Raport Badawczy
Research Report

RB/17/2007

Sequential signals on a control chart based on nonparametric statistical tests

O. Hryniewicz, A. Szediw

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2007

Sequential Signals on a Control Chart Based on Nonparametric Statistical Tests

Olgierd Hryniewicz, Anna Szediw

Systems Research Institute, ul. Newelska 6, 01-447 Warsaw, Poland
{hryniewi, szediw}@ibspan.waw.pl

1 Introduction

Statistical process control (SPC) is a collection of methods for achieving continuous improvement in quality. This objective is accomplished by continuous monitoring of the process under study in order to quickly detect the occurrence of assignable causes. The Shewhart control chart, and other control charts - like CUSUM, MAV, and EWMA - are the most popular SPC methods used to detect whether observed process is under control. Their classical, and widely known, versions are designed under the assumption that process measurements are described by independent and identically distributed random variables. In the majority of practical cases these assumptions are fulfilled at least approximately. However, there exist production processes where consecutive observations are correlated. This phenomenon can be frequently observed in chemical processes, and many other continuous production processes (see Wardell et al. [25] or Alwan and Roberts [2] for examples). The presence of correlations between consecutive measurements should be taken into account during the design of control charts. This need was noticed in the 1970s, see for example the papers by Johnson and Bagshaw [9] and by Vasilopoulos and Stamboulis [24], but the real outburst of papers related to this problem took place in the late 1980s and in the 1990s.

There exist several approaches for dealing with serially correlated SPC data. First approach, historically the oldest one, consists in adjusting control limits of classical control charts. This approach was used, for example, in papers [11],[12],[13],[18],[19],[23],[24],[30]. The second approach, which is represented by a seminal paper by Alwan and Roberts [1], profits from the knowledge of the correlation structure of a measured process. Alwan and Roberts [1] propose to chart so called residuals, i.e. differences between actual observations and their predicted, in accordance with a prespecified mathematical model, values. This approach has its roots in the statistical analysis of time series used in automatic control, originated by the famous book of Box and Jenkins [3]. Among many papers on control charts for residuals we can mention e.g. the following: [11],[12],[13],[16],[19]. Lu and Reynolds [11],[12],[13] compared control charts for original observations (with corrected limits) with control charts for residuals. They have shown that the best practical results may be obtained while using simultaneously EWMA (or CUSUM) modified charts for original observations and Shewhart charts for residuals. According to the third approach, introduced in the area of SPC by Yourstone and Montgomery [28],[29], a process is monitored by charting statistics, such as the coefficient of serial correlation, that reflect the correlation structure of the monitored process. Another original approach of this type was proposed in the paper by Jiang et al. [8] who proposed a new type of a chart for monitoring of autocorrelated data an ARMA chart.

SPC procedures for autocorrelated data have been usually proposed for charting individual observations that are typical for continuous production processes. Relatively few papers have been proposed for the analysis of SPC procedures in case of autocorrelation within the sample. A good overview of those papers together with interesting original results can be found in the paper by Knoth et al. [10]. Another important problem which has attracted only few authors is related to the control of short-run processes. Some interesting results in this area can be found in papers [21],[27].

The number of papers devoted to the problem of charting autocorrelated processes is quite large. Therefore the readers are encouraged to look at those papers for further references. The review of first papers devoted to the problem of SPC with correlated data can be found, for example, in the aforementioned paper by Wardell et al. [25], and in a short overview paper by Woodall and Faltin [26]. A good overview of the papers published in 1990s can be found in [10],[12].

While discussing different SPC methods used for the analysis of correlated data we have to take into account their efficiency and as it was pointed out by Lucas [14] in the discussion of [25], simplicity. The results obtained by Wardell et al. [25] for the case of Auto Regressive Moving Average (ARMA) time series that describe correlated measurements have shown that Alwan Roberts type control charts for residuals in certain cases of positive correlation may be outperformed even by a classical Shewhart control chart with unmodified control limits. However, in other cases the application of these rather complicated procedures (their usage requires the software for the analysis of time series) may be quite useful. Also the results obtained by Timmer et al. [22] show that in the case of Auto Regressive AR(1) processes the application of simple control charts based on the serial autocorrelation coefficient as the only SPC tools may be not effective.

While dealing with correlated data we cannot rely, even in the case of classical control charts, on the methods used for the estimation of their parameters in case independent observations. Some corrections are necessary, as it was mentioned e.g. in the paper by Maragah and Woodall [15]. Another problem with the application of the procedures designed to control autocorrelated data is the knowledge of the structure of correlation. In the majority of papers it is assumed that the type of the stochastic process that describes the process data is known. Moreover, it is also assumed that the parameters of this stochastic process are also known. However, Lu and Reynolds [12],[13] have shown that precise estimation of such parameters requires at least hundreds of observations. Taking into account that all computations required for designing and running SPC procedures for autocorrelated data are not easy for an average practitioner, the problem arises then: how to verify the hypothesis of correlation in a simple way? The answer to this question is very important, as it indicates the amount of possible future difficulties with running SPC procedures. It is quite obvious that practitioners would like to avoid these problems as it can be only possible. The simplest solution to this problem is to use the serial autocorrelation control chart, as it was proposed in [28],[29]. However, the coefficient of serial autocorrelation performs well for processes with Gaussian random error components. For more general processes specialists in time series analysis suggest to apply nonparametric statistical tests. An interesting review of such tests can be found in the paper by Hallin and Mélard [7]. Unfortunately, the majority of those tests are either complicated or unsuitable for process control. However, one of the recent papers [5] on the application of Kendall's tau statistics for testing serial dependence seems to be promising in the context of SPC. We consider this possibility in this paper.

