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 3(302), 2014

*Magdalena Chmielińska**

USING STABLE DISTRIBUTIONS FOR MONITORING PROCESSES, WITH UNKNOWN DISTRIBUTIONS

Abstract. The control chart is a tool of statistical quality control widely used in production. The fulfillment of its basic assumptions, guarantees flawless assessment of correctness of the monitored process. The purpose of this paper is to pay attention to the need to verify the assumptions of the used method and the effects of its unauthorized use, in case of not meeting its assumptions. In the paper a method that uses a family of stable distributions to estimate the unknown probability density of monitored diagnostic variable, is proposed. The estimated density function is the basis for determining the control limits.

Keywords: Control chart, stable distribution

I. INTRODUCTION

The control chart which is a simple and effective tool for statistical quality control is widely used in factories to monitor manufacturing processes. Uniqueness of the diagnosis process guarantees the fulfillment of its basic assumptions. Incompatibility of the empirical distribution of variable diagnostic to normal distribution is sometimes a source of incorrect assessment the correctness of the process being monitored. This fact makes it necessary to search for methods resistant form of the empirical distribution of the diagnostic variable. The purpose of this paper is to pay attention to the need to verify the assumptions of the used control chart and the effects of its unauthorized use, in case of not meeting its assumptions. The accepted hypothesis states that using a family of stable distributions to estimate the unknown probability density of monitored diagnostic variable increases the effectiveness of the monitoring of the production process.

II. CONTROL CHART – BASIC INFORMATION

The control chart as a tool of statistical quality control, allows to monitor the manufacturing process by providing information on its progress. Its characteristic features are the efficiency and simplicity of design. Classical

* M.Sc., Department of Statistic, Katowice University of Ecomonics.

control chart based on the assumption that the variable diagnostic observed during the monitoring process has a normal distribution with expected value μ and standard deviation σ (Kończak, 2007). It also requires independent measurements in subsequent time periods. The control chart is a plot designed to record the results of ongoing quality control of products containing the central line, the control lines and the warning lines. Control lines are determined at such a level that the probability of occurrence of values below the lower or above the upper line of control, in case of the in-control process, is appropriately small.

Inference about the correctness of the process is based on the analysis of the points plotted on a control chart in relation to the control and warning lines. In this work, however, as the signal of process disturbance is considered only a point above (below) the upper (lower) control line.

Characteristics of control charts is the ARL (the average run length). Distinguished ARL_0 used for process being in – control, and ARL_1 for process being out – of – control (Coleman et al. 2008). ARL_0 is the inverse of the probability of making a mistake I - type, ie the inverse of the probability of the signal of process disturbance for process of the correct course. This relationship is given by the formula:

$$ARL_0 = \frac{1}{\alpha} = \frac{1}{p_s} \tag{1}$$

where:

α – the probability of an error I - it kind,

p_s – the probability of signal for a single observation (sample).

In the case of process being out – of – control the average run length is the expected number of samples examined before detection of disturbance. ARL_1 for the process of dysregulated statistically expressed by the formula:

$$ARL_1 = \frac{1}{1-\beta} \tag{2}$$

where β is the probability of not detecting the disturbance in the first sample after the disturbance.

III. STABLE DISTRIBUTIONS – BASIC INFORMATION

The class of stable distributions was introduced during investigations of the behavior of sums of independent random variables. (Borak, 2005). A complete description of the α – stable distribution is possible using four parameters: the shape parameter α, the skewness index $\beta \in \langle -1,1 \rangle$, the scale parameter $\gamma > 0$ and the location parameter $\delta \in \Re$. The most important of them is the shape parameter $\alpha \in (0,2]$, also called the tail index, exponent characteristic, or the index of stability. α – stable distributions can be defined using the characteristic function and the inverse Fourier transform.

For a random variable with α – stable distribution, the characteristic function is given by (Trzpiot, 2010):

$$\ln \varphi(t) = \begin{cases} i\delta t - \lambda^{\alpha} |t|^{\alpha} \left[1 - i\beta \operatorname{sgn}(t) tg\left(\frac{\alpha\pi}{2}\right)\right], & \text{for } \alpha \neq 1 \\ i\delta t - \lambda |t| \left[1 + i\beta \operatorname{sgn}(t) \frac{2}{\pi} \ln|t| \right], & \text{for } \alpha = 1 \end{cases}$$

where sgn (t) is the function of the trade.

All distributions belonging to the class of stable distributions are continuous distributions, but for the most part they do not have the analytical form of the density function. They are known – we can save them for the following three special cases:

– the normal distribution when $\alpha = 2$, β *insignificant,*
– the Cauchy distribution, when $\alpha = 1$, $\beta = 0$,
– the Levy distribution of when $\alpha = \frac{1}{2}$, $\beta = 1$.

IV. DATA DESCRIPTION

The underlying data presented in this paper's considerations are shared by a production plant operating in the province of Silesia and representing the automotive industry. These data contain information about the values of certain characteristics of an item produced there and obtained at the stage of the middle operating control. They include information about inspections conducted during the period from 10.09.2012 to 15.09.2012, which together make up 10, 969 records.

