearregression, as well as ARIMA model for timeseriesstudy. Based on the actual dataprovided by the consideredenterprise, concerning the sale of the company'sflagshipproduct, twomodelswereidentified and estimated, obtainedresultswereverified and theirreliabilityassessed. Finally, obtainedforecastswerecompared. At the request of the company, itremainedanonymous.

## 1.Research procedure

Forecastsshould be constructed on the basis of dependencies. Theirdefinitionmustbe preceded by ananalysis of the collectedempirical data. First, a visualinspectioniscarried out with the use ofa linegraph (in the case of one-dimensionaltimeseries) and a box plot to identifyuncertainobservations. Thesechartsfor the salesprocessunderconsiderationareshown in Figure 1 and Figure 2.

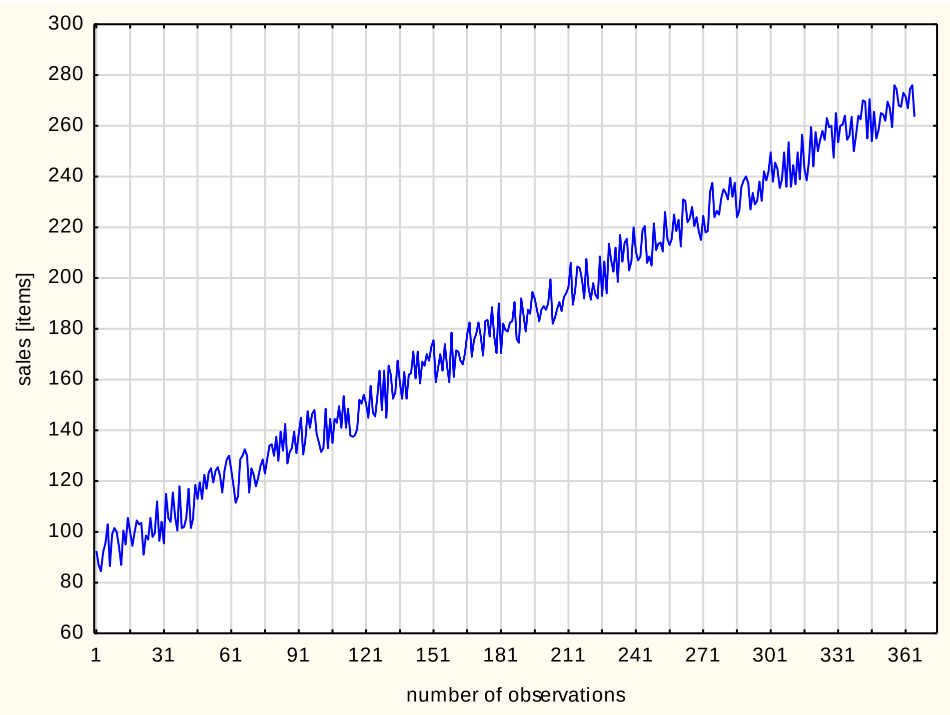


**Fig. 1.** Line graphof product sales

Source: the author's ownstudy.

The linegraph (Fig. 1) indicates a clear trend and increase in the value of salesovertime, while the box plot (Fig. 2) does not show the existence of outliers, as confirmed by the Grubbs test, for which the value of empiricalstatisticsturned out to be lowerthan the tablevalueat the significancelevel of$\alpha$=0.05. Therefore, thereis no need to interferewith theempirical data. Due to the strongcorrelationbetweensalesvalue and time, the next step was to checkitsstrength and confirm the direction. For thispurpose, a correlationbetweenvariableshasbeencalculated, whoseresultsare presented in Table 1 and Figure 3.

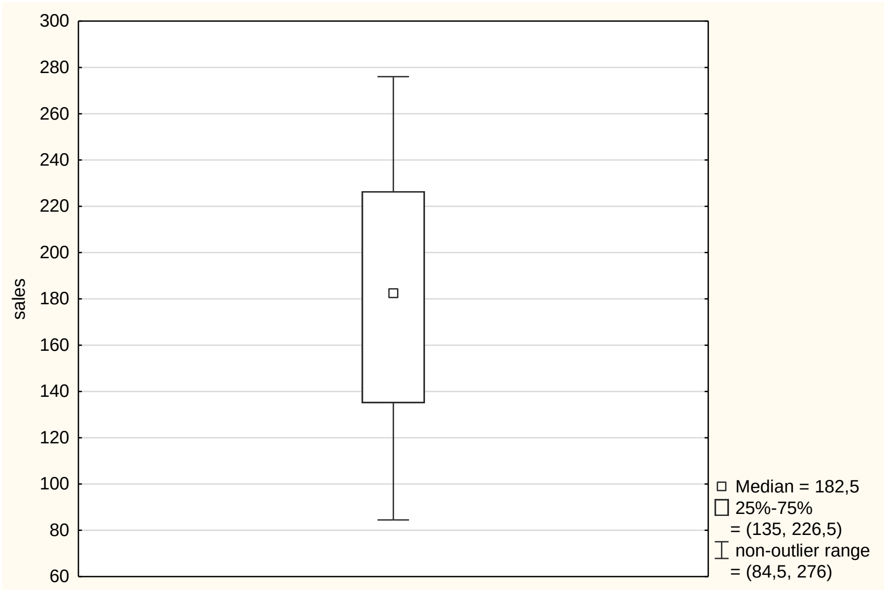


**Fig. 2.** Box plot of product sales

Source: the author's own study.