In the second section of the paper we present some basic information on the Kendall's tau statistic when it is used for the analysis of autocorrelated data. Using basic properties of this statistic we propose a relatively simple control chart based on this statistic. Statistical properties of this control chart have been investigated using Monte Carlo simulations. In the third section of the paper we present the results of Monte Carlo experiments in the case when two consecutive observations are described by a two-dimensional normal distribution, i.e. in the case of a simple autoregressive model. We compare the behavior of our chart with the behavior of a chart based on the serial autocorrelation coefficient. In the fourth section we analyze the results of simulations when the dependence structure is more complicated. We consider the case when two consecutive observations are described by a two-dimensional copula with different marginal probability distribution functions. In the last section of the paper we formulate some conclusions, and propose the directions for further investigations.

2 Control chart based on Kendall's tau statistic for serially autocorrelated data

Many of problems experienced when applying traditional SPC to monitoring processes are caused by the violation of the basic assumption of statistical independence of consecutive observations. However, in practice this condition is very often not fulfilled, and consecutive observations are correlated. It should be stressed that some small disturbances of independence conditions may be natural and even desirable (e.g. in case of the existence of favourable trends of process parameters). However, in the majority of cases autocorrelation of process parameters should be considered either as an obstacle in monitoring the process or even as unwanted feature, when it increases process variation. In such situation special statistical methods for detecting dependencies (autocorrelations) between consecutive process observations are strongly recommended. For this purpose we propose to use the Kendall's τ statistic, which is a fundamental statistical measure of association.

Let $Z_1, Z_2, \ldots, Z_n$ denote a random sample of n consecutive process observations and (X_i, Y_i), where $X_i = Z_i$ and $Y_i = Z_{i+1}$ for $i = 1, 2, \ldots, n-1$ is a bivariate random vector. Then, the Kendall's τ sample statistic measuring the association between random variables X and Y is given by the following formula

$$\tau_n = \frac{4}{n-1} \sum_{i=1}^{n-1} V_i - 1, \tag{1}$$

where

$$V_i = \frac{card\{(X_j, Y_j) : X_j < X_i, Y_j < Y_i\}}{n-2}, i = 1, \ldots, n-1. \tag{2}$$

In terms of the original observations Kendall's tau can be represented as a function of the number of disconcordances M, i.e. the number of pairs (Z_i, Z_{i+1}) and (Z_j, Z_{j+1}) that satisfy either $Z_i < Z_j$ and $Z_{i+1} > Z_{j+1}$ or $Z_i > Z_j$ and $Z_{i+1} < Z_{j+1}$. In these terms we have

$$\tau_n = 1 - \frac{4M}{(n-1)(n-2)}, \tag{3}$$

where

$$M = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} I(Z_i < Z_j, Z_{i+1} > Z_{j+1}), \tag{4}$$

and $I(A)$ represents the indicator function of the set A. In case of mutually indepeudent pairs of observations (X_i, Y_i), $i = 1, 2, \ldots, n-1$ the probability distribution of (1) is well known. However, in case of time series, even in the case of mutual independence of $Z_1, Z_2, \ldots, Z_n$ pairs of observations (X_i, Y_i) become dependent, and the probability distribution of τ_n for small values of n has been obtained only recently [5]. Ferguson et al. [5] obtained precise probabilities $P_n(M \leq m)$ for $n = 3, \ldots, 10$, and approximate probabilities for $n > 10$. In Table 1 we present the probabilities of $\tau_n \geq \tau_{crit}$ for some selected small values of n. This type of presentation is more useful for the discussion of the applicability of Kendall's tau in SPC. First of all, from Table 1 it is clearly seen that except for the case $n = 6$ and $\tau_{crit} = 1$ it is not possible to construct a one-sided statistical test of independence against the alternative of positive dependence with the same probability of false alarms as in the case of a Shewhart control chart. By the way, this exceptional case is equivalent to a well known supplementary pattern test signal on a classical control chart: "six observations in a row are either increasing or decreasing". Moreover, all critical values of τ_n that are equal to one correspond to sequential signals of the type "n observations in a row are either increasing or decreasing", and for the values τ_{crit} that are close to one it is possible to formulate similar pattern rules. Thus, for small values of n it is in principle impossible to make precise comparisons of control charts based on Kendall's tau with classical three-sigma Shewhart control charts. Close investigation of the probability distribution of M presented in [5] shows that due to a discrete nature of the Kendall's tau this situation is the same in case of a two-sided test and also similar even for larger values of n.

Table 1. Critical values for Kendall's tau statistic in presence of dependence between pairs of observations

n=6		n=7		n=8		n=9	
τ_{crit}	$P(\tau_n \geq \tau_{crit})$	τ_{crit}	$P(\tau_n \geq \tau_{crit})$	τ_{crit}	$P(\tau_n \geq \tau_{crit})$	τ_{crit}	$P(\tau_n \geq \tau_{crit})$
1	0.00267	1	0.00042	1	0.00006	1	0.00001
0.8	0.00834	0.866	0.00119	0.904	0.00014	0.857	0.00007
0.6	0.03056	0.733	0.00477	0.809	0.00069	0.785	0.00021
		0.600	0.01356	0.714	0.00178	0.714	0.00071
				0.619	0.00565	0.642	0.00185
						0.571	0.00514

Consecutive values of τ_n are dependent even for independent original observations. Therefore, it is rather difficult to obtain the values of ARLs. In Table 2 we present such values, each based on over 1 million simulations in case of small ARLs and over 10 000 simulations for very large ARLs, obtained for the case of mutual independence of normally distributed observations, when the alarm signal is generated when $\tau_n \geq \tau_{crit}$. The results presented in Table 2 confirm our claim that the construction of a test having statistical properties similar to the properties of a Shewhart control chart is hardly possible. Due to a discrete character of τ_n ARLs in case of independence are either very large, and this suggests poor discrimination power of the test, or rather low, resulting in a high rate of false alarms.

