Raport Badawczy
Research Report

RB/15/2010

Performance of the Shewhart control chart in the presence of dependent data

O. Hryniewicz, J. Karpiński,
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 2010

Performance of the Shewhart control chart in the presence of dependent data

Olgierd Hryniewicz, Janusz Karpinski, Anna Szediw

Abstract Shewhart control charts were originally designed under the assumption of independence of consecutive observations. In the presence of dependence the authors usually assume dependencies in the form of autocorrelated and normally distributed data. However, there exist many other types of dependencies which are described by other mathematical models. The question arises then, how classical control charts are robust to different types of dependencies. This problem has been sufficiently well discussed for the case of autocorrelated and normal data.In the paper we use the concept of copulas to model dependencies of other types. We use Monte Carlo simulation experiments to investigate the impact of type and strength of dependence in data on the value of the *ARL* of Shewhart control charts.

Key words: Shewhart control charts, correlated data, copulas; ARL

1 Introduction

Statistical process control (SPC) is a collection of statistical methods used by thousands of practitioners who are striving to achieve 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 $\bar{X}$ control chart, known for more than eighty years, is the most popular SPC method used to detect whether observed process is under control. Its classical and internationally standardized version is 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 obviously correlated, e.g. in case of certain continuous production processes. The presence of correlations between consecutive measurements should be

Olgierd Hryniewicz, Janusz Karpinski, Anna Szediw
Systems Research Institute, Newelska 6, 01-447 Warsaw, Poland

taken into account during the design of control charts. This need was already noticed in the 1970s, see e.g. the papers by Johnson and Bagshaw [10] and by Vasilopoulos and Stamboulis [25], but the problem was widely discussed in papers published in the late 1980s and in the 1990s.

One of the visible effects of autocorrelation in observed process data is the significant difference between statistical properties of control charts designed for independent and dependent data. There exist several approaches for dealing with this problem. First approach, historically the oldest one, consists in dealing with original data and adjusting control limits of classical control charts. This approach was used, for example, in papers by Reynolds Jr. and co-authors [13],[12],[14],[24], Schmid [21],[22], Vasilopoulos and Stamboulis,[25], and Zhang [29]. Other approaches are based on the concept of residuals (see the papers by Alwan and Roberts [2] or by Montgomery and Mastrangelo [16]) or on monitoring statistics related to autocorrelations (see the papers by Yourstone and Montgomery [28] or by Jiang *et al.* [9]). There also exist more sophisticated methods for dealing with SPC autocorrelated data. An overview of SPC methods used for autocorrelated data can be found in papers by Wardell *et al.* [26], Lu and Reynolds [12], and Knoth *et al.* [11].

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 a 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],[14] have shown that precise estimation of such parameters requires at least hundreds of observations.

All these problems, noticed by many authors, make the SPC with dependent data very difficult, especially for not well-trained in statistics practitioners who need efficient tools to discriminate between complicated problems with dependent data and relatively simple problems when observed data are independent. This problem was considered in the paper by Hryniewicz and Szediw [8] who proposed a relatively simple and efficient nonparametric tool, named by them the Kendall control chart, for testing hypotheses about independence of SPC data. While discussing the properties of this tool they noticed that the type of existing dependence plays a crucial role. In this paper we continue the work along that line by analyzing the properties of Shewhart control charts when data are generated by different variants of a simple autoregression model. The mathematical model that describes serial dependence between consecutive observations of a process in terms of copulas is described in the second section of the paper. In the third section we present the results of Monte Carlo simulation experiments which show very strong dependence of statistical properties of control charts upon the type and the strngth of dependence. Conclusions derived from these results are presented in the last section of the paper.

2 Mathematical models of dependence between consecutive observations on a control chart

Mathematical models used for the description of dependent random variables are well known for many years. In the simplest two-dimensional case we are interested in the description of dependence between two random variables X and Y having marginal distributions described by cumulative probability functions $F(x)$ and $G(y)$, respectively. In the context of the considered in this paper time-dependent observations we can, in the simplest case, set $X = X_t$ and $Y = X_{t+1}$, where $X_t, t = 1, 2, \ldots$ is the time series representing consecutive observations of the process under consideration. In his fundamental work Sklar [23] showed that for a two-dimensional probability distribution function $H(X,Y)$ with marginal distribution functions $F(X)$ and $G(Y)$ there exists a copula C such that $H(x,y) = C(F(x), G(y))$. This result has been later extended to the case of multivariate probability distributions. For more information about copulas the reader should refer e.g to the book by Nelsen [17].

All well known multivariate probability distributions, the multivariate normal distribution included, can be generated by parametric families C_α of copulas, where real- or vector-valued parameter α describes the strength of dependence between the components of the random vector. Thus, copulas have found many interesting practical applications. The number of papers devoted to the theory and applications of copulas is still growing rapidly, thanks to the increasing interest coming from e.g. the analysis of financial risks and the survival analysis. For more recent results the reader should consult already mentioned book by Nelsen [17].

