

ON THE METHOD OF IDENTIFICATION OF ATYPICAL OBSERVATIONS IN TIME SERIES

Maciej Oesterreich

West Pomeranian University of Technology, Szczecin, Poland
Department of Applied Mathematics in Economics, Faculty of Economics
e-mail: maciej.oesterreich@zut.edu.pl

ORCID: 0000-0003-4750-3636

© 2020 Maciej Oesterreich

This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivs license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

DOI: 10.15611/ead.2020.2.01

JEL Classification: C32, C53

Abstract: The paper presents a method of detecting atypical observations in time series with or without seasonal fluctuations. Unlike classical methods of identifying outliers and influential observations, its essence consists in examining the impact of individual observations both on the fitted values of the model and the forecasts. The exemplification of theoretical considerations is the empirical example of modelling and forecasting daily sales of liquid fuels at X gas station in the period 2012-2014. As a predictor, a classic time series model was used, in which 7-day and 12-month cycle seasonality was described using dummy variables. The data for the period from 01.01.2012 to 30.06.2014 were for the estimation period and the second half of 2014 which was the period of empirical verification of forecasts. The obtained results were compared with other classical methods used to identify influential observations and outliers, i.e. standardized residuals, Cook distances and DFFIT. The calculations were carried out in the R environment and the Statistica package.

Keywords: forecasts, identification, multiple regression, time series, outliers.

1. Introduction

In the process of modelling economic phenomena, on the basis of data in the form of time series showing both seasonal fluctuations and without such fluctuations, the observations with values much lower or significantly higher than the others become a serious problem. These observations are called atypical observations. Their occurrence affects the properties of estimators, which as a consequence, distorts the description of the phenomenon expressed by the overstatement or undervaluation of both parameter assessments and theoretical values and forecasts.

Among atypical observations, the outliers and influential observations are distinguished (Dittmann, Dittman, Szabela-Pasierbińska, and Szpulak, 2009, p. 46).

An outlier is an observation whose value clearly differs from the others (Anderson, Sweeney, and Williams, 2011, p. 678; Cook, 1980, p. 1). Due to the type of impact i -th observation on the series, one can additionally distinguish (Tsay, 1988; Watson, Tight, Clark, and Redfern, 1991; Zahari, Zainol, Sopian, Zaharim, and Ibrahim, 2010):

- additive outliers,
- innovative outliers,
- level changes,
- permanent level changes,
- transient level changes,
- variance changes.

Influential observations, in turn, can be defined as observations that individually or in combination with other observations have a greater impact on estimated values (coefficients) than other ones (Belsley, Kuh, and Welsch, 2004).

In the paper by Chatterjee and Hadi (1986), a comprehensive description of detection methods for outliers and influential observations was made. The authors classified them into methods of:

- residuals,
- projection matrix ,
- volume of confidence,
- influence functions,
- partial influence.

The purpose of the work is to present an iterative method of the detection of atypical observations (influential observations, outliers) by analysing the strength and a direction of impact of individual observations in the estimation and the forecast period, as well as comparing its accuracy with the classic methods of outliers detection. As the main research hypothesis, it was assumed that atypical observations affect the descriptive properties of the model and the accuracy of forecasts differently.

The following parts present: classic methods for detecting unusual observations (Chapter 2), proposed procedure (Chapter 3), empirical example regarding the application of presented methods (Chapter 4) and analysis of the impact of the effects of unusual observations on prognostic properties (Chapter 5). The paper ends with a summary.

2. Methods of detection of influential observations and outliers

A synthetic description of the most commonly used classic methods of detecting atypical observations:

2.1. Projection matrix

Projection matrix (PM) is determined according to the following formula (Dittmann, et al., 2009, p. 48):

$$H = X(X^T X)^{-1} X^T, \quad (1)$$

where: X – the matrix of explanatory variables.

The elements of the H matrix lying on the main diagonal are called influential observations (h_i). They fulfill the following properties:

$$\frac{1}{N} \leq h_i \leq 1, \quad (2)$$

where: N – the number of observations of the estimation period.

The authors consider as influential observations those for which h_i values meet the condition (Hoaglin and Welsch, 1978):

$$h_i \geq \frac{2p}{N}, \quad (3)$$

where:

$$p = \sum_{i=1}^N h_i. \quad (4)$$

Observations which exceed (3) are called leverage points.

2.2. Standardized residuals

Standardized residuals of the model (SR) are determined as:

$$SR_i = \frac{e_i}{S_e}, \quad (5)$$

where: e_i – the i -th residue of the model, S_e – the standard deviation of a component of the random model.