Bearing in mind the requirement of simplicity we can now propose a Shewhart type control chart based on τ_n with control limits of the following form:

$$LCL = \max\left(E(\tau_n) - k\sigma(\tau_n), -1\right), \tag{5}$$

$$UCL = \min\left(E(\tau_n) + k\sigma(\tau_n), 1\right), \tag{6}$$

where LCL and UCL are the lower and upper limit, respectively. In the remaining part of this paper we will name it *the Kendall control chart.* To calculate the limits of the Kendall control chart we use the following formulae for the expected value and the variance of τ_n given in [5]:

$$E(\tau_n) = -\frac{2}{3(n-1)}, n \geq 3, \tag{7}$$

$$V(\tau_n) = \frac{20n^3 - 74n^2 + 54n + 148}{45(n-1)^2(n-2)^2}, n \geq 4. \tag{8}$$

It is worth noting that for small values of n the probability distribution of τ_n is not symmetric. Therefore, the properties of the proposed control chart for testing independence of consecutive observations from a process may be improved by using control lines that are asymmetric around the expected value of τ_n. However, for sake of simplicity, in this paper we will not consider this possibility. The properties of the proposed control chart are investigated in the next section of the paper.

Table 2. Values of ARL of Kendall's test for independent observations

n=6		n=7		n=8		n=9	
τ_{crit}	ARL	τ_{crit}	ARL	τ_{crit}	ARL	τ_{crit}	ARL
1	422.0	1	2885.8	1	22586.0	1	> 150000.0
0.8	151.4	0.866	1013.5	0.904	7799.0	0.857	15395.0
0.6	47.5	0.733	270.6	0.809	1891.2	0.785	5618.0
		0.600	106.5	0.714	705.9	0.714	1784.7
				0.619	246.0	0.642	724.9
						0.571	291.4

3 Properties of the Kendall control chart in case of dependencies described by a multivariate normal distribution

In order to analyze basic properties of the proposed Kendall control chart let us consider the simplest case when two consecutive observations are described by a bivariate normal distribution. Let $X_i = Z_i$ and $Y_i = Z_{i+1}$, $i = 1, 2, \ldots, n-1$ denote the random variables describing two consecutive observations in a sample of size n. We want to model the stochastic dependence between them. In order to do it we assume that the join probability distribution of the random vector (X, Y) is the bivariate normal distribution with the following probability distribution

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp(-Q), \tag{9}$$

where

$$Q = \frac{1}{1-\rho^2}\left\{\frac{(x-m_X)^2}{\sigma_X^2} - 2\rho\frac{(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right\}, \tag{10}$$

and parameter ρ is the coefficient of correlation. If the variables X and Y are independent, then we have $\rho = 0$.

The conditional cumulative distribution function of Y given $X = x$ is the normal distribution with the mean $m_Y + \rho(\sigma_X/\sigma_Y)(x - m_X)$ and the variance of $\sigma_Y\sqrt{1-\rho^2}$. In a particular case when $m_X = m_Y = 0$ and $\sigma_X = \sigma_Y = 1$ it is the normal distribution with mean of $x\rho$ and variance of $\sqrt{1-\rho^2}$. Thus, the proposed model is the well known autoregression model.

The basic characteristic that describes the performance of control charts is the Average Run Length (ARL). ARL is calculated as the average number of samples (or individual observations) plotted on a control chart up to and including the point that gives rise to a decision that a special cause is present. In Table 3 we present the results of simulation (each entry of the table is calculated as the average from one million simulation runs) for different sample sizes (numbers of considered consecutive points) n, and different values of the correlation coefficient ρ .

Table 3. Values of ARL of the Kendall control chart with k=3 for observations described by a bivariate normal distribution

				n			
ρ	6	7	8	9	10	20	50
0.8	47.36	63.71	67.81	50.56	44.96	26.67	50.11
0.5	90.94	165.44	221.86	165.35	154.87	66.41	59.94
0.2	140.82	416.60	702.13	610.14	689.44	473.52	328.82
0.1	146.23	520.36	1156.77	859.95	1593.03	1026.06	1058.76
0	141.90	595.74	1623.15	1043.36	1497.27	1502.60	2597.57
-0.1	129.64	607.00	2003.14	1020.18	1597.13	1011.92	1053.87
-0.2	111.47	577.85	2286.24	796.90	1257.70	465.59	327.97
-0.5	56.18	220.44	952.73	194.62	247.45	61.30	58.82
-0.8	23.97	53.98	145.83	41.04	43.43	23.78	50.05

The results presented in Table 3 reveal that a simple Kendall control chart with a simple to remember three-sigma decision rule, and a small sample size

n, is not a good tool for finding dependencies between consecutive observations. Obviously, the common value $k = 3$ cannot be used for all values of n. Moreover, the discrimination ability of the Kendall control chart for the sample sizes n smaller than 10 seems to be insufficient, even if we decrease the value of k. For larger value of n the situation looks better, but the discrimination power of this simple Kendall chart is still insufficient.