In this paper we focus our attention on three types of copulas. First is the normal copula, which in the two-dimensional case is defined as follows:

$$C(u_1, u_2; \rho) = \Phi_N(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho) \tag{1}$$

where $\Phi_N(u_1, u_2)$ is the cumulative probability distribution function of the bivariate normal distribution, $\Phi^{-1}(u)$ is the inverse of the cumulative probability function of the univariate normal distribution (the quantile function). Parameter ρ in case of marginals described by the normal distribution is equal to the well known coefficient of linear correlation introduced by Pearson. It is worth noticing that the values of the linear correlation coefficient depend upon the type of marginals. Therefore, for the same value of the parameter ρ of the normal copula, the values of the Pearson's correlation may be different for different distributions of X and Y.

Second copula considered in this paper is the Farlie-Gumbel-Morgenstern (FGM) copula who is frequently used for modelling weak dependencies. This copula is defined by the following formula:

$$C(u_1, u_2; \theta) = u_1 u_2 + \theta u_1 u_2 (1 - u_1)(1 - u_2), |\theta| \leq 1 \tag{2}$$

The remaining three copulas considered in this paper belong to a general class of symmetric copulas, named the Archimedean copulas. They are generated using a class Φ of functions $\phi : [0,1] \to [0,\infty]$, named generators, that have two continuous derivatives on $(0,1)$ and fulfill the following conditions: $\phi(1)=1$, $\phi'(t)<0$, and $\phi''(t)>0$ for all $0<t<1$ (these conditions guarantee that ϕ has an inverse ϕ^{-1} that also has two derivatives). Every member of this class generates a multivariate distribution function. In this paper we consider three two-dimensional Archimedean copulas defined by the following formulae (copulas and their respective generators):

- Clayton's

$$C(u,v)=\max\left(\left[u^{-\alpha}+v^{-\alpha}-1\right]^{-1/\alpha},0\right),\alpha\in[-1,\infty)\setminus 0 \tag{3}$$

$$\phi(t)=(t^{-\alpha}-1)/\alpha,\alpha\in[-1,\infty)\setminus 0 \tag{4}$$

- Frank's

$$C(u,v)=-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1}\right),\alpha\in(-\infty,\infty)\setminus 0 \tag{5}$$

$$\phi(t)=\ln\left(\frac{1-e^{-\alpha}}{1-e^{-\alpha t}}\right),\alpha\in(-\infty,\infty)\setminus 0 \tag{6}$$

- Gumbel's

$$C(u,v)=\exp\left(-\left[(-\ln u)^{1+\alpha}+(-\ln v)^{1+\alpha}\right]^{\frac{1}{1+\alpha}}\right),\alpha\in(0,\infty) \tag{7}$$

$$\phi(t)=(-\ln(t))^{\alpha+1},\alpha\in(0,\infty) \tag{8}$$

In case of independence the dependence parameter α_{ind} adopts the value of 0 (in Clayton's and Frank's copulas as an appropriate limit). The copulas mentioned above are sometimes presented using different parametrization, and in such cases independence is equivalent to other values of α.

As it has been already mentioned above, a well known coeficient of linear correlation cannot be used for measuring the strength of dependence between random variables whose dependence is described by a given copula. Nonparametric measures of dependence, such as Spearman's ρ or Kendall's τ can be used for this purpose. For the copulas considered in this paper the values of Kendall's τ are easier to calculate, and for this reason we use this measure of dependence in further analyses.

Genest and MacKay [6] considered the population version of the Kendall's coefficient od dependence (association) τ. This characteristic can be used for the description of the strength of dependence in copulas, and its importance in characterizations of copulas has been shown recently in papers by Nelsen *et. al.* [18]. Let $K(t)$ be the cumulative probability function of the random variable $T=C(U_1,U_2)$, where U_1 and U_2 are random variables uniformly distributed on $[0,1]$. The following relation links a copula with Kendall's τ:

$$\tau=3-4\int_0^1 K(t)dt \tag{9}$$

Estimation of $K(t)$ for the case of two-dimensional copulas, and thus the estimation of τ, was considered by Genest and Rivest [7].