The outliers are those observations for which the following relation (Dittmann et al., 2009, p. 49) occurs:

$$|SR_i| > 3. \quad (6)$$

The selected limit level for standardized residues should not be less than 2.

2.3. Cook distances

Cook distances (CD) are determined according to the formula (Cook, 1977; Kannan Senthamarai and Manoj, 2015):

$$CD_i = \frac{e_i^2 h_i}{(m+1) S_e^2 (1-h_i)^2}, \quad (7)$$

where: h_i – the i -th element of the main diagonal of the projection matrix, e_i^2 – the i -th residue of the model, S_e^2 – the variance of a random component of the model, m – the number of explanatory variables.

The atypical observations are those ones that meet the following relation (Algur and Biradar, 2017):

$$CD_i \geq \frac{4}{N}. \quad (8)$$

High values of CD_i indicate observations that have impact on parameters of the constructed model.

2.4. DFFIT

DFFIT values are determined according to the following formula (Cousineau and Chartier, 2010; Paul, 2018):

$$DFFIT_i = \frac{h_i e_i}{1 - h_i}, \quad (9)$$

where: h_i – the i -th element of the main diagonal of the projection matrix, e_i – the i -th residue of the model.

The outliers are those observations for which the DFFITS values meet the following condition (Belsely, Huh and Welsch, 2004, p. 28):

$$|DFFIT_i| \geq 2\sqrt{\frac{m}{N}}, \quad (10)$$

where: m – the number of independent variables.

The basic drawback of methods based on the values of the main diagonal of the projection matrix, as well as standardized residuals, is the fact that they only take into account the impact of observations from the estimation period on the theoretical values of the model, completely ignoring the fact of their impact on the forecast values. This means that the effect of removing atypical observations from the estimation period may improve the descriptive properties of the model, while reducing the accuracy of forecasts (cf. Oesterreich, 2017; Zawadzki, 1999, pp. 42-43; Zawadzki, 2003, pp. 29-30).

3. The description of the method

The essence of the proposed procedure of detecting atypical observations is based on examining the impact of individual observations of the series on the fitted values of the model and forecast. The method consists of three stages:

- Stage 1: the estimation of the model for the full series (without gaps) and the calculation on its basis fitted values (\hat{Y}_{teo_t}) and the of forecasts (\hat{Y}_{fc_t}) for the previously set horizon ($h = 1, \dots, r$).
- Stage 2: the construction of N models for the series, from which one observation (i) was successively eliminated and the calculation, on their basis, N series of: fitted values ($\hat{Y}_{teo_{it}}^*$) and forecasts ($\hat{Y}_{fc_{it}}^*$) ($i = 1, \dots, N$, N – number of observations in the estimation period).
- Stage 3: the comparison of fitted values for estimation and forecasts intervals, calculated on the basis of the full series and the series from which individual observations were eliminated:

$$CT_i = \frac{1}{N} \sum_{t=1}^N (\hat{Y}_{teo_t} - \hat{Y}_{teo_{it}}^*) \quad (11)$$

$$CF_i = \frac{1}{h} \sum_{t=N+1}^{N+h} (\hat{Y}_{fc_t} - \hat{Y}_{fc_{it}}^*), \quad (12)$$

where: \hat{Y}_{teo_t} – the fitted value for t -th observation calculated on the basis of a model built on the basis of the full series, $\hat{Y}_{teo_{it}}^*$ – the fitted value for t -th observation calculated on the basis of the model built after removing i -th observation for the estimation period, \hat{Y}_{fc_t} – the t -th forecast based on a model built on the basis of a full series, $\hat{Y}_{fc_{it}}^*$ – the t -th forecast based on the model built after the removal i -th observation for the estimation period, N – the number of observations of the estimation period, h – the forecast horizon.

Equation (11) describes the average differences of the fitted values of the model caused by the removal of i -th observation in the test period (CT), while formula (12) the average differences in forecast values caused by the removal of i -th observation in the sample time interval (CF). $CT_i = 0$; $CF_i = 0$ when deleting the i -th observation does not affect the fitted value or forecast.

- Stage 4. Checking whether there are observations in the CT and CF series that meet the following properties:

$$|CT_i| > \bar{x}_{CT} + p \cdot S_{CT} \quad (13)$$

$$|CF_i| > \bar{x}_{CF} + p \cdot S_{CF}, \quad (14)$$

where: \bar{x}_{CT} – the average for the CT series, \bar{x}_{CF} – the average for the CF series, S_{CT} – a standard deviation for the CT series, S_{CF} – a standard deviation for the CF series, p – the value adopted in advance. In the paper it was set at 3 (see Osborne and Overbay, 2004).