Now, let us analyze the behavior of a simple autocorrelation chart in similar circumstances. To design this chart we assume that the expected value of the plotted statistic and its variance are equal to their asymptotic values. Hence, we set $E(\rho_n) = 0$, and for the calculation of the variance we use an approximate simple formula proposed by Moran [17]

$$V(\rho_n) = \frac{n-1}{n(n+2)}. \tag{11}$$

Now, let us define control limits of this simple autocorrelation chart as $\pm k\sigma(\rho_n)$, and set $k = 3$. The values of ARLs for different values of n and ρ are presented in Table 4.

Table 4. Values of ARL of the autocorrelation control chart with k=3 for observations described by a bivariate normal distribution

	n						
ρ	6	7	8	9	10	20	50
0.8	2713.4	88.7	39.0	30.0	26.5	24.0	50.0
0.5	1539.3	251.7	115.8	87.6	75.7	51.7	57.0
0.2	732.8	796.3	494.7	415.5	386.1	352.9	301.1
0.1	400.4	991.1	793.0	720.3	693.6	822.7	1018.9
0	400.6	1004.8	1183.0	1190.2	1143.7	1427.0	2797.7
-0.1	287.3	834.3	1454.2	1662.3	1485.7	1121.7	1093.6
-0.2	200.1	590.9	1326.9	1677.6	1253.8	499.2	319.6
-0.5	61.3	133.9	266.9	297.2	185.3	57.8	56.8
-0.8	19.5	28.3	38.0	37.6	29.3	22.8	50.0

The results given in Table 4 look rather surprisingly. We would expect that for the assumed model of dependence the behaviour of the autocorrelation chart should be much better than the behaviour of the Kendall chart which is based on a nonparametric statistic. Surprisingly though, the behaviour of a simple autocorrelation chart does not seem much better. For small values of n the coefficient of autocorrelation is obviously biased, and the simple Moran's approximation may influence the results in a negative way. The direct comparison of the both charts using the data given in Table 3 and Table 4 is, of course, impossible. In order to make this comparison relevant we have compared both charts for $n = 10$ and $n = 50$. In case of $n = 10$ for the Kendall chart we set k to 2.7, and for the autocorrelation chart we set k to 2.65. In case of $n = 50$ for the Kendall chart we set k to 2.2, and for the autocorrelation chart we set k to 2.16. For these values of the parameters the ARLs in case of independence are nearly the same in both cases. The results of the comparison (each value based on 10^5 simulations) are presented in Table 5 and Figure 1.

The results presented in Table 5 confirm our previous finding that a simply designed autocorrelation chart (without a correction for bias) for small values of n does not perform better than the Kendall chart. For large value of n, such as $n = 50$, surprisingly though, for the assumed dependence model the autocorrelation chart does not perform better, as it is expected to do. The effect of bias is still observed, and this results in better discrimination of negative dependence, and visibly worse discrimination for positive dependence. It is also worth noticing that fast detection of small correlations requires samples even much larger than $n = 50$.

Table 5. Values of ARL for equivalent Kendall and autocorrelation control chart

	n=10		n=50	
ρ	Kendall	autocorrelation	Kendall	autocorrelation
0.8	30.9	19.8	50	50
0.5	84.3	43.3	52.1	51.9
0.3	189.1	97.9	73.9	75.8
0.2	271.2	156.8	117.5	127.4
0.1	344.6	249.4	226.6	259.1
0	351.2	351.9	350.7	350.9
-0.1	286.4	385.6	226.3	189.8
-0.2	200.6	305.8	117.1	99.3
-0.3	132.2	193.3	73.3	66
-0.5	55.7	67.0	51.8	50.9
-0.8	19.4	19.4	50	50

4 Properties of the Kendall control chart in case of other models of dependence

Let us consider now the sensitivity of the proposed Kendall chart and the simple autocorrelation chart to the change of some basic assumptions. First, we check how changes the performance of these charts when we assume that the consecutive observations in a sample are not independent. To describe diffrent kinds of dependencies between them we need model different from autoregression.

For the bivariate normal distribution the correlation coefficient completely defines the dependence structure between random variables. However, it is

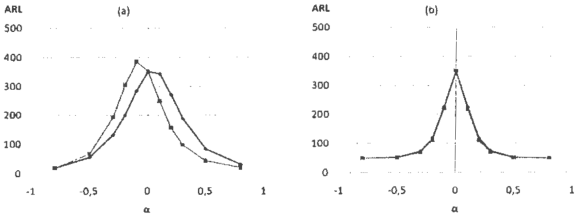

Fig. 1. ARL comparison for Kendall control chart (black line) and autocorrelation control chart (grey line) (a) $n = 10$, (b) $n = 50$

worth to notice that the random vector (X, Y) can be described by any two-dimensional probability distribution function. In such case, the information given by a correlation coefficient may be not sufficient to define the dependence structure between random variables. Therefore, to fully capture this structure one may consider another type of dependence described by a so called copula. The copula contains all of the information on the nature of the dependence between random variables. The joint cumulative distribution function $F_{12}(x, y)$ of any pair (X, Y) of continuous random variables may be written in the form $F_{12}(x, y) = C(F_1(x), F_2(y))$, $x, y \in R$, where $F_1(x)$ and $F_2(y)$ are the marginal density functions of X and Y, respectively, and $C : [0,1]^2 \to [0,1]$ is a copula. In this paper, to model another type of dependence between random variables we use, as examples, the Farlie-Gumbel-Morgenstern (FGM) copula (C_{FGM}), Plackett copula (C_P) and Frank copula (C_F). Using the probability integral transformations $u_1 = F_1(x)$, $u_2 = F_2(y)$, where u_1, u_2 have a uniform distribution on the interval $[0, 1]$, we can write the particular copulas as