Closed formulae for Kendall's τ are available only for some copulas. In the case of the normal copula we have the following expression

$$\tau_{Norm} = arcsin(\rho)/(\pi/2). \tag{10}$$

For the FGM copula we can compute Kendall's τ from a very simple formula

$$\tau_{FGM} = 2\theta/9. \tag{11}$$

For the family of Archimedean copulas there exists the following general formula that links Kendall's τ with the generator function ϕ:

$$\tau_{Arch} = 1 + 4\int_0^1 \frac{\phi(v)}{\phi'(v)}dv. \tag{12}$$

For specific cases of the considered in this paper Archimedean copulas we have:

- Clayton's copula

$$\tau = \frac{\alpha}{\alpha+2}, \tag{13}$$

- Frank's copula

$$\tau = 1 + 4\left(\frac{1}{\alpha}\int_0^{\alpha}\frac{t}{e^t-1}dt - 1\right)/\alpha, \tag{14}$$

- Gumbel's copula

$$\tau = \frac{\alpha}{\alpha+1}. \tag{15}$$

Each copula can be looked upon as a multivariate probability distribution whose all marginal distributions are uniform. However, by using an inverse probability distribution function (a quantile function) we can transform each uniformly distributed random variable to a variable with any continuous probability distribution. In this paper we will consider the case when such transformation will lead to marginals described by the standard normal distribution $N(0,1)$. This assumption definitely restricts generality of inferred conclusions, but - on the other hand - allows to compare our new results with those presented by other authors who usually made this assumption.

3 Basic properties of the Shewhart control chart in case of dependencies of different types

The most frequently used statistical characteristic of a control chart is its Average Run Length *ARL*. This characteristic describes the expected number of observations (points plotted on a chart) until the occurence of an alarm (e.g. when the first point beyond 3-sigma control limits has been observed). When consecutive observations are independent, and their probability distribution is known, the random variable which describes the waiting time till the moment of the first observation beyond the control limits is distributed according to the geometric distribution, and the value of the *ARL* can be calculated analytically. However, when observations are dependent (serially correlated) and/or their probability distributions are only partially known (e.g. the class of the distribution is known, but its parameters are estimated) this characteristic usually cannot be calculated from a closed formula. Therefore, we need to use statistical Monte Carlo simulation in order to evaluate the value of the *ARL*.

In our simulation experiments we have generated consecutive observations using conditional probability distributions derived from two-dimensional copulas. In order to arrive at comparable results we have generated serially correlated processes described by a fixed in advance value of Kendall's τ. By having the same normal marginal distributions, and the same values of the measure of the strength of dependence we can detect a possible influence of the type of dependence related to the type of the underlying copula. In the following two subsections we will present the results of experiments for two cases:

- Parameters of the normal distribution (design parameters) are known,
- Parameters of the normal distribution (design parameters) are estimated from an initial sample.

In both cases we consider only one type of the process deterioration: the shift of the process level by $k\sigma$. When $k = 0$ (i.e. when there is no shift) the value of the *ARL* represents the average time to a false alarm. When $k = 1$ we have the case of a small deterioration. Significant deterioration of the process is in our experiment modelled by setting $k = 3$.

3.1 Known design parameters

The results of the simulation experiment for known design parameters are presented in Tables 1–3 for different values of the shift of the process level (mean value). Each number in these tables has been obtained after averaging the results of 200 000 simulation runs. The maximal length of each simulation run varied from 10000 to 100 000 observations (for strongly dependent observations).

In Table 1 we present the average times to a false alarm. In all considered cases the expected time to a false alarm in presence of dependent data is always larger than in the case of independence, and this difference increases with increasing strength of dependence. However, the way how the *ARL* depends on the value of τ strongly depends on the type of the copula that describes the data. It is interesting to see that for all considered copulas, with a noticeable exception of Gumbel's copula, the values of the *ARL* change insignificantly for weakly dependent data. However, for moderate and strong dependencies these values are changing in a completely different way depending on the type of a copula. For the normal copula (i.e. for an ordinary Gaussian autoregression $AR(1)$ process) the value of the ARL increases for the increasing absolute value of the strength of dependence measured by Kendall's τ. The dependence of the ARL on the value of τ is symmetrical and these values become very large for large values of τ. In case of the FGM copula, which is used for modelling weak dependencies, the influence of the value of τ on the ARL is practically non-existing. The similar situation, but extended to larger values of τ, has been observed for Frank's copula.

Table 1 *ARLs*–Shewhart control chart(design parameters known), TEST 1: "3-σ" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	1391,0	x	759,3	385,4	1301,8
0,5	466,8	x	622,4	373,1	633,57
0,3	389,0	x	496,0	373,4	528,6
0,1	371,1	369,6	384,4	370,5	456,4
0,05	370,5	369,3	374,7	372,1	443,3
0,01	370,6	369,7	371,6	372,2	430,7
0	**370,5**	**370,5**	**370,5**	**370,5**	**370,5**
-0,01	372,31	369,28	370,19	372,97	x
-0,05	371,3	370,8	370,0	371,7	x
-0,1	371,3	368,9	370,6	371,2	x
-0,3	390,0	x	384,3	374,1	x
-0,5	468,4	x	433,9	375,0	x
-0,8	1379,4	x	898,6	384,0	x