The observations satisfying the inequality (13) should be considered as atypical observations (outliers) of the first order. They will affect the descriptive properties of

the model. When the observations meet inequality (14), they will be called atypical observations (outliers) of the second order. They will affect the predictive properties of the model.

4. The empirical example

The variable describing daily sales of liquid fuels (in litres) at X gas station in 2012-2014 was analysed. The sale of liquid fuels is understood as the sum of sales of Pb95 and Pb98 unleaded gasoline as well as diesel. The data for the period from January 1, 2012 to June 30, 2014, which consisted of 912 observations, constituted the model estimation period. The period of empirical verification of forecasts covered the second half of 2014 (184 observations). Figure 1 presents the analysed variable.

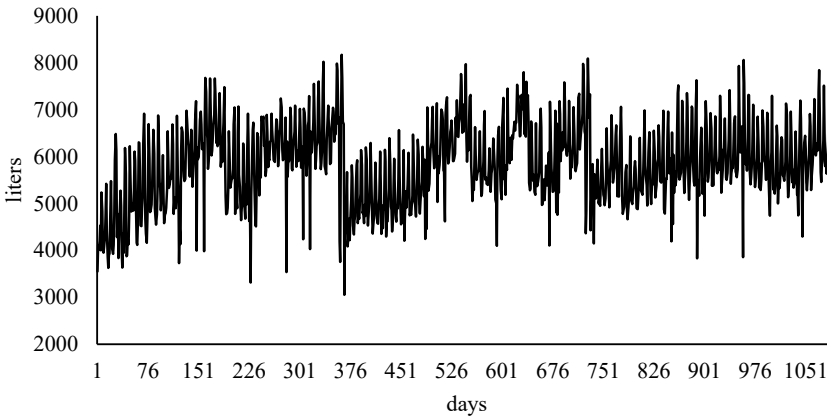


Fig. 1. Daily sales of liquid fuels at gas station X in the period from 01/01/2012 to 31/12/2014

Source: own study based on Department data.

For modelling and forecasting of the variable, a classic time series model with complex seasonal fluctuations of a 7-day and 12-month cycle, described by means of zero-one variables, was applied:

$$Y_t = \alpha_0 + \alpha_1 t + \sum_{k=1}^{12} b_{0k} M_{kt} + \sum_{j=1}^7 c_{0j} D_{jt} + U_t \quad (15)$$

$$t = 1, 2, \dots, N; k = 1, 2, \dots, 12; j = 1, 2, \dots, 7$$

assuming that:

$$\sum_{k=1}^{12} b_{0k} = 0; \sum_{j=1}^7 c_{0j} = 0, \quad (16)$$

where: α_0, α_1 – coefficient associated with a linear trend, b_{0k}, c_{0j} – coefficient associated with seasonality, M_{kt} – dummy variable that takes 1 for k month and 0 for the other, D_{jt} – dummy variable that takes 1 for j day and 0 for the other, U_t – random component.

In order to check the effectiveness of the proposed procedure, the results of detection of influential and outlier observations were compared for four selected methods:

- projection matrix (*PM*),
- Cook distances (*CD*),
- DFFITS (*DF*),
- standardized residuals (*SR*).

Calculations were made in the R environment using the functions: *hatvalues()*, *rstandard()*, *cooks.distance()*, *dffits()* and the proprietary function *gap_test()*.

Figure 2 shows the values of the main diagonal of the projection matrix.

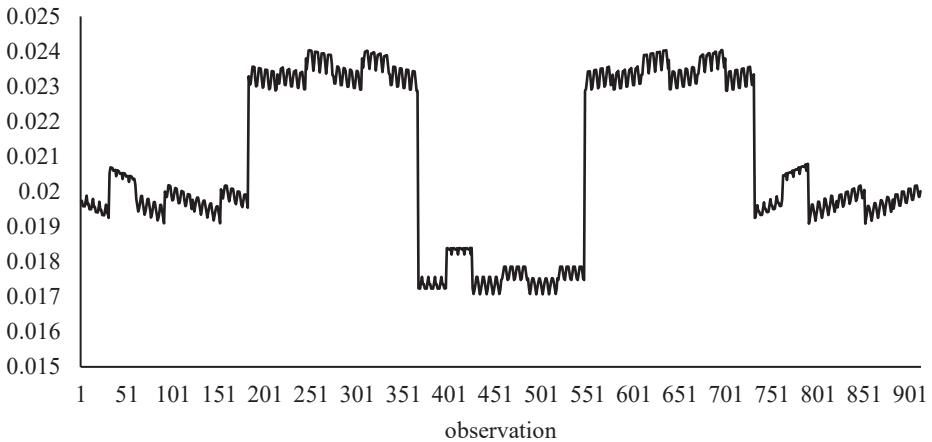


Fig. 2. Values of the main diagonal of the projection matrix

Source: own study.