**Table 1.**Correlation matrix between variables

| Variable | Correlation. The significant correlation coefficients are underlined. p <0.05 N=365 | | | |
|---|---|---|---|---|
| | Mean | Standard deviation | sales | t |
| sales | 181.5740 | 52.8866 | 1.000000 | <u>0.993470</u> |
| *t* | 183.0000 | 105.5107 | <u>0.993470</u> | 1.000000 |

Source: the author's own study.

The Pearson correlationcoefficientobtained amounts tor(x, y)=0.99 and itisstatisticallysignificant, with the adoptedsignificancelevel of$\alpha$=0.05. Visual analysis of the correlation graph (Fig. 3) also indicates a linear relation. It is clear that with time, the value of sales increases. Due to such a strongcorrelation, itwas decided to use the cause-and-effect model for prediction, i.e. regression function. Itsusewillallow distinguishing the developmentaltendency and estimating dependent variables in the forecasted period.

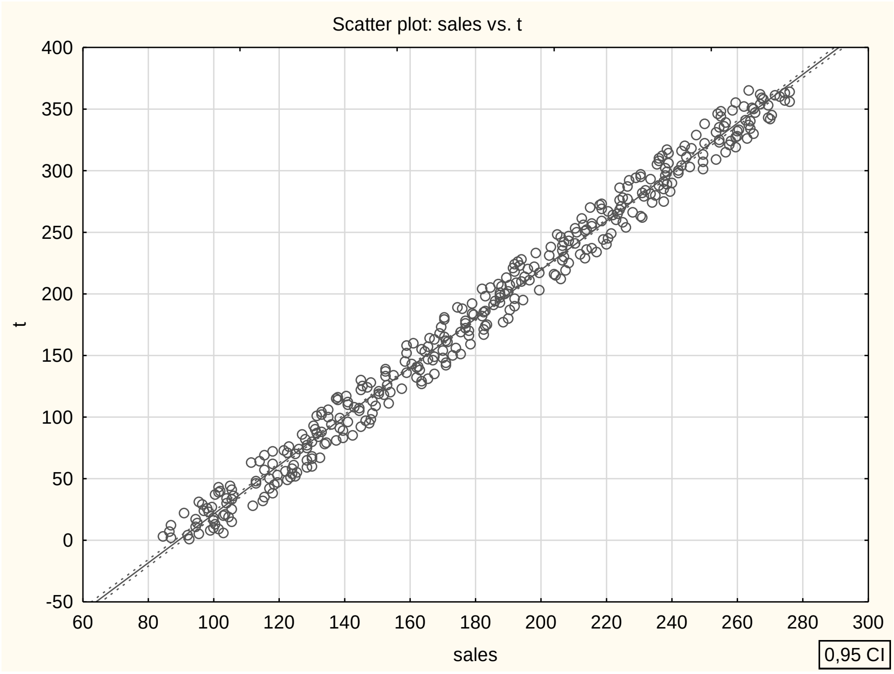


**Fig. 3.** Scatter plot sales versus time

Source: the author's own study

## 2. The linear regression model

Regressionanalysisis one of the most importantmathematicaltoolsused in economicanalyses, whichallows creating afunctionaldependencybetweenstudiedvariables in order to forecastvalues of one feature, assumingcertainvalues of one orseveralotherfeatures(Bielińska2007; Dittmann2000). The general form of the regressionfunctionis as follows (formula1):

$$y_i = b_0 + b_1 * x_{1i} + b_2 * x_{2i} + \cdots + b_k * x_{ki} \quad (1)$$

where:
$y_i$ – forecastfeature, dependent variable;
$x_{1i}$, $x_{2i}$ ,..., $x_{ki}$– independent variables;
$b_0$ – absolute term;
$b_1$,$b_2$,..., $b_k$– coefficients for independent variables.
The structuralparameters*b1,b2,...,* *bk*areoftendeterminedusing the Gauss method of the leastsquares, according to whichtotalsquares of deviations in observedvalues of variables

dependent on theoreticalvalues (determined on the basis of a createdfunction) should be the smallest(Maciąg, Pietroń, Kukla, 2013).

In the analyzedcase, the simplestregression model, i.e. simple linear regression (formula 2) was used. The dependent variable y is a forecast feature, i.e. expected demand for studied goods, while the independent variable x is time:

$$y = b_1 x + b_0 \tag{2}$$

Structural parameters $b_0$ and $b_1$ were estimated using Statistica computer program. The obtainedresultsarepresented in Table 2.

**Table 2.**Results of structural parameters estimation (Statistica)

| N=365 | regression model<br>$R^2$= 0.9869, Adjusted $R^2$= 0.9869, Standard error 6.0422 | | | |
|---|---|---|---|---|
| | b | Standard error b | t(363) | p |
| absolute term | 90.445 | 0.634 | 142.698 | 0.00 |
| t | 0.498 | 0.003 | 165.905 | 0.00 |

Source: the author'sstudy.

The standardestimation error is 6.04,whichmeansthat the foreseeable sales values differ from empirical values on average by 6 items. Determination coefficient $R^2$,whichmeasures the quality of model-fitting to empirical data, is 99%, which means very good model-fitting. This indicates what part of variability of a dependent variable is explained by the model. Thus, the variability of sales was explained in 98%.According to the aboveresults, the relationbetweenthe quantity of sold items and timecan be described by equation (formula3):

$$y \ = 0.49797 * time + 90.44528 \pm 6.04 \tag{3}$$

whichmeansthat the daily increase in number of sold items is about 0,5.

The next step isthe verification of the model. According to results in Table 3, the linearity of regression model isimportant (test probability$p$<0), and the estimatedregressioncoefficientsare alsoimportant.