$$C_{FGM}(u_1, u_2) = u_1 u_2 \{1 + \alpha(1 - u_1)(1 - u_2)\}, \ \alpha \in [-1, 1], \tag{12}$$

$$C_P(u_1, u_2) = \begin{cases} u_1 u_2, & \alpha = 1, \\ \frac{[1+(\alpha-1)(u_1+u_2)] - \sqrt{[1+(\alpha-1)(u_1+u_2)]^2 - 4u_1 u_2 \alpha(\alpha-1)}}{2(\alpha-1)}, & \alpha \in R_+ \setminus \{1\}, \end{cases} \tag{13}$$

$$C_F(u_1, u_2) = -\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1}\right), \ \alpha \in R \setminus 0. \tag{14}$$

The parameter α in above formulas for C_{FGM}, C_P and C_F describes the power and the direction of association between X and Y. The variables X and Y are independent if and only if $\alpha = 0, \alpha = 1, \alpha \cong 0$, respectively for C_{FGM}, C_P, C_F.

When $F_1(x)$ and $F_2(y)$ are the univariate cumulative distribution functions of the normal distribution, the marginal distributions of the FGM copula are normal, but the structure of dependence is different than in the case of the bivariate normal distribution. For example, when $\alpha \in [-1, 1]$, there exists a limit on the coefficient of correlation, namely $\rho \leq \alpha/\pi$ (see [20]). The similar fact is also true for Plackett and Frank copulas (see [4], [6] for more details).

The Kendall control chart is based on a nonparametric statistic. Thus, its performance should not depend not only upon the type of a marginal distribution of observations, but upon the type of dependence as well. Using Monte Carlo simulations we investiagte the performance of the Kendall and autocorrelation charts. We assume that the dependence between pairs of observations Z_i, Z_{i+1}, $i = 1, \ldots, n-1$ in a sample of size n is described by a copula and we generate the random numbers sample. If the Kendall τ statistic calculated for this sample does not fall outside of the limits established for the Kendall chart, we move to the next process observation, i.e. we genarate random number Z_{n+1} and we calculate the Kendall τ for the sample $Z_1' = Z_2, \ldots, Z_n' = Z_{n+1}$. We repeat this step as long as the Kendall τ falls outside the limits. The number of process observations Z_i plotted on the control chart up to and including the last observation in a sample for which the Kendall τ is outside the limits defines ARL.

In Tables 6-7 and Figures 2-4 we present the average ARL values, each based on 10^5 simulations, obtained in case of normally distributed observations, where the dependence between consecutive observations is described by FGM, Plackett and Frank copula.

In Table 6 there are given the ARL values for the Kendall chart in case of FGM copula. On the basis of these results we can observe that the Kendall chart for FGM copula with normal marginal distributions behaves similarly as in case of bivariate normal distributions (some slight differences may come from two different applied generators of random numbers). This fact arises from nonparametric character of Kendall's tau statistic. It confirms fact that the performance of the Kendall control chart based on a nonparametric statistic should not be dependent on the type of dependence. Obviously, results for autocorrelation control chart for simple autoregressive model (see Table 4) and the same chart for FGM copula (see Table 7) do not confirm this fact and they are differ from each other in a meaningful way.

Now, we compare both control charts applied for FGM copula. In order to do it, in case of $n = 10$ for Kendall chart we set k to 2.71 and for autocorrelation chart we set k to 2.73. In case of $n = 50$ we set k to 2.17 for both charts. The results of those experiments are presented in Figure 4. For $n = 10$ in case of positive dependence we obtain lower ARL values for autocorrelation chart than for Kendall chart, while in case of negative dependence the situation is inverse. It is difficult to specify which of these two charts is better. For $n = 50$ both control charts are symmetric and corresponding ARL values are nearly the same.

We have repeated the analogical research for Plackett and Frank copulas and we have obtained similar results. It is worth to notice that in case of these two copulas for extreme or high values of parameter α, ARL are the minimum and equal to the sample size. In case of Plackett copula for $n = 10$ and $\alpha \cong 0$ or $\alpha > 30$ ARL attains the minimal value which is close to 10, while in case

Table 6. ARL for the Kendall control chart with $k = 3$ for observations described by FGM copula with normal marginal distributions

		n						
α	ρ	6	7	8	9	10	20	50
1	0.318	128.83	297.60	470.94	371.06	389.99	191.19	112.55
0.942	0.3	129.50	315.19	510.86	395.31	420.79	216.42	127.04
0.628	0.2	136.43	402.49	729.07	578.95	651.87	443.14	288.60
0.314	0.1	145.80	490.94	1049.37	808.29	1016.84	969.05	955.37
0	0	144.74	578.02	1506.35	1009.02	1475.79	1481.33	2593.94
-0.314	-0.1	132.16	593.57	2101.18	994.87	1637.61	950.73	962.21
-0.628	-0.2	115.96	553.11	2469.10	809.24	1327.59	444.13	287.80
-0.942	-0.3	91.57	459.41	2325.11	566.49	875.34	213.47	124.05
-1	-0.318	92.87	438.73	2238.41	538.43	816.31	188.24	110.40