None of the values found on the main diagonal of the projection matrix exceed 0.025. At the limit value of 0.04167 (formula 3), it should be considered that the analysed variable does not contain any influential observations.

Figure 3 shows the Cook distance (*CD*) for each of the observations of the estimation period. The horizontal line indicates the limit value of 0.004386 (formula 8).

For Cook’s distance, the limit indicating occurrence of outliers, was exceeded 44 times.

Figure 4 presents the DFFITS (*DF*) values for subsequent observations of the estimation period. Limit levels of ± 0.28097 were marked with horizontal lines (formula 10).

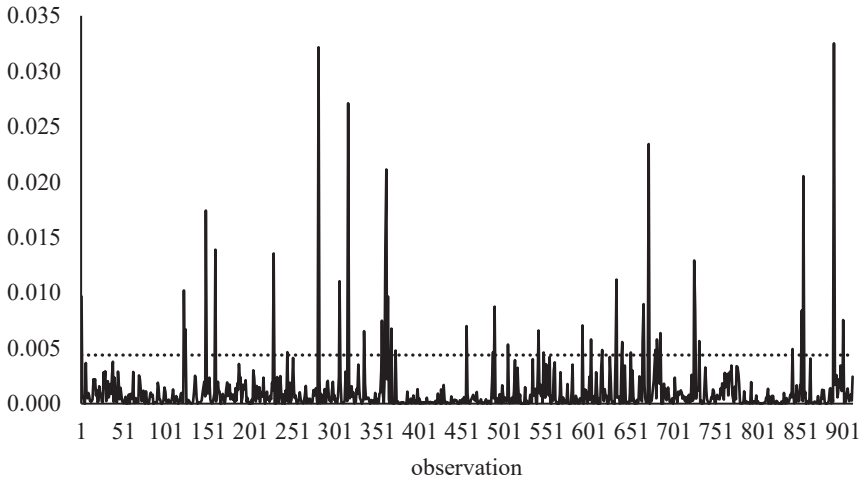


Fig. 3. Cook distance values for the observations from the estimation interval

Source: own study.

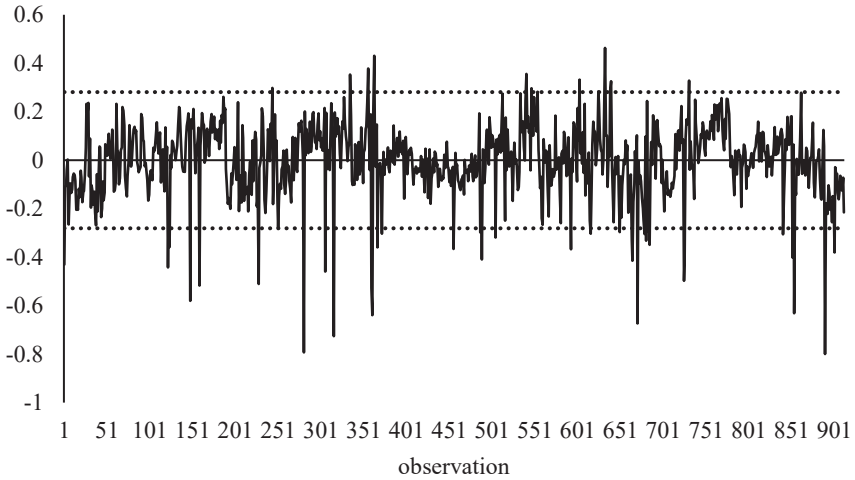


Fig. 4. DFFITS values

Source: own study.

The DFFITS limit was exceeded for 48 observations.

Figure 5 presents the values of standardized residuals (*SR*) for the observation of the estimation interval together with the limits of ± 3 standard deviations.