During this period, the analyzed production process can be considered as statistically regulated, since all the values obtained during the audit, were set in norms for the controlled characteristics. The analyzed data can be treated in the same way as the data from the process of the correct course. Therefore, these data can be used to determine the control limits of the test sheet.

V. ANALYSIS OF PRODUCTION ACCURACY

The analysis of the production process, which is enabled by control charts, should start by checking (usually based on historical data) the assumptions of the method. The data shared and analyzed in this study are characterized by a distribution significantly different from a normal distribution, which is shown by Chmielińska.

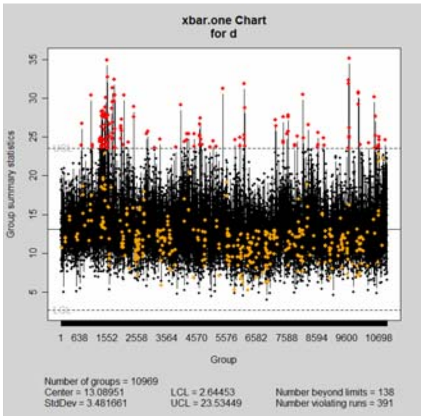

Figure 1. The classical control chart plotted for analyzed data

Source: Own elaboration.

The classical control chart removed for the analyzed data (Fig. 1), due to the lack of compatibility of distribution of tested diagnostic variable with a normal distribution, leads to incorrect assessment of the correctness of the production process. It shows several times to dysregulation of the production process – a lot of points in the field above the upper control line – when it is known that this process runs smoothly.

In the case of diagnostic variable of distribution significantly different from a normal distribution, to eliminate the risk of an erroneous assessment of the accuracy of the manufacturing process, the control limits should be set in a different way. This problem is the subject of considerations of among others Chakraborti et al (2004), Chmielińska.

In this work, the possibility of the use of a class of stable distributions for estimating the probability distribution of values of the unknown tested variable diagnostic and to request a proper conduct of the manufacturing process being analyzed based on the obtained probability distribution function of the form.

For shared data empirical probability distribution of the value of the diagnostic variable are approximated by stable distribution with parameters $\alpha = 1.719$, $\beta = 0.999$, $\gamma = 2.054$, $\delta = 12.328$. Matching the estimated distribution to the empirical distribution is shown in Figure 2 and the results of the compatibility test are shown in Table 1.

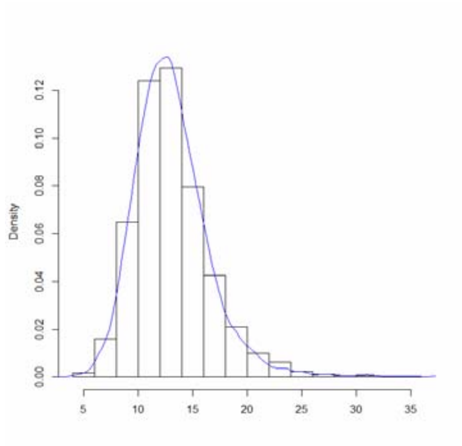

Figure 2. The histogram of analyzed data with the curve of matched to the data stable

Source: Own elaboration.

The Anderson – Darling compatibility test leads to the conclusion that there is no reason to reject the hypothesis of compatibility of the empirical distribution of the postulated α – stable distribution. For this reason, the estimated probability function can be used to monitor the production process, setting the control limits at the level of the corresponding quantile of the estimated distribution $(q_{0,001}, q_{0,999})$.

Table 1. The Anderson – Darling compatibility test result for the empirical distribution of the data and matched to the data stable distribution

	Statistic	*p*-Value
Anderson - Darling	1,28012	0,238907

Source: Own elaboration.

Prepared in this way control chart, plotted for the analyzed data is shown in Figure 3. It is clear that the upper control limit has been overestimated. Its too high value is due to a higher probability of extremely high value in the used distribution than a normal distribution, which indicates a parameter $\alpha = 1.71882$ taking the value of less than 2, and the fact that the distribution is characterized by a strong asymmetry of right-hand ($\beta = 0.999$).

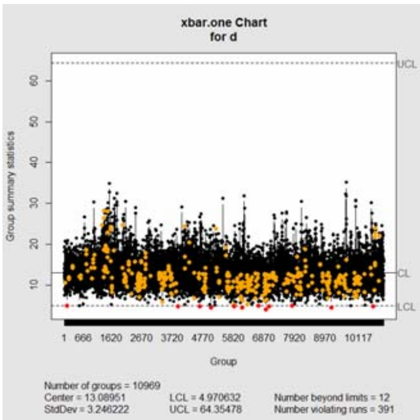

Figure 3. The control chart based on quantile of matched to the data stable distribution.

Source: Own elaboration.