Another step is to study the distribution of residuals. In a properlyconstructed model, the residualsshould be random and have a normal distribution. The following histogram (Fig. 4) and the graph of residualsnormality (Fig. 5) show thatthisdistributiondeviates from the normaldistribution, whichisconfirmed by the Shapiro-Wilk test, for which the value$W$ of statisticsat the significancelevel$\alpha$=0,05 turned out to be statisticallysignificant.

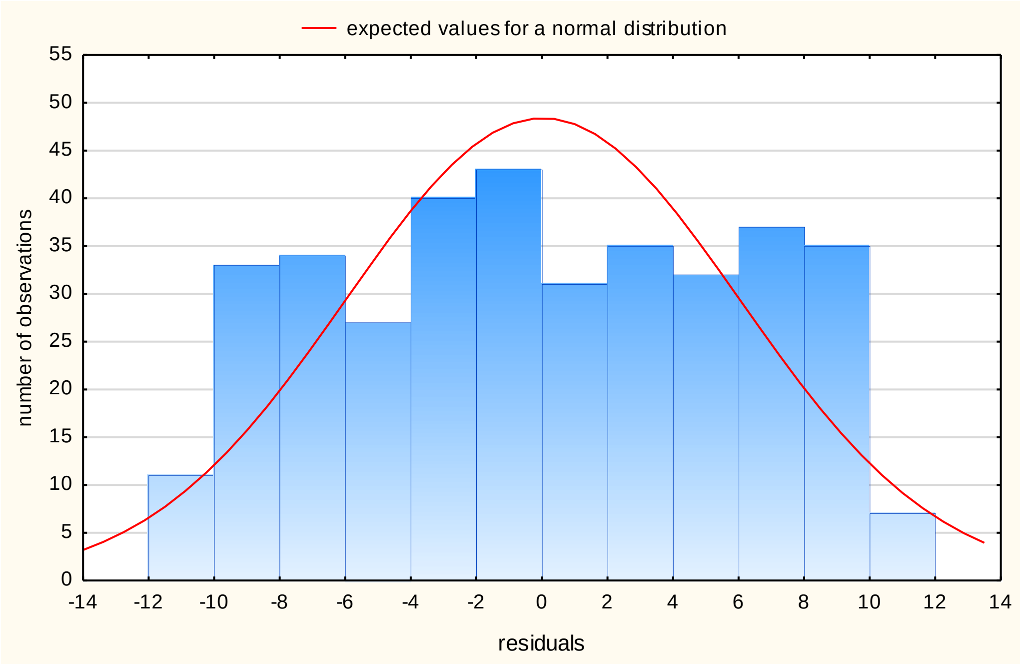


**Fig. 4.** Histogram of model residuals

Source: the author’s ownstudy.

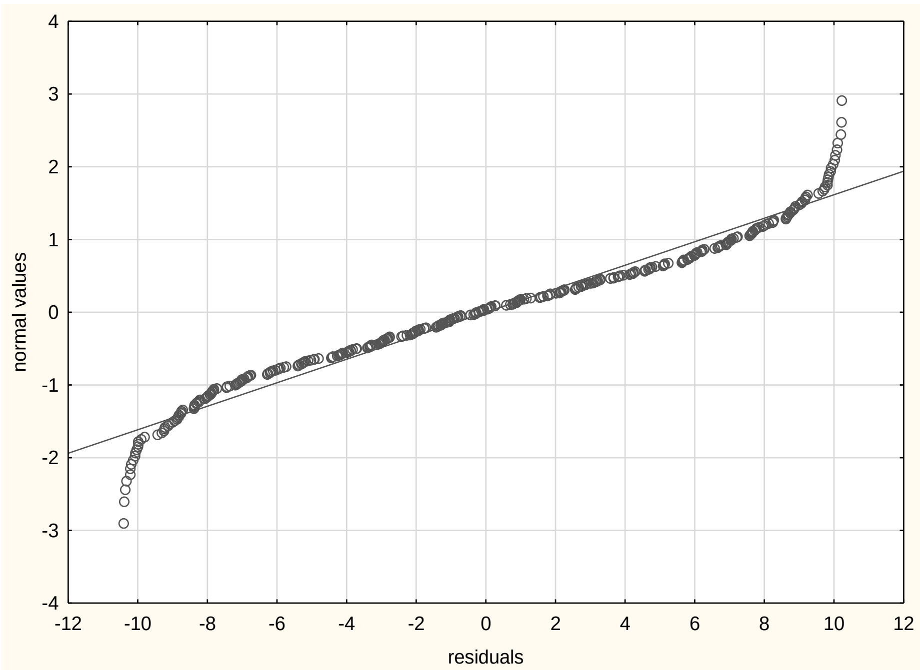


**Fig. 5.** Normal plot of residuals

Source: the author’s ownstudy.

The lack of normality of the distribution of residualsresults from dailyfluctuations in sales, whichoscillatearoundthe meanvalue, and thelinearregressionfunctioncannot accuratelyreflect the existingvariability, as illustrated by the graph of forecast and empirical data (Fig. 6). Therefore, the ARIMA model was proposed to compare the effectiveness of predictions.