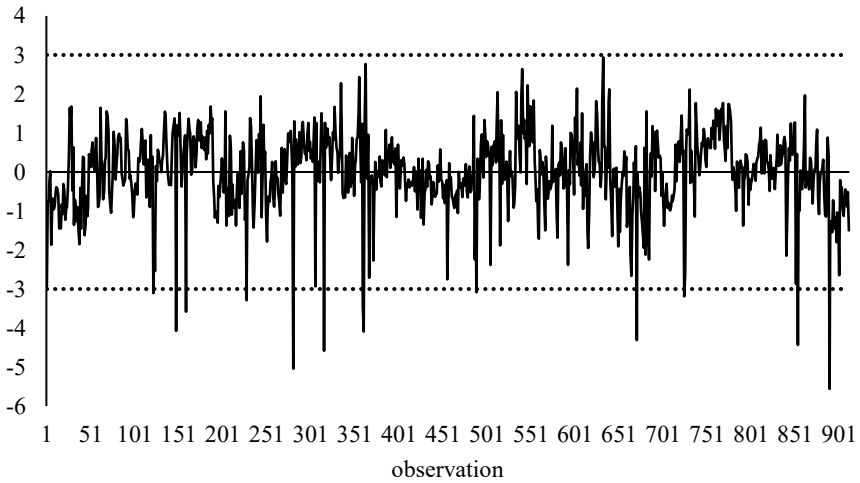


Fig. 5. Standardized residual values for the observations from the estimation period

Source: own study.

The values of standardized residuals indicate the occurrence of outliers during the estimation period. The limit (minus) of three standard deviations (formula 6) was exceeded for 14 observations.

Figure 6 presents the changes in level of fitted values (CT) as a result of deleting the i -th observation.

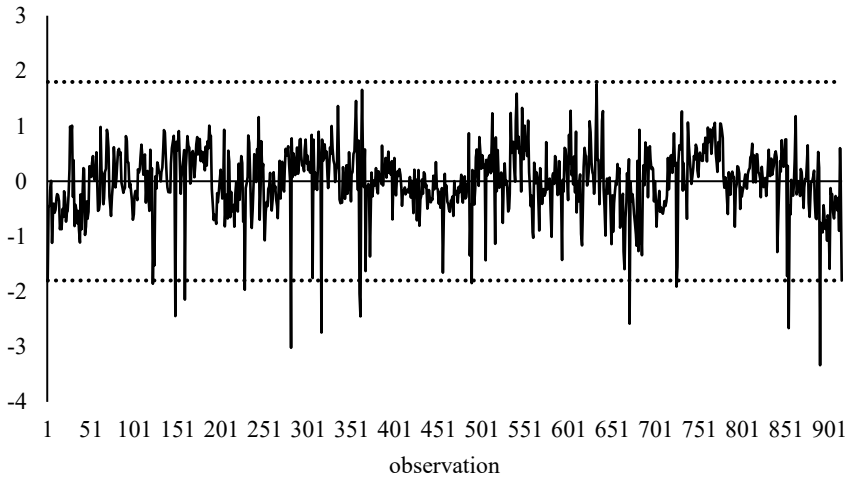


Fig. 6. Changes in level of fitted values of the model (CT) as a result of deleting the i -th observation of the estimation period

Source: own study.

The CT_i value series has exactly the same form as the series of standardized (SR) residuals. The limit of three standard deviations (formula 13) was exceeded for 14 observations. Their absence caused an increase in the level of fitted values and the excess of the limit of minus three standard deviations.

Figure 7 presents the changes in forecast values (CF_i) as a result of deleting the i -th observation of the estimation period.

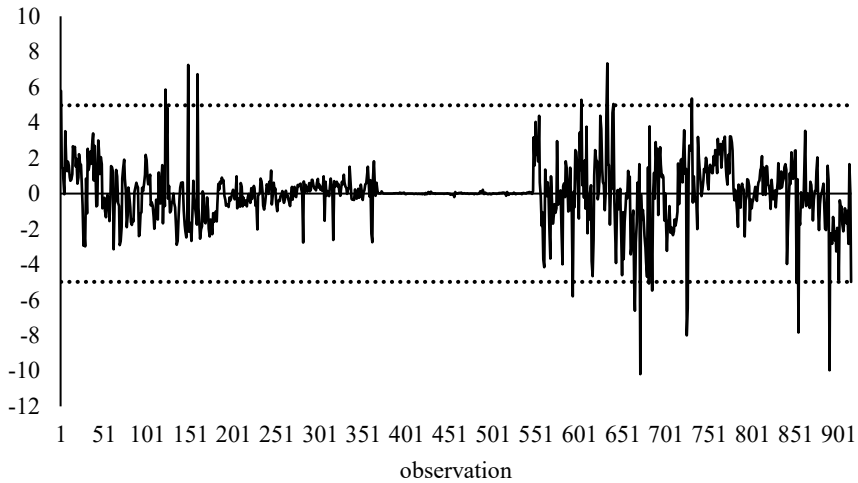


Fig. 7. Changes in the level of forecasts as a result of deleting the i -th observation of the sample time interval

Source: own study.