Table 7. ARL for the autocorrelation control chart with $k = 3$ for observations described by FGM copula with normal marginal distributions

		n						
α	ρ	6	7	8	9	10	20	50
1	0.318	1048.79	561.54	287.43	225.15	202.95	153.47	115.72
0.942	0.3	1021.87	592.45	311.75	246.39	222.83	170.98	129.73
0.628	0.2	813.13	794.85	472.06	397.84	373.21	348.94	303.60
0.314	0.1	566.77	961.89	750.39	670.50	657.14	793.17	999.54
0	0	392.76	998.18	1137.59	1155.81	1143.13	1459.37	2805.56
-0.314	-0.1	261.16	771.92	1443.81	1706.57	1512.51	1083.89	1079.53
-0.628	-0.2	177.08	531.78	1263.12	1661.35	1233.48	493.23	321.80
-0.942	-0.3	124.47	344.45	842.86	1102.55	736.39	226.63	133.82
-1	-0.318	116.06	317.42	772.82	1003.40	661.75	198.12	117.73

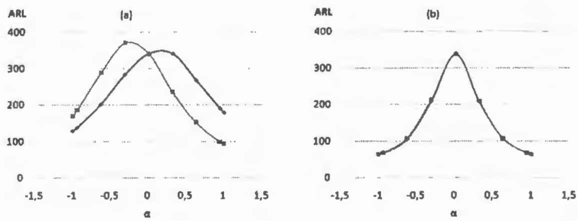

Fig. 2. ARL comparison for the Kendall chart (black line) and autocorrelation chart (grey line) for FGM copula with normal marginal distributions (a) $n = 10$, (b) $n = 50$

of Frank copula ARL is minimal and equal to 10 for $|\alpha| > 13$. For sample size $n = 50$ in case of Plackett copula ARL=50 for $\alpha \cong 0$ or $\alpha > 8$. In case of Frank copula the ARL=50 for $\alpha > 4$. Analysing results presented in Figures 2-4 we get the conclusion that the autocorrelation control chart does not perform better than the Kendall chart based on a nonparametric statistic.

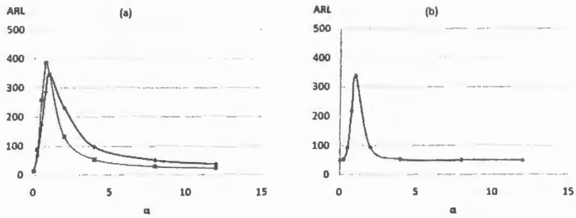

Fig. 3. ARL comparison for the Kendall chart (black line) and autocorrelation chart (grey line) for Plackett copula with normal marginal distributions (a) $n = 10$, (b) $n = 50$

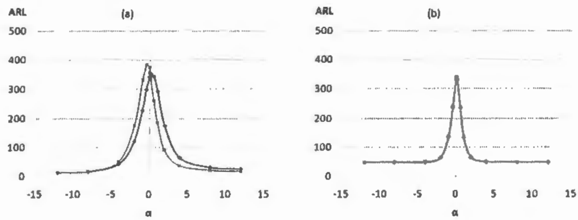

Fig. 4. ARL comparison for Kendall control chart (black line) and autocorrelation control chart (grey line) for Frank copula with normal marginal distributions (a) $n = 10$, (b) $n = 50$

It seems to be interesting to make the comparison of the Kendall and autocorrelations control charts in case of non-normal distributions. To verify the influence of character of distribution on the performance of the considered charts we make the analogical research under assumption that marginal distribution function is not normal. We consider two cases, first - we do not know the distribution of our data is not normal, and second - we know the character of non-normal distribution function of our data. In the last case we modify the width of the autocorrelation control chart and set a proper value of parameter k to get nearly the same ARLs for this chart and the Kendall chart in case of independence. The results obtained for FGM, Plackett and Frank copulas with exponential and uniform marginal distributions are presented in Figures 5-10.

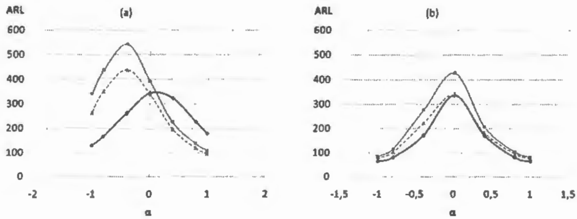

Fig. 5. ARL comparison for Kendall control chart (black line), autocorrelation control chart (grey line) and modified autocorrelation chart (grey dotted line) for FGM copula with exponential distributions (a) $n = 10$, (b) $n = 50$

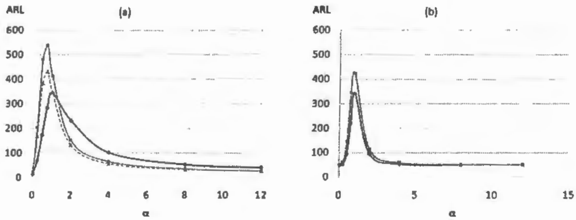

Fig. 6. ARL comparison for Kendall control chart (black line), autocorrelation control chart (grey line) and modified autocorrelation chart (grey dotted line) for Plackett copula with exponential marginal distributions (a) $n = 10$, (b) $n = 50$

Let's assume that we do not know the type of marginal distribution and we use the autocorrelation chart with determined for normal distribution value of $k = 2.73$. Then, if the marginal distributions are really non-normal, we see that ARL values are different from ARL values obtained in case of normal marginal distri-butions. So, in such case we have to set a suitable value of