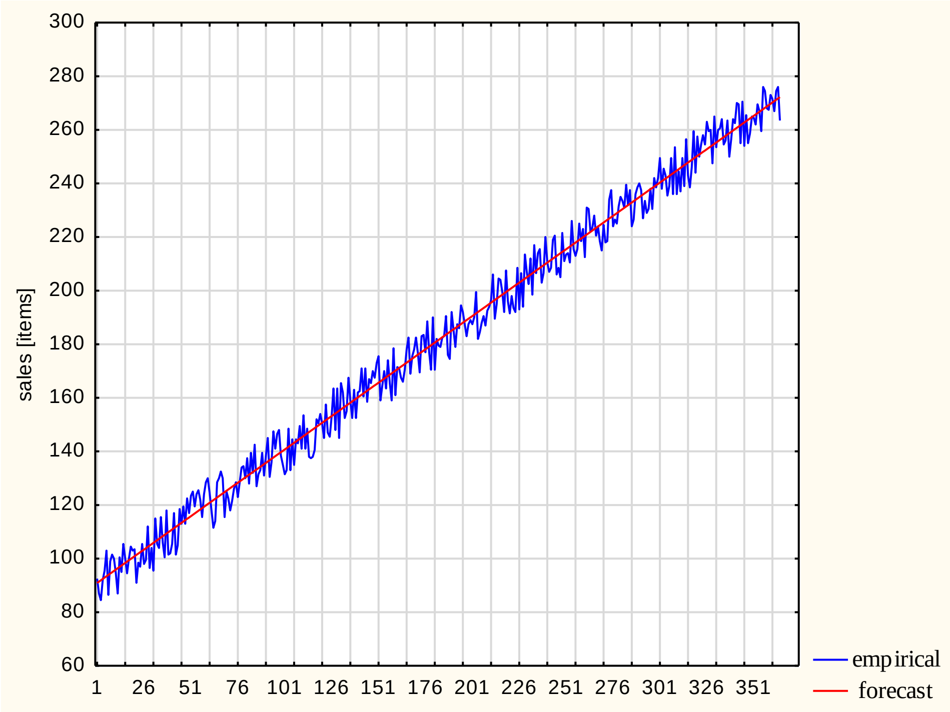


**Fig. 6.** Chart of empirical and forecasted data in regression model

Source: the author's own study.

## 3. ARIMA model

Autoregressive integrated moving average model (ARIMA) belongs to a group of forecasting methodsbased on time-seriesanalysis. Due to itsgreatercomplexity – compared to the normalcause-and-effectregression model – additionalrequirementsareinvolved. This model providesbetterefficiency and flexibility in matching, but canonly be used for stationaryor non-stationarymodelling, whichisreduced to stationary. Itscomponentsareautoregressivemodels and a movingaverage. The ARMA model isobtained from theircombination, based on the assumptionthat the value of a predictedvariableisaffected by its past values, as well as differencesbetween the real past values of forecastvariable and itsvaluesobtained from the model, i.e. so-called forecast errors. The form of ARMA model isshownbelow (formula4) (Dittmann,Szabela-Pasierbińska,Dittmann,Szpulak,2011;Sokołowski 2016):

$$y_t = \alpha_0 + \alpha_1 \, y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p \, y_{t-p} + \varepsilon_t \, \beta_0 + \varepsilon_t - \beta_1 \, \varepsilon_{t-1} -$$

$$-\beta_2 y_{t-2} - \dots - \beta_q \varepsilon_{t-q} \quad (4)$$

where:
$y_t, y_{t-1}, y_{t-2}, y_{t-p}$-values of forecasted variable at the moment or period $t, t-1, t-2, \dots, t-p$;
$\alpha_0, \alpha_1, \alpha_2, \alpha_p$- parameters for autoregressive part of model (AR);
$p$- delay value.
$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-q}$- error (residuals) of the model at the moment or period t$t, t-1, t-2, \dots, t-q$;
$\beta_0, \beta_1, \beta_2, \beta_q$- parameters for moving average part of model (MA)
$q$- delay value.

The use of ARMA modelsislimitedonly to stationaryseries. In caseswherethe analyzedseriesis not stationary, but stationarityisachievable, the ARIMA model can be used(Bielińska,2007). The additionalletter‘I’ in the nameindicatesthatthe studiedtimeseries was subjected to differentiation in order to obtainstationary form. Parameter d indicates how many such actions should be performed. The estimation of the ARIMA model requiresanappropriateprocedure, named afteritsauthors, the Box and Jenkins methodology, whichis based on the followingstages: identification, estimation, and forecasting.According to the above, the first step of the analysisis to studythe stationarity of a series. Theanalysis of the course of timeseries (Fig. 1) alreadyexcludesa stationarycharacterdue to the existence of a trend, indicating a need to bring the series to stationary form. The ACF autocorrelation function (Fig. 7) and PACF partialautocorrelationfunction (Fig. 8) arealsohelpful in the study of stationarity.

