A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 286, 2013

Dominika Polko[*], *Grzegorz Kończak*[**]

ON TIME SERIES PREDICTION BASED ON CONTROL CHART

Abstract. Control charts are the most commonly used quality control tools. These tools are dedicated to monitoring processes characteristic over time. Control charts may be successfully applied in other statistical areas. The non-classical use of control charts for time series prediction has been presented by Z. Pawłowski in the paper *Predykcja za pomocą kart kontrolnych* (*Control Chart Based Prediction*, 1969). The forecasts obtained by this method are quantitative or qualitative. The modification of this method is presented in the paper. It leads to quantitative predictions in all cases. The proposal was compared to some well-known classical prediction methods in the Monte Carlo study.

Key words: time series, prediction, control chart.

I. INTRODUCTION

Control charts be statistical tools for detecting shifts in the process parameters. Control charts are based on the assumption that the process data are normally distributed and the monitored characteristic is independently distributed (there is no autocorrelation in the production process). The central line is located at a level equal to the mean value of the monitored characteristic. In addition, there could be plotted two control limits (upper and lower) and two warning lines (upper and lower).

Control charts are designed to detect instability of the monitored processes. There are some rules that help to detect shifts or trends in the analyzed time series. Some of them are as follows (D.C. Montgomery, 1996):

- one point plots above (below) the upper control limit (the bottom control limit),
- two out of three consecutive points plot beyond the two-sigma warning lines,
- a run of *r* consecutive points (usually eight points) plot on one side of the center line (above or below),

[*] MA, Department of Statistics, Katowice University of Economics, dpolko@gmail.com.
[**] Ph.D., Associate Professor, Department of Statistics, Katowice University of Economics, grzegorz.konczak@ue.katowice.pl.

• a run of r consecutive points (usually eight points) plot in an increasing (decreasing) trend.

There are many other sets of rules that can be used in monitoring processes. Lists of various types of such rules are presented in D.C. Montgomery (1996) and G. Kończak (2007). This article will concentrate only on the rules considered in the paper by Z. Pawłowski (1969).

II. PREDICTION BASED ON CONTROL CHARTS

Let $y_1, y_2, \ldots, y_k$ are the consecutive time series observations. Z. Pawłowski (1969) proposed a time series prediction method that is based on the construction of control chart. He proposes to evaluate indexes w_t, and on the basis of these results, he uses the process control chart to predict the value of the next period. The index of period t can by written using the formula

$$w_t = \frac{y_t}{y_{t-1}}, \text{ for } t = 2, 3, \ldots \quad (1)$$

Z. Pawłowski suggests plotting the warning lines at the level equal to the mean of the process plus the value of the standard deviation (the upper warning line) and the mean minus the value of the standard deviation (the lower warning line) and the upper and the lower control limits respectively at the following levels: the mean plus doubled value of the standard deviation and the mean minus doubled value the standard deviation. Since the monitored characteristic is an index (the ratio of two successive measurements) the level of the control and warning lines has to be estimated. It can be made on the basis of a large sample taken from a stable process. Location of successive points on the control chart determines the prediction method for the next period. Z. Pawłowski suggested the following possible options (variants) of the forecasts for the *t+1* period construction:

a) if successive points corresponding to the last r measurements are between the warning lines then it should be concluded that the process is stable and the forecast could be stated at the level of the average of the last r measurements (r is the number of recent points between the control lines), that is for $r \geq 1$

$$y_{t+1,p} = \frac{1}{r}\sum_{i=0}^{r-1} y_{t-i} \quad (2)$$

b) two successive points that are above the upper (lower) warning line indicate a tendency of an increase (or decrease) and the forecast should be described in the following form

$$y_{t+1,p} > \frac{1}{r}\sum_{i=2}^{r+1} y_{t-i} \text{ for } r \geq 1. \quad (3)$$

c) a single point above the upper (lower) control limit determines the forecasts in the following form

d)

$$y_{t+1,p} > \frac{1}{r}\sum_{i=1}^{r} y_{t-i} \quad r \geq 1 \quad (4)$$

e) r points on one side of the center line. In this case Z. Pawłowski defines three possible variants of prediction depending on the number of points. A run of points above the central line leads to three possible formulas of the prediction

$$y_{t+1,p} > y_t,\ y_{t+1,p} = y_t\overline{w} \text{ or } y_{t+1,p} > y_t\overline{w} \text{ where } \log\overline{w} = \frac{1}{r}\sum_{i=0}^{r-1}\log w_{t-i} \quad (5)$$