Figure 4 shows quantile – quantile graph of compared distributions. It confirms all the previously drawn conclusions: the relative goodness of fit of the estimated probability function for the empirical distribution of the data and the disparity between the values of quantiles higher orders.

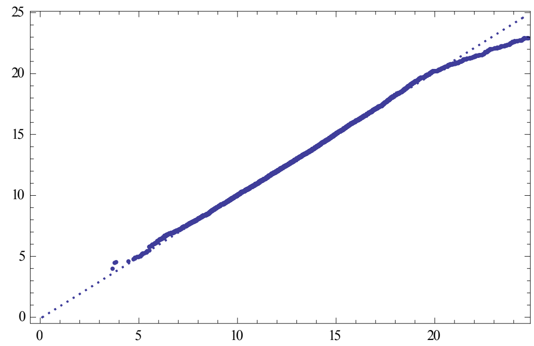

Figure 4. The quantile – quantile graph of empirical distribution of analyzed data and matched to the data stable distribution

Source: Own elaboration.

The use of the stable distribution matched to the output data to determine the levels of control lines causes the proposed method, in this case, to be ineffective. The analyzed control chart (Figure 3) indicates 12 times dysregulation of the process, the same giving the probability of the signal level $\left(p_s = \dfrac{12}{10969} = 0{,}0011\right)$ and ARL for process being in – control (in this case we deal with this kind of process) at the level of $ARL_0 = \dfrac{1}{0.0011} = 914$. The value of the probability of the signal and the value of ARL_0 for the analyzed chart seem to be better than those, that guarantee the classical control chart under its assumptions ($(p_s = 0.013)$, $ARL_0 = 370$) to which during constructing the non-classical method it is pursued. However, the high value ARL_0 of the proposed method is the result of incorrectly specified level of the upper control line. The idea of control chart permits for process being in – control for the one in the 1000 registered values were above the upper (lower) control limit, therefore, in non-classical methods it is based on the value of quantile of order of 0,999 (about 0,001) adopted probability distribution. Bad fit the tails of the empirical distribution of the data and the estimated probability distribution makes that the control chart determined based on the matched to data stable distribution does not have this feature. In addition, the adoption of border control at such a high level (much higher than the maximum of empirical data) causes a hypothetical deregulation of the manufacturing process, not to be detected on time.

VI. CONCLUSION

Process monitoring which has an unknown distribution significantly different from the normal distribution should be carried out using an appropriate non-classical method. Probability density function which is the basis of this method must be well matched to the empirical data and take into account all the properties of the empirical distribution of the diagnostic variable.

Considered in this article, the control chart based on matching (good) to data stable distribution, has proved to be ineffective. The empirical distribution of the analyzed diagnostic variables is characterized by right-sided asymmetry, but the force of asymmetry is significantly less than the matched to data a stable distribution. The differences in the thickness of the right tail of both distributions caused that the upper control line was set at the wrong level. Removed in this way, control chart, due to the overvaluation of the upper limit of the control should not continue to be used to monitor the analyzed production process, as any of its dysregulation has been signalled to the expected short time.

However, the considered method can be effective in the case of monitoring a process whose unknown diagnostic variable is characterized by a distribution of strong asymmetry.

ACKNOWLEDGEMENTS
The research was supported by Polish National Science Centre grant DEC-2011/03/B/HS4/05630

REFERENCES

Borak S., Hardle W., Weron R. (2005): *Stable Distributions,* SFB 649 Discussion Papers, Springer, Berlin.

Chakraborti S, van der Laan P., van de Wiel M. A. (2004): A Class of Distribution-free Control Charts, *Journal of the Royal Statistical Society*, No. 3, Vol. 53, s. 443-462.

Chmielińska M.: Wykorzystanie estymacji jądrowej do monitorowania procesów o nieznanym rozkładzie. W druku.

Coleman S., Greenfield T., Stewardson D., Montgomery D. C. (2008): *Statistical Practice in Business and Industry,* John Wiley&Sons, Inc, New York.

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

Trzpiot G. (red.) (2010): *Wielowymiarowe metody statystyczne w analizie ryzyka inwestycyjnego,* PWE, Warszawa.

Magdalena Chmielińska

WYKORZYSTANIE ROZKŁADÓW STABILNYCH DO MONITOROWANIA PROCESÓW O NIEZNANYM ROZKŁADZIE

Karta kontrolna jest powszechnie stosowanym w zakładach produkcyjnych narzędziem statystycznej kontroli jakości. Spełnienie podstawowych jej założeń, gwarantuje bezbłędną ocenę poprawności monitorowanego procesu produkcyjnego.

Celem niniejszej pracy jest zwrócenie uwagi na konieczność weryfikacji założeń stosowanej metody i skutki nieupoważnionego jej stosowania, w przypadku braku ich spełnienia. W pracy proponuje się metodę wykorzystującą rodzinę rozkładów stabilnych do szacowania nieznanej gęstości prawdopodobieństwa monitorowanej zmiennej diagnostycznej. Oszacowana funkcja gęstości stanowi podstawę wyznaczania granic kontrolnych.