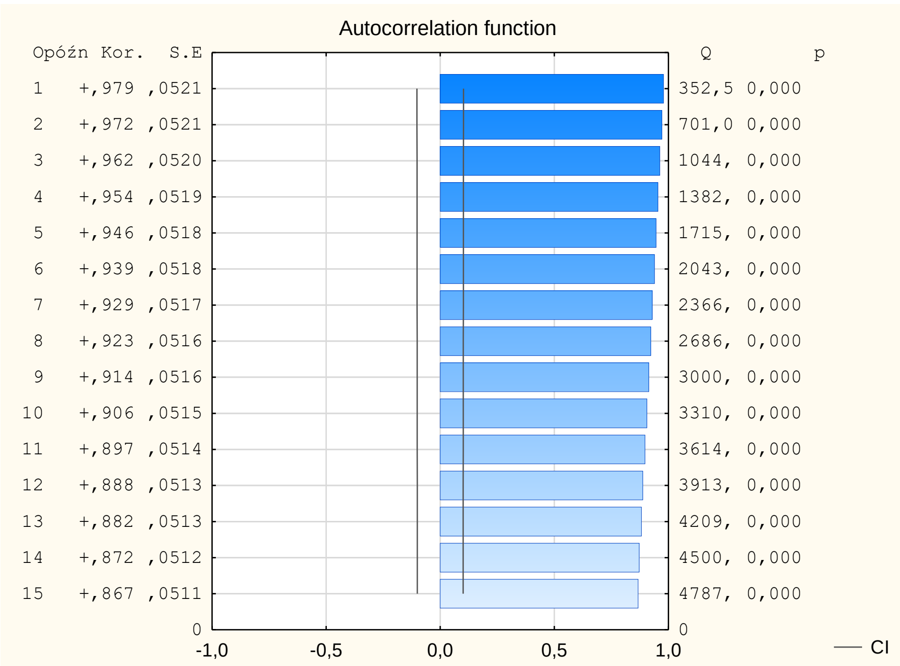


**Fig. 7.** Chart of the autocorrelation function for sales variable

Source: thr author’s ownstudy.

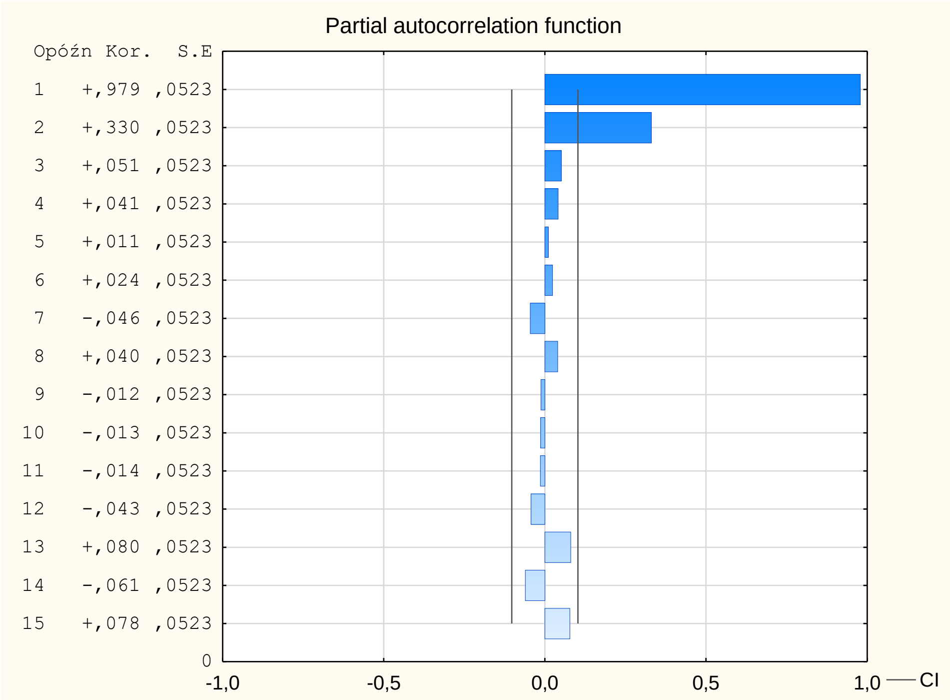


**Fig. 8.** Chart of partial autocorrelation function for sales variable

Source: the author's own study.

The autocorrelationgraphreveals a strongcorrelation of the currentobservation with the previous one, whichindicates the necessity to carry out differentiation with a delayequal to -1. Such a procedurewill not onlyeliminate the trend, but willalsoaffect the stationarity of the series. The results of the variabletransformationarepresented in Figure 9.

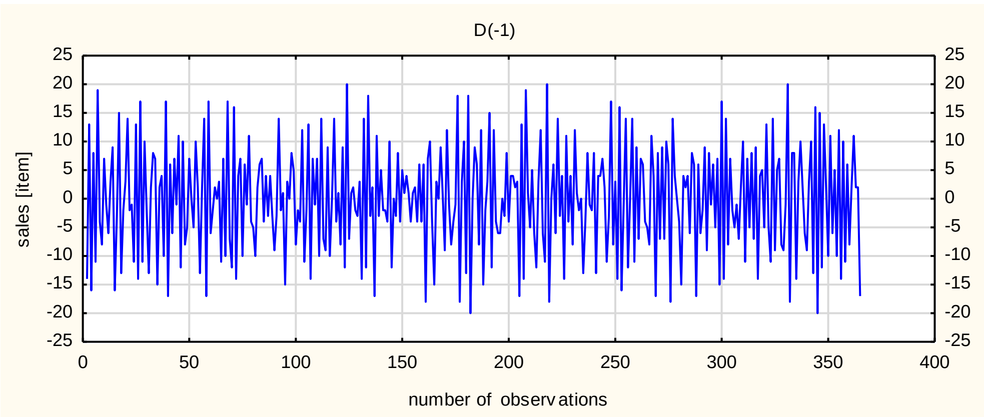


**Fig. 9.** Results of variable transformation

Source: the author's own study.

The analysis of the autocorrelation and partialautocorrelationfunctionisalsohelpful in estimating the parameters of the ARIMA model. Since the value of the timeseriesiscorrelated with itspreviousvalue, as shown in the ACF graph, the analyzedprocessisan autoregression. The order of the autoregressiveprocessisindicated by the PACF function, which for the AR(p) model takesvaluesequal to zero for delaysgreaterthan p (preciselyindicatingthatfragmentaryautocorrelationcoefficients for partialdelaygreaterthan p arestatisticallynot significantlydifferent from zero). Therefore, the surveyedseriesis a series with normalautoregression of at most the second order.