Moreover Z. Pawłowski indicates the possibility of two consecutive values of indexes in periods *t-1* and *t* relatively small and relatively large (or alternatively large and small), which leads to the detection of a turning point and the forecast has the form $y_{t+1,p} > \frac{1}{r}\sum_{i=1}^{r} y_{t-i}$ $\left(y_{t+1,p} < \frac{1}{r}\sum_{i=1}^{r} y_{t-i}\right)$ for $r \geq 1$. This paper concentrates on the systematic process changes detection in time series, therefore this case will not be considered.

If the process is stable then the forecast should be calculated based on the moving average formula given in (2). In this paper only the qualitative forecast will be considered (cases in point b to d). If the hypothesis that the process is stable is rejected then an increase or a decrease of the value of the observed characteristics is expected.

III. MODIFICATION OF QUALITATIVE PREDICTIONS

It could be seen that some of the cases indicating the changes described by Z. Pawlowski may occur at the same time. It is possible for example plots two points above the upper warning line where the last one is above the upper control limit. Taking into account the relationship between the arithmetic mean ($\bar{x}$) and the geometric ($\bar{x}_G$) expressed by inequality $\bar{x}_G \le \bar{x}$, the forecast will be set for all possible types of signals listed above. For the stable process the moving average (arithmetic mean) as a forecast should be calculated. The case of crossing the control limit can be used the geometric mean as a forecast. On this basis, the following predictions of quantitative forecasts in place of qualitative forecasts

a) no signals – the forecast can be calculated on the basis of a moving average (2)

b) two points in a row plots above the warning limit or four points in a row above the central line should be indicated as a dynamic stabilization and this allows to forecast on the basis of the moving averages the last *r*-1 values, that is

$$y_{t+1,p} = \frac{1}{r-1}\sum_{i=0}^{r-2} y_{t-i} \tag{6}$$

c) points plotted above the control limit and at the same time one of the other two signals (two consecutive points above the warning line or four consecutive points above the center line) suggest that crossing the upper control line was the result of random fluctuations and the value of the forecast is to be somehow reduced using the geometric mean, that is

$$y_{t+1,p} = \sqrt[3]{y_{t-2}y_{t-1}y_t} \tag{7}$$

d) the occurrence of all three considered signals simultaneously informs about systematic changes and suggests forecasting using the last (highest) values, that is

$$y_{t+1,p} = \sqrt{y_{t-1}y_t} \tag{8}$$

The described forecasts were compared to the forecast obtained by the naive, the moving average, and the exponential smoothing methods (see Cieślak et all, 1993, Kohler and College, 1988) in the Monte Carlo study.

IV. MONTE CARLO STUDY

The comparison of the proposed method and some well-known forecasting methods has been done in the computer simulation study. One hundred thousand replications of 50 observations of the random variable Y_t from the normal distribution were generated. The first 10 observations (t=1,2,...,10) were generated from the normal distribution $N(m, \sigma)$, where m = 100 and σ = 1. The next 40 observation (t =11,12,...,50) were generated from the normal distribution where m was changed. There were 4 variants of the changes considered. The first one is associated with a step change in the expected value (I), the second one with a linear change in the average level (II), while the last two of the different forms of quadratic function of trend deflection (III and IV).The formulas of the considered random variables could be written in the following form

I) $Y_t \sim N(m+\sigma,\sigma)$

II) $Y_t \sim N\left(m+\frac{t-10}{40}\sigma,\sigma\right)$

III) $Y_t \sim N\left(m+\left(\frac{t-10}{40}\right)^2\sigma,\sigma\right)$

IV) $Y_t \sim N\left(m+\sigma-\left(\frac{50-t}{40}\right)^2\sigma,\sigma\right)$

where t =11,12,…,50.