Figure 7 shows that the observations at both the beginning and the end of the estimation period strongly influenced the forecast values. The changes in the values above or below three standard deviations of the mean (formula 14) were recorded for 19 observations. This indicates that there are atypical observations (outliers) of the second order.

Table 1 summarizes, for all analysed methods, the observation numbers that were found to be outliers (of the 1st and 2nd order). The observations for which the limit values have been exceeded were marked: (-) for the lower and (+) for the upper limit. In the case of the Cook distance method, the exceeding of the limit value was indicated by the symbol (*).

The analysis of the data contained in Table 1 shows that the direction of the influence of atypical observations on the theoretical values of the model and the forecast may be different (observations No. 1; 122; 148; 159). This confirms the legitimacy of the distinction, in addition to atypical observations of the first order, and also atypical observations of the second order.

Table 1. Numbers of identified atypical observations of estimation period by methods

Number of observation	<i>SR</i>	<i>CD</i>	<i>DF</i>	<i>CT</i>	<i>CF</i>
1	2	3	4	5	6
1	–	*	–	–	+
122	–	*	–	–	+
124		*	–		
148	–	*	–	–	+
159	–	*	–	–	+
228	–	*	–	–	
244		*	+		
281	–	*	–	–	
306		*	–		
316	–	*	–	–	
335		*	+		
356		*	+		
360	–	*	–	–	
361	–	*	–	–	
363		*	+		
367		*	–		
372		*	–		
456		*	–		
487		*	–		
489	–	*	–	–	
505		*	–		
541		*	+		
547		*	+		
554			+		
593		*	–		–
603		*	+		+
616		*	–		
625			+		
632			+		
633		*	+		+
639			+		
640		*	+		+
650		*	–		
664		*	–		–

Table 1, cont.

1	2	3	4	5	6
665		*	–		–
671	–	*	–	–	–
679		*	–		
681		*	–		–
683		*	–		
685		*	–		–
725	–	*	–	–	–
726		*	–		–
731		*	+		+
841		*	–		
852		*	–		–
854	–	*	–	–	–
890	–	*	–	–	–
901		*	–		

Source: own study.

5. The impact of atypical observation on forecasting properties

Table 2 presents the values of the parameters of ‘goodness’ of fit, stochastic structure and the significance of structural parameters of models built on the basis of a full series and a series where atypical observations were removed.

Table 2. The assessment of parameters of the stochastic structure of models built on the basis of a complete series and after the removal of atypical observations

Methods	Number of removed observations	R^2	S_e	V_{se} [%]	Number of statistically significant coefficients
Complete series	–	0.628	551.921	9.538	17
SR	14 (1.54%)	0.703	480.144	8.257	18
CD	44 (4.82%)	0.759	417.769	7.178	18
DF	48 (5.26%)	0.763	412.414	7.095	18
CT	14 (1.54%)	0.703	480.144	8.257	18
CF	19 (2.08%)	0.690	493.588	8.506	17

Source: own study.

The characteristics of the model built on the basis of the original model were empirical data matching at the level of 62.8% and a random variation coefficient of 9.538%. The removal of 48 observations (DFFITS method) improved the assessment of fit measures and the stochastic structure to 76.3% and 7.095%, respectively. This means that deleting 5.26% of the series observations enabled an increase in the determination coefficient rating by 13.5 percentage points (pp), i.e. 21.5%.

The removal of the observations detected by the CT method enabled the determination coefficient (R^2) to increase by 7.5 pp and to reduce the level of random variation by 1.281 pp. Identical results were obtained for the SR method. Whereas removing the outlier observations of the second order (CF method) improved the accuracy of the model description by 6.2 pp (approximately 9.93%).

Table 3 presents the assessment of forecast errors *ex post* built with the use of the model for the full time series and one where atypical observations were removed.

Table 3. The evaluation of *ex-post* forecast errors (MAPE) for the full series and the refined series according to methods for the 184 days horizon (in %)

Method	Complete series	SR	CD	DF	CT	CF
MAPE	8.380	8.591	9.100	8.869	8.591	8.668

Source: own study.