Ananalogousprocedureshould be carried out also for delayedvariable D(-1), due to the factthatremoval of autocorrelation of the higher order oftenrevealscorrelations of a lower order and, for example,a previouslyinvisible seasonalrelation. The ACF functionafterdifferentiationis shown in Figure 10.

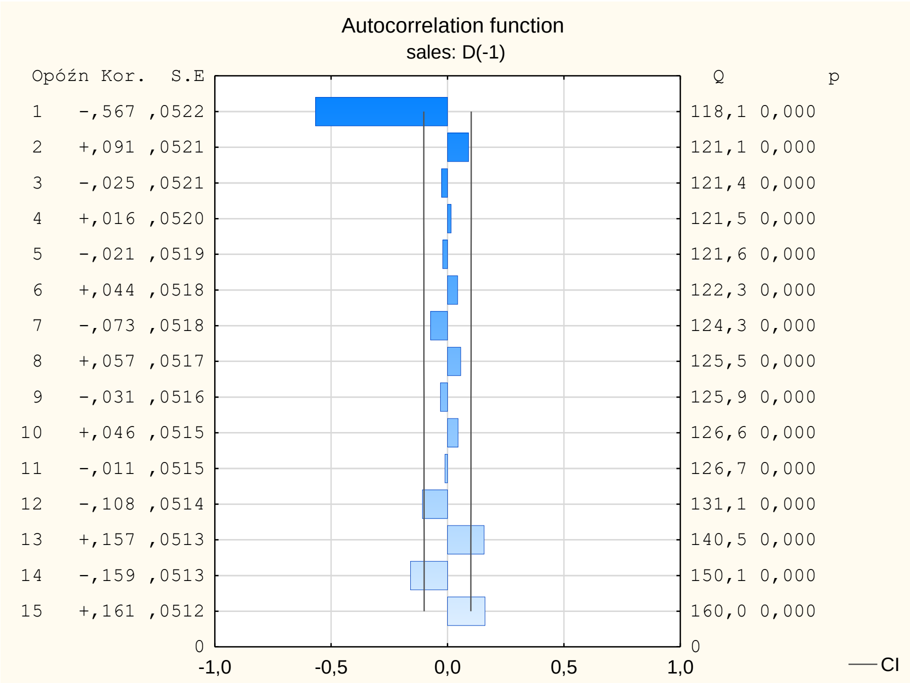


**Fig. 10.** Autocorrelation function after differentiation

Source: the author’s own study.

The correctness of ACF autocorrelationgraph for a differentiatedvariablewillallow for the nextstage, i.e. estimation. Severalmodelsareproposed, as shown in Table3, whichis a commonprocedure. In most cases, severaldifferentpossibilitiesareproposed in order to make a finalselection of the best ones (based on the analysis of selected criteria such as e.g. significance of model parameters, forecast error or information criteria).

**Table 3**. Summary of estimation results

| | **Model: (1,1,1)** | **Model: (0,1,2)** | **Model: (0,1,2)** | **Model: (0,1,1)** |
|---|---|---|---|---|
| Transformation | ln(x)D(1) | ln(x)D(1) | ln(x)D(1) | ln(x)D(1) |
| Constant | 0.00299 | 0.00299 | | 0.00299 |
| p | -0.0946 | | | |
| q(1) | 0.85272 | 0.94453 | 0.84180 | 0.87283 |
| q(2) | | -0.0795 | -0.1273 | |
| MS | 0.00164 | 0.00164 | 0.00182 | 0.00165 |

Source: the author's own study.

Only two of the above models have all estimated statistically significant parameters. However, the analysis of residuals in bothmodelsshowedthat in the ARIMA (0,1,2) model in the correlogramstillindicatessignificantfunctionvalues, suggestingthatthe distribution of residualsis not normal and thereareunexplaineddependencymodels. However, in the case of the ARIMA (0,1,1) model,such relations have not beenrevealed (Fig. 11 and Fig. 12), whichallows considering the residuals as a process of whitenoise (residualsare not correlated).

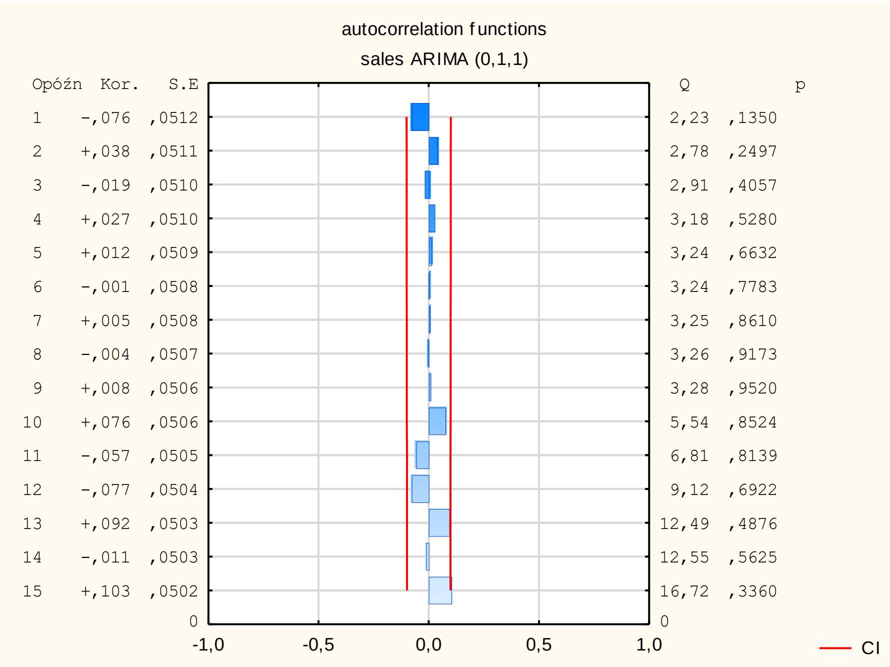


**Fig. 11.** The ARIMA (0,1,1) residuals autocorrelation function

Source: the author's own study.