The expected values of the considered random variables Y_t are schematically presented in Figure 1.

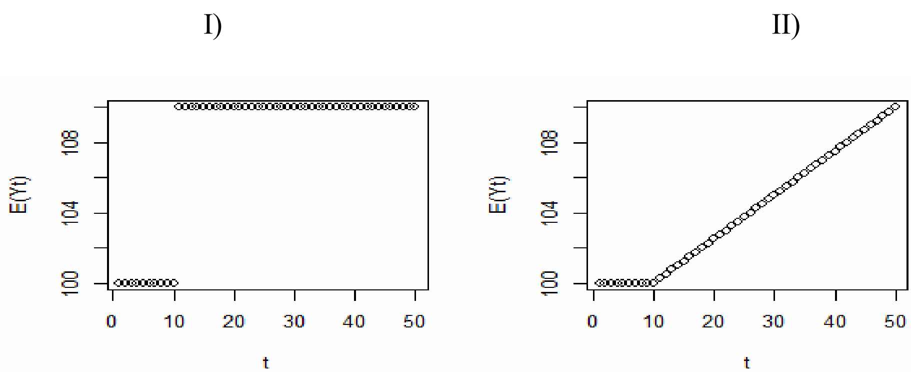

Fig. 1. The expected value of a random variable Y_t for 4 variants of parameter changes

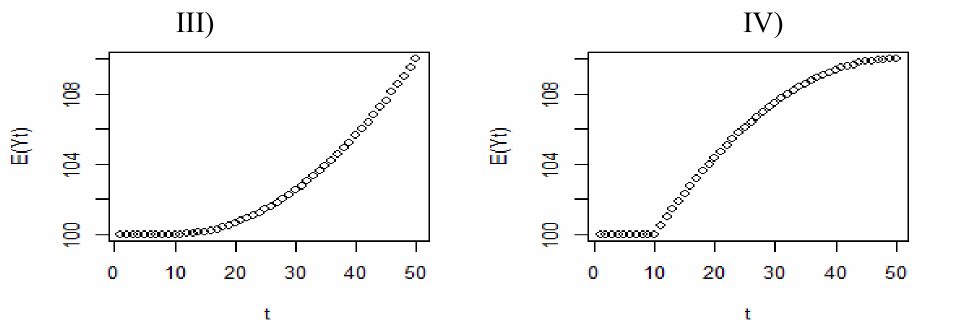

Fig. 1. The expected value of a random variable Y_t for 4 variants of parameter changes (cont.)

100 000 simulations were performed for each described case. The prediction errors MSE were estimated and the results are presented in Table 1. The assessments of errors for forecasts were made on the basis of the methods (see Cieślak et all, 1993, Wywiał, 1995): the naive, the exponential smoothing and the moving averages were presented in the same table.

There were used following formulas:

- naive: $y_{t+1,P}^{(n)} = y_{t,P}$,
- exponential smoothing: $y_{t+1,P}^{(w)} = \alpha y_{t,P} + (1-\alpha) y_{t,P}^{(w)}$, where $y_{1,P}^{(w)} = y_1$,
- moving averages: $y_{t+1,p}^{(r)} = \frac{1}{r}\sum_{i=0}^{r-1} y_{t-i}$ (in the simulations $r = 4$).

Table 1. Estimation of prediction errors

Prediction method	Variant of changes in time series			
	I	II	III	IV
Naive $y_{t+1,P}^{(n)}$	4.499	2.062	2.081	2.084
Exponential smoothing $y_{t+1,P}^{(w)}$	4.269	1.763	1.793	1.795
Moving average $y_{t+1,P}^{(r)}$	5.932	1.626	1.731	1.749
Proposed method $y_{t+1,P}$	5.594	1.622	1.72	1.736

Source: Monte Carlo study

The comparison results of the considered forecasting methods are presented in Figure 2. Computer simulation showed that in the case of a linear and quadratic form of changes in time series, the proposed method results lead to smaller errors than the estimates determined on the basis of a naive method, exponential smoothing method and the moving average method.