Despite a clear improvement in the descriptive properties of the models, the removal of outliers obstructed the accuracy *ex post*. The model with the highest level of fit and the lowest level of the coefficient variation (time series reduced by the DFFITS method) was characterized by an error of 0.489 pp higher than the model for full data. Even worse results – an error of 9.100% – were obtained by removing from the series the observations indicated with the use of the Cook distance method. The analysis of Table 3 shows that the removal of atypical second order observations (CF method) did not improve either the accuracy of the forecasts. The obtained forecast errors were higher than for the SR or CT method. The reason for this is explained by a more detailed analysis of the data presented in Figure 8 and Table 1. The appearance of strong fluctuations in the initial and final part of the estimation interval suggests a change in the dynamics of the analysed phenomenon. These conclusions are confirmed by the average value of the CF series for 10% of the final observations of the estimation interval, amounting to -0.5846 . It also indicates that the value of forecasts is being underestimated (formula 11).

Due to a different impact direction of atypical observations of the second order on forecast values (see Table 1), Table 4 summarizes the average values of the CF series according to impact: positive (+) or/and negative (-).

The removal of all atypical observations contributed to the underestimation of the already underestimated forecasts, which consequently led to an increase in forecast error assessment. This means that in the case of the analysed the time series,

it is appropriate to delete only atypical observations of the second order for which the values of the CF series are greater than 0 (+). This will result in an increase in the level of constructed forecasts and thus a reduction of the average error.

Table 4. The average value of the CF series according to the impact direction of atypical observations of the second order

Impact direction of atypical observations	Average value of CF
–	–6.878
+	6.090
–/+	–1.418

Source: own study.

For comparative purposes, Table 5 presents the assessment of parameters of the stochastic structure and model's goodness-of-fitting and *ex post* forecast errors built on the basis of a series, where the atypical second order observations were removed, with a breakdown into positive (+) and negative (–) CF values.

Table 5. The assessment of parameters of goodness of fit and the stochastic structure of models built on the basis of a series after removal of atypical observations of the second order

Methods	Number of removed observations	R^2	S_e	V_{se} [%]	MAPE
CF(–)	11 (1.21%)	0.671	515.526	8.887	9.088
CF(+)	8 (0.88%)	0.625	531.831	9.188	8.038

Source: own study.

The removal of atypical observations of second order with CF values greater than 0 resulted in a decrease in the average relative forecast error to 8.038%, i.e. 4.1% lower than the one obtained for full data. This contributed to the deterioration of the descriptive properties of the model built on the basis of this series, whereas the removal of atypical observations with CF lower than 0 clearly deteriorated the quality of the obtained forecasts.

6. Conclusion

The analyses carried out at the paper confirmed that the direction in which observations affect the theoretical values of the model and forecast may be different. The refining of time series from atypical observations of the first order improved the descriptive properties of the models, which resulted in an increase in the level of the model's goodness-of-fit and a reduction in the level of variation of the random

component. At the same time this process reduced predictive properties of models and increased the *ex post* forecast error rate.

This confirms the legitimacy of the identification of the atypical observations of the second order, in addition to atypical observations of the first order. The elimination of this type of observations improved forecasting properties of the model, but also resulted in a deterioration of the model's descriptive properties.

The main advantage of the presented method is that it enables ex-ante analysis without knowing the real values in the forecasted period. The analysis showed that elimination of atypical observations of the second order should be preceded by the analysis of values of CF series. The appropriate methods of its analysis should increase the precision of detection of this type of observation. This will be one of the directions of future research.

The conducted research focused on identifying atypical observations and analyzing their impact for models constructed using the least squares method (linear or linearized). Therefore, future research will focus on the possibility of application of the presented method to other types of models.

References

- Algur, S. P., and Biradar, J. G. (2017). Cooks distance and mahabobis distance outlier detection methods to identify review spam. *International Journal of Engineering and Computer Science*, 6(6), 21638-21649. doi:10.18535/ijecs/v6i6.16
- Anderson, D. R., Sweeney, D. J., and Williams, T. A. (2011). *Statistics for business and economics 11e*. Mason: South-Western Cengage Learning.
- Atkinson, A. C., Koopman, S. J., and Shephard, N. (1997). Detecting shocks: Outliers and breaks in time series. *Journal of Econometrics*, (80), 387-422. Retrieved from <https://www.sciencedirect.com/science/article/pii/S030440769700050X>
- Belsely, D. A., Huh, E., and Welsch, R. E. (2004). *Regression diagnostics. Identifying influential data and sources of collinearity*. Hoboken: John Wiley & Sons.
- Bonham, C. D. (1971). Testing for outlying observations in a sample group. *Journal of Range Management*, 24(4), 310-312. doi:10.2307/3896952
- Chatterjee, S., and Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, 1(3), 379-393. Retrieved from <http://www.mathstat.ualberta.ca/~wiens/stat578/papers/Chatterjee%20&%20Hadi.pdf>
- Cook, D. R. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15-18. doi:10.2307/1268249
- Cook, D. R. (1980). *Identification of outliers*. Chapman & Hall. doi:10.1007/978-94-015-3994-4
- Cousineau D., and Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1), 58-67.
- Dittmann, P., Dittman, I., Szabela-Pasierbińska, E., and Szpulak, A. (2009). *Prognozowaniu w zarządzaniu przedsiębiorstwem*. Kraków: Wolters Kluwer Polska Sp. z o.o.
- Hoaglin, D. C., and Welsch, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17-22. Retrieved from <http://www.stat.ucla.edu/~cocteau/stat201b/handout/hat.pdf>