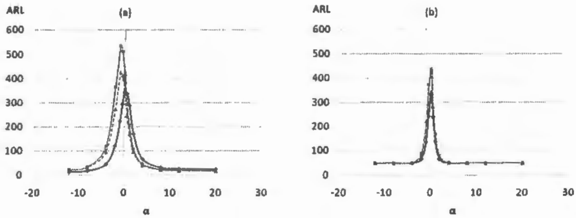

Fig. 7. ARL comparison for Kendall control chart (black line), autocorrelation control chart (grey line) and modified autocorrelation chart (grey dotted line) for Frank copula with exponential marginal distributions (a) $n = 10$, (b) $n = 50$

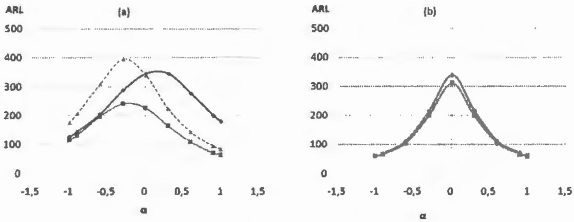

Fig. 8. ARL comparison for Kendall control chart (black line), autocorrelation control chart (grey line) and modified autocorrelation chart (grey dotted line) for FGM copula with uniform marginal distributions (a) $n = 10$, (b) $n = 50$

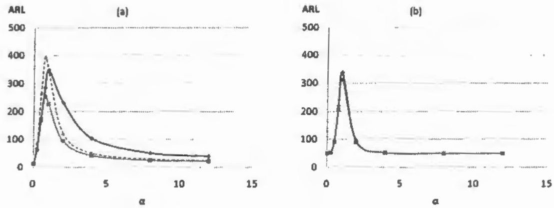

Fig. 9. ARL comparison for Kendall control chart (black line), autocorrelation control chart (grey line) and modified autocorrelation chart (grey dotted line) for Plackett copula with uniform marginal distributions (a) $n = 10$, (b) $n = 50$

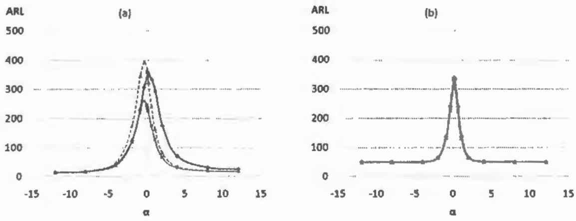

Fig. 10. ARL comparison for Kendall control chart (black line), autocorrelation control chart (grey line) and modified autocorrelation chart (grey dotted line) for Frank copula with uniform marginal distributions (a) $n = 10$, (b) $n = 50$

k, but for that purpose we need the information about our distribution. For $n = 10$, in case of positive dependence the ARL values are lower or nearly the same and in case of negative dependence they are greater than ARL values for Kendall chart. Whereas for $n = 50$ the Kendall control chart is not worse than autocorrelation control chart. In case of uni-form marginals it works almost like the autocorrelation chart, but in case of expo-nential marginal distributions it works better.

So, now let's assume that we know the type of our marginal distributions. Then we can set a proper value of k. In our research we set $k = 2.68$ and $k = 2.85$, respectively, for exponential and uniform distributions. Let's notice that even if we profit from this knowledge the autocorrelation chart does not perform better then the Kendall chart.

The results given in Figures 5-10 show, how sensitive is the autocorrelation chart to the assumption of the underlying distribution. The compared distributions have been chosen deliberately so different in order to magnify the differences. For small sample sizes n, in both cases of a heavy-tailed distribution (uniform) and a skewed distribution (exponential), the rate of false alarms is unacceptable high. Therefore, the autocorrelation chart should be specially tailored in case of different probability distribution of the plotted observations. This conclusion is hardly unexpected, but the observed differences in the behavior of the autocorrelation control chart in case of different distributions are very significant from a practical point of view. So, we should apply the autocorrelation chart very carefully if we do not know that our distribution is really the normal distribution. On the other hand, if we know the type of our distribution and can set a proper value of k, we do not obtain the control chart with better performance than Kendall control chart.

5 Conclusions

Mutual dependencies (correlations) between consecutive observations of processes may influence properties of SPC procedures in a dramatic way. This phenomenon has attracted the attention of many researchers for the last over twenty years. Many new or modified SPC tools have been pro-posed for dealing with this problem. However, their usage requires addi-tional skills, specialized software and is usually much more complicated than in the case of classical tools used for mutually independent observa-tions. Thus, from a practical point of view, it is very important to identify situations when an additional treatment of data is really necessary. In other words, there is a need to have a simple SPC tool, like a Shewhart control chart, which could be used for the detection of dependencies (autocorrela-tions) between observations of processes.

In this paper we have proposed such a tool based on a nonparametric Kendall's tau statistic. This tool is similar to a Shewhart control chart with plotted values of the Kendall's tau. We compared this new tool to a known autocorrelation chart. The results of the comparison show that in the case of dependencies described by autoregressive processes with a normal error component the new tool performs nearly as well as the autocorrelation chart. However, when the assumption of multivariate normality of consecu-tive observations is not fulfilled, the newly proposed Kendall control chart performs much better due to its distribution-free character.

The Kendall chart, in its simplest "three-sigma" form, does not perform well for small and moderate sample sizes. It has unnecessarily high values of ARL in case of independence, and hence, a very low rate of false alarms. Unwanted consequences of this feature are the high values of ARL in case of the existence of weak dependencies between observations. This situation can be improved by changing the limits of a control chart, but such im-provements may not be sufficient for the effective detection of weak de-pendencies. For such cases large sample sizes are required, and this un-pleasant from a practical point of view situation does not depend upon a statistical tool used for the detection of autocorrelation.