The Monte Carlo study has shown that the results of the proposed method in general are similar to the results obtained by moving average prediction. In the first case of the analyzed changes (step change – Fig. 1) the results of the proposal are a bit worse that those obtained by the naïve method and by the exponential smoothing method.

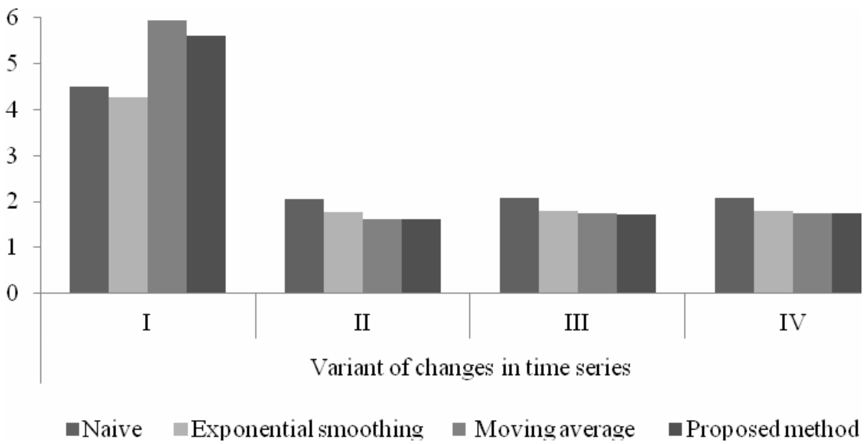

Fig. 2. Estimation of prediction errors for 4 variants of the parameter changes

The greatest advantage of the proposal is that it is easy to use. This method could be expanded by adding more rules which are commonly used in monitoring processes (see D.C. Montgomery, 1996). Depending on the anticipation of various types of changes in time series the different sets of control chart rules could be used.

V. CONCLUDING REMARKS

This paper presents a proposal of modification of a forecasting method using control charts presented by Z. Pawłowski. The forecasts obtained by this method are quantitative or qualitative. The proposed modification in each case leads to a quantitative prediction. The forecast is based on a moving average and for nonstable processes it uses the arithmetic or geometric mean of the corresponding number of the last observation. The proposal was compared to some well-known classical prediction methods in the Monte Carlo study.

REFERENCES

Cieślak M., Dittman P., Kania-Gospodarowicz A., Krupowicz J., Kwiatkowska D., Radzikowska B. (1993) *Prognozowanie gospodarcze*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.

Kohler H., College A. (1988) *Essentials of Statistics*, Scott, Foresman & Company, Glenview, Illinois.

Kończak G. (2007) *Metody statystyczne w sterowaniu jakością produkcji*, Wydawnictwo Akademii Ekonomicznej w Katowicach.

Montgomery D.C. (1996) *Introduction to Statistical Quality Control*, John Wiley & Sons, Inc., New York.

Pawłowski Z. (1969) *Predykcja za pomocą kart kontrolnych*, Przegląd Statystyczny, Tom XVI, Zeszyt 3-4, str. 203-220.

Wywiał J. (1995) *Weryfikacja hipotez o błędach predykcji adaptacyjnej*, Wydawnictwo Ossolineum, Wrocław.

Dominika Polko, Grzegorz Kończak

O PROGNOZOWANIU SZEREGÓW CZASOWYCH Z WYKORZYSTANIEM KART KONTROLNYCH

Metody statystyczne opracowane na potrzeby kontroli jakości produktów z powodzeniem mogą być stosowane w analizie innych zagadnień. Do najczęściej wykorzystywanych narzędzi kontroli jakości należy zaliczyć karty kontrolne. Nieklasyczne zastosowanie kart kontrolnych związane z wykorzystaniem ich do prognozowania przedstawił Z. Pawłowski w artykule *Predykcja za pomocą kart kontrolnych* (1969). Prognozy otrzymywane tą metodą mają charakter ilościowy lub jakościowy. W artykule przedstawiono propozycję modyfikacji tej metody w celu uzyskania wszystkich prognoz o charakterze ilościowym. Proponowaną metodę porównano symulacyjnie z wybranymi metodami predykcji.