In the case of Clayton's copula the dependence of the ARL upon the value of τ is not symmetric. In case of positive dependence ($\tau > 0$), and small and moderate strength of dependence, the ARL in this case is larger than in the case of the normal copula. However, in case of very strong positive dependence this value of the ARL is significantly smaller than in the normal case. In case of negative dependence ($\tau < 0$) the ARL for Clayton's copula is always smaller than the ARL in the normal case. The case of Gumbel's copula requires special comments. This copula describes only positive dependence, and even for very weak dependencies the corresponding values of the ARL are significantly greater than in the case of independence. Only in case of very strong dependence the behaviour of the Shewhart control chart seems to be similar to that described by the normal copula. It means that the Shewhart control chart is very sensitive to this type of dependence, even if this dependence is very weak, and thus difficult to be confirmed.

In case of small shifts (equal to 1σ) of the process level the dependence of the ARL upon τ looks different. The values of the ARL in presence of dependent data are nearly always greater than in the case of independence. It means that the dependence in data has negative impact on discrimination abilities of the Shewhart control chart, and this unpleasant feature does not depend upon the type of dependence. In case of strongly dependent data the values of the ARL may be so large (especially for normal and Gumbel's copulas) that the chart becomes practically insensitive to relatively

Table 2 *ARLs*–Shewhart control chart (design parameters known), TEST 1: "3-σ" rule, shift of 1σ

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	273,21	x	90,28	86,74	653,04
0,5	71,95	x	48,55	54,18	126,65
0,3	52,25	x	45,18	47,7	73,07
0,1	45,11	43,97	43,74	45,42	50,39
0	**43,78**	**43,78**	**43,78**	**43,78**	**43,78**
-0,1	43,62	43,73	43,74	44,75	x
-0,3	44,36	x	44,63	44,55	x
-0,5	50,96	x	50,15	45,53	x
-0,8	135,18	x	101,04	52,44	x

small deterioration of the process. However, in the case of Frank's copula the value of the *ARL* remains reasonable even for strongly dependent data (especially in case of negative dependence).

Table 3 *ARLs*–Shewhart control chart (design parameters known), TEST 1: "3-σ" rule, shift of 3σ

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	10,58	x	24,94	9,75	9,1
0,5	3,29	x	4,05	4,51	3,16
0,3	2,49	x	2,59	3,55	2,44
0,1	2,1	1,99	2,12	3,13	2,11
0	**1,99**	**1,99**	**1,99**	**1,99**	**1,99**
-0,1	1,92	2,0	1,92	2,91	x
-0,3	1,79	x	1,83	2,76	x
-0,5	1,69	x	1,77	2,65	x
-0,8	1,57	x	1,71	2,57	x

When the shift of the process is large (e.g. equal to 3σ) the situation is different. First of all, in case of negative dependence described by the normal and Clayton's copulas the chart reacts faster than in the case of independence. Positive dependence in all considered cases has negative influence on the ability of the chart to detect shifts. The worse situation is in the case of Frank's copula, and this is somewhat unexpected because for small shifts this copula seems to be the most favourable. A similar situation is with Clayton's copula which usually behaves quite well except for the case of large shifts and strong positive dependence.

3.2 Estimated design parameters

Let us consider the case when parameters of the probability distribution (mean value and standard deviation) that are used for the design of a control chart are estimated from a process (its Phase I, as the sampling period is sometimes called) with possibly dependent consecutive observations. This assumption leads to significant consequences. First of all, random character of control lines which are estimated from a sample adds some variability resulting in wider (on average) in-control area on a control chart. This problem has been considered by many authors, and some conclusions from that research may be found in the paper by Woodall and Montgomery [27] or in the paper by Albers and Kallenberg [1]. Second, the autocorrelation between sample observations influences the properties of estimators, as it was noticed already in the paper by Vasilopoulos and Stamboulis [25]. Variability related to both these two sources is difficult to be assessed analytically. Thus, simulation experiments are needed in order to evaluate the properties of control charts designed in such a way.

Our simulation experiment has two phases as in actual applications. First we simulate a sample of n elements, and the results of this simulation are used for the design of a control chart. The minimal number of observations which is suggested for designing a chart should be, according to many authors, such as e.g. Quesenberry [20], not smaller than 300. However, in the majority of popular textbooks on quality control this minimal value is proposed to be equal to 100. Having in mind our main purpose, i.e. to investigate the influence of different types of dependence on the performance of control charts actually used in practice, in our experiments we set the sample size (the number of consecutive observations that are used for the design of a chart) as equal to 100. In the experiment we have simulated 500 different control charts, and for each of them we have simulated 500 production runs. Thus, for each experiment described by the chosen copula and the given value of Kendall's τ we have had altogether 250 000 simulation runs. These runs have been used for the estimation of the *ARL*, and other statistical properties of the chart.

Table