As it was mentioned above, it seems to be impossible to propose a very simple design (like e.g. using a "three-sigma" rule) of the Kendall control chart. Additional research is required in order to work-out guidance how to design an effective chart for different practical situations. Special attention should be paid to situations, when the type of dependence differs from a simple autoregressive process. There is also a need to investigate properties of the Kendall chart in case of shifts in a process mean value, and for other types of process deterioration. Preliminary investigations, not reported in this paper, suggest, however, that the Kendall control chart is not a good tool for the detection of such deteriorations.

References

1. L.C. Alwan and H.V. Roberts. Time-series modeling for statistical process control. *Journal of Business & Economic Statistics*, 6:87–95, 1988.
2. L.C. Alwan and H.V. Roberts. The problem of misplaced control limits. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 44:269–306, 1995.
3. G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis, Forecasting and Control.* Prentice-Hall, Englewood Cliffs, 3 edition, 1994.
4. D. Conway. Plackett family of distributions. *Encyclopedia of Statistical Sciences*, pages 6164–6168, 2006.
5. T.S. Ferguson, C. Genest, and M. Hallin. Kendall's tau for serial dependence. *The Canadian Journal of Statistics*, 28:587–604, 2000.
6. C. Genest. Frank's family of bivariate distributions. *Biometrika*, 74:549–555, 1987.
7. M. Hallin and G. Mélard. Rank-based tests for randomness against first-order serial dependence. *Journal of the American Statistical Association*, 83:1117–1128, 1988.
8. W. Jiang, K.L. Tsui, and W.H. Woodall. The new SPC monitoring method: The ARMA chart. *Technomerics*, 42:399–410, 2000.
9. R.A. Johnson and M. Bagshaw. The effect of serial correlation on the peformance of CUSUM tests. *Technomerics*, 16:103–112, 1974.
10. S. Knoth, W. Schmid, and A. Schone. Simultaneous Shewhart-type charts for the mean and the variance of a time series. *Frontiers in Statistical Quality Control VI*, pages 61–79, 2001.
11. C.W. Lu and M.R. Reynolds Jr. Control charts for monitoring the mean and variance of autocorrelated processes. *Journal of Quality Technology*, 31:259–274, 1999.
12. C.W. Lu and M.R. Reynolds Jr. EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology*, 31:166–188, 1999.
13. C.W. Lu and M.R. Reynolds Jr. CUSUM charts for monitoring an autocorrelated process. *Journal of Quality Technology*, 33:316–334, 2001.
14. J.M. Lucas. Discussion to the paper „Run-length distributions of special-cause control charts for correlated processes" by Wardell et al. *Technometrics*, 36:17–19, 2001.
15. H.D. Maragah and W.H. Woodall. The effect of autocorrelation on the retrospective X-chart. *Journal of Statistical Computation and Simulation*, 40:29–42, 1992.
16. D.C. Montgomery and C.M. Mastrangelo. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, 23:179–193, 1991.
17. P.A.P. Moran. Some theorems on time series I. *Biometrika*, 34:281–291, 1947.
18. W. Schmid. On the run length of a Shewhart control chart for correlated data. *Statistical Papers*, 36:111–130, 1995.
19. W. Schmid. On EWMA charts for time series. pages 115–137. Physica Verlag, Heidelberg, 1997.
20. W. R. Schucany, W. C. Parr, and J. E. Boyer. Correlation structure in Farlie-Gumbel-Morgenstern distributions. *Biometrika*, 65:650–653, 1978.
21. A. Snoussi, M. El Ghourabi, and M. Limam. On SPC for short run autocorrelated data. *Communications in Statistics Simulation and Computation*, 34:219–234, 2005.

22. D.H. Timmer, J. Pignatello Jr, and M. Longnecker. The development and evaluation of CUSUM-based control charts for an AR(1) process. *IIE Transactions*, 30:525–534, 1998.
23. L.N. VanBrackle III and M.R. Reynolds Jr. EWMA and CUSUM control charts in presence of correlations. *Communications in Statistics Simulation and Computation*, 26:979–1008, 1997.
24. A.V. Vasilopoulos and A.P. Stamboulis. Modification of control limits in the presence of correlation. *Journal of Quality Technology*, 10:20–30, 1978.
25. D.G. Wardell, H. Moskowitz, and R.D. Plante. Run-length distributions of special-cause control charts for correlated processes. *Technometrics*, 36:3–17, 1994.
26. W.H. Woodall and F.W. Faltin. Autocorrelated data and SPC. *ASQC Statistics Division Newsletter*, 13:18–21, 1993.
27. C.M. Wright, D.E. Booth, and M.Y. Hu. Joint estimation: SPC method for short-run autocorrelated data. *Journal of Quality Technology*, 33:365–378, 2001.
28. S.A. Yourstone and D.C. Montgomery. A time-series approach to discrete real-time process quality control. *Quality and Reliability Engineering International*, 5:309–317, 1989.
29. S.A. Yourstone and D.C. Montgomery. Detection of process upsets sample autocorrelation control chart and group autocorrelation control chart applications. *Quality and Reliability Engineering International*, 7:133–140, 1991.
30. N.F. Zhang. Statistical control chart for stationary process data. *Technometrics*, 40:24–38, 1998.