- Kannan Senthamarai, K., and Manoj, K., (2015). Outlier detection in multivariate data. *Applied Mathematical Sciences*, 47(9), 2317-2324
- Oesterreich, M. (2017). Symulacyjne badanie wpływu liczby i rozmieszczenia luk na dokładność prognoz w szeregu czasowym dla danych dziennych. *Ekonometria*, 1(55), 57-68. doi:10.15611/ekt.2017.1.05
- Osborne, J. W., Overbay, A. (2004). The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessment, Research & Evaluation*, 9(6). Retrieved from <https://pareonline.net/getvn.asp?v=9%26n=6>
- Paul, R. K. (2018). *Some methods of detection of outliers in linear models*. Retrieved from IASRI: <http://iasri.res.in/seminar/AS-299/ebooks/2005-2006/Msc/trim1/4.%20Some%20Methods%20of%20Detection%20of%20Outliers%20in%20Linear%20Regression%20Model-Ranjit.pdf>
- Tsay, R. S. (1988). Outliers, level shifts and variance changes in time series. *Journal of Forecasting*, (7), 1-20. doi:10.1002/for.3980070102
- Watson, S. M., Clark, S., Tight, M., and Redfern, E. (1992). An influence method for outliers detection applied to time series traffic data. *ITS Working Paper*, (365). Retrieved from http://eprints.whiterose.ac.uk/2206/1/ITS365_WP365_uploadable.pdf
- Watson, S. M., Tight, M., Clark, S., and Redfern, E. (1991). Detection of outliers in time series. *ITS Working Paper* (362). Retrieved from http://eprints.whiterose.ac.uk/2209/1/ITS261_WP362_uploadable.pdf
- Zahari, S. M., Zainol, M. S., Sopian, K. B., Zaharim, A., and Ibrahim, K. (2010). Additive outliers (AO) and innovative outliers (IO) in GARCH (1, 1) processes. *AMERICAN-MATH'10 Proceedings of the 2010 American conference on Applied mathematics*, 471-479. Retrieved from https://www.researchgate.net/publication/228664831_Additive_outliers_AO_and_innovative_outliers_IO_in_GARCH_1_1_processes
- Zawadzki, J. (1999). *Ekonometryczne metody predykcji dla danych sezonowych w warunkach braku pełnej informacji*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Zawadzki, J. (2003). *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*. Szczecin: Wydawnictwo Akademii Rolniczej w Szczecinie.

O METODZIE IDENTYFIKACJI OBSERWACJI NIETYPOWYCH W SZEREGACH CZASOWYCH

Streszczenie: W pracy zaproponowano metodę wykrywania obserwacji nietypowych w szeregach czasowych z wahaniami sezonowymi oraz bez tych wahań. Jej istota jej polega na badaniu wpływu poszczególnych obserwacji szeregu na wartości teoretyczne modelu oraz wielkości prognoz zbudowanych na jego podstawie. Egzemplifikacją rozważań o charakterze teoretycznym jest przykład empiryczny dotyczący modelowania i prognozowania dziennej sprzedaży paliw płynnych na stacji paliw X w latach 2012-2014. Dane za okres od 1.01.2012 do 30.06.2014 stanowią okres estymacyjny, a za II półrocze 2014 r. okres empirycznej weryfikacji prognoz. Wyniki otrzymane za jej pomocą zostały porównane z wynikami uzyskanymi innymi metodami służącymi do identyfikacji obserwacji wpływowych oraz odstających, w tym m.in.: reszt standaryzowanych, odległości Cooka oraz DFFIT. Obliczenia przeprowadzono w środowisku R oraz pakiecie Statistica.

Słowa kluczowe: prognozy, identyfikacja, regresja wieloraka, szeregi czasowe, obserwacje odstające.

Quote as: Oesterreich, M. (2020). On the method of identification of atypical observations in time series. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, 24(2).