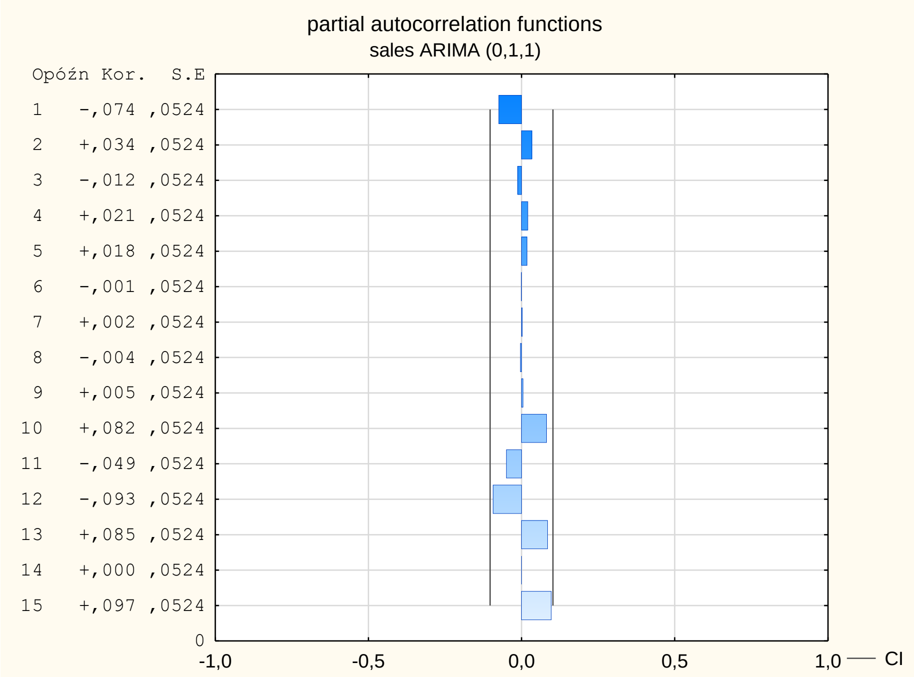


**Fig. 12.** Model ARIMA (0,1,1) residuals partial autocorrelation function

Source: the author's own study.

The graph of forecast and empirical data ispresented in Figure13.

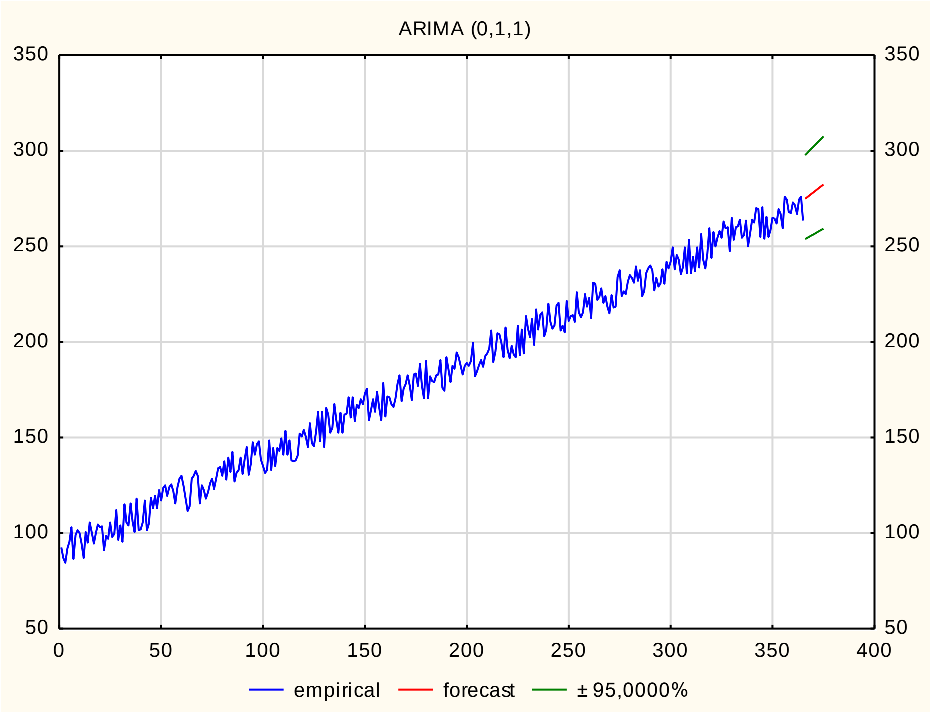


**Fig. 13.** Chart of empirical and forecasted data in the ARIMA (0,1,1) model

Source: the author's own study.

At the end of the study, twoproposedmodelswerecompared with empirical test observations (Table4), whichwere not used to constructany of them. It turns out thatthe forecasts do not differsignificantly and the predictedvaluesarecharacterized by a small relativeforecast error. The resultsobtained for bothmodelsweresatisfactorilydescribed by empirical data, but the regression model is much easier to estimate and does not requirecomplextransformationsor calculations, nor the use of specialized software. In the analyzedcase, demandforecastingbased on the linearregression model issufficient and reflects the nature of the studiedphenomenon.

**Table 4**. Comparison between regression and ARIMA model

| Numer of observation | Regression model | ARIMAmodel | Empirical data | Mean standard error $\Psi$[%]Regression model | Relative forecast error$\Psi$[%] ARIMA model | Comparison between regression and ARIMA model |
|---|---|---|---|---|---|---|
| 366 | 272.7027 | 274.9032 | 266 | -2.5198 | -3.34706 | -2.2005 |
| 367 | 273.2006 | 275.7270 | 275.5 | 0.834615 | -0.08238 | -2.5263 |
| 368 | 273.6986 | 276.5532 | 276 | 0.833838 | -0.20044 | -2.8546 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 369 | 274.1966 | 277,3820 | 278.5 | 1.545215 | 0.401448 | -3.1854 |
| 370 | 274.6945 | 278.2132 | 265 | -3.65832 | -4.98611 | -3.5186 |
| 371 | 275.1925 | 279.0469 | 284.5 | 3.271522 | 1.916729 | -3.8544 |
| 372 | 275.6905 | 279.8831 | 269 | -2.48717 | -4.04577 | -4.1926 |
| 373 | 276.1885 | 280.7218 | 284.5 | 2.921455 | 1.328001 | -4.5334 |
| 374 | 276.6864 | 281.5631 | 282 | 1.884244 | 0.15494 | -4.8766 |
| 375 | 277.1844 | 282.4068 | 279.5 | 0.828478 | -1.04001 | -5.2224 |

Source: the author's own study.

Linearregressionmodels and ARIMA models are among the short-term forecasting methods, but such predictions must be closely monitored and verified. It is not possible to make clear decisions on their basis; their task is only to support management processes and judicial proceedings on the future values of forecasted phenomena

## CONCLUSIONS

Demandforecastsareaneffectivetoolfor supporting the planningprocess in a company. Theircompetent and reasonableusecan be a support for managers in shaping the supplychain, deciding on necessaryorders andschedulingproductiondates. It alsoallows detection and quick response to changes in the market, whichisoften akeyfactorthat determines the future of the wholecompany. Therearemanymethods to describeupcomingphenomena, characterized by a differentdegree of complexity and estimation difficulties. They often require appropriate mathematical software.

The article presents models belonging to two different categories. Theyarethe regressionfunction, whichis a classicexample of a cause-and-effect model, and the ARIMA model for timeseriesanalysis. Theresultsobtained for bothmethodsproved to be satisfactorilyreliable, but the construction of a regression model is much simpler and does not requireanyadditionalassumptions. Therefore, itisworthwhile to try out the simple and equallyeffectivetoolsbeforeusingadvancedtechniques, as itturns out thatthanks to them we canensure not onlyoptimization of results in the company, but alsocorrective action where this is necessary.

### References

Bielińska, E. (2007).*Prognozowanieciągówczasowych*.Gliwice: WydawnictwoPolitechnikiŚląskiej.

Dittmann, P. (2000).*Metodyprognozowaniasprzedaży w przedsiębiorstwie*.Wrocław: WydawnictwoAkademiiEkonomicznej.

Dittmann, P., Szabela-Pasierbińska, E., Dittmann, I., Szpulak, A. (2011).*Prognozowanie w zarządzaniusprzedażąifinansamiprzedsiębiorstwa*.Warszawa: Wolters Kluwer Polska Sp. Z o.o.

Maciąg, A., Pietroń, R., Kukla, S. (2013).*Prognozowanieisymulacja w przedsiębiorstwie*.Warszawa, PolskieWydawnictwoEkonomiczne.

Sokołowski, A. (2016).*Prognozowanieianalizaszeregówczasowych*. Materiałyszkoleniowe.Kraków: StatSoftPolska.

## ANALIZA EFEKTYWNOŚCI WYBRANYCH MODELI PROGNOSTYCZNYCH POPYTU

### Streszczenie

**Cel**. W artykule, wykorzystując dane dotyczące sprzedaży, zaproponowano dwie metody predykcji.popytu. Dokonano identyfikacji i estymacji modeli, wyznaczono prognozy, sprawdzono ich wiarygodność a następnie porównano wartości otrzymane dla każdej z metod.

**Metoda**. W artykule zaprezentowano modele należące do dwóch różnych kategorii. Funkcję regresji, będącą klasycznym przykładem modelu przyczynowo – skutkowego, oraz służący do analizy szeregów czasowych model ARIMA.

**Wyniki**.Wyznaczone prognozy nie różnią się zdecydowanie między sobą, a przewidywane wartości charakteryzuje niewielki, względny błąd prognozy. Otrzymane wyniki dla obu modeli satysfakcjonująco opisały analizowane dane empiryczne, jednak model regresji jest zdecydowanie łatwiejszy do estymacji i nie wymaga skomplikowanych przekształceń i obliczeń, a także wykorzystania specjalistycznego oprogramowania. W analizowanym przypadku, prognozowanie popytu w oparciu o model regresji liniowej jest wystarczające i oddaje charakter badanego zjawiska.

**Słowa kluczowe:**prognozowanie, model ARIMA, model regresji liniowej, popyt

**Klasyfikacja JEL**: C2, C22

Anna Borucka
Wojskowa Akademia Techniczna
ul. Gen. Witolda Urbanowicza 2
00-908 Warszawa 46
anna.borucka@wat.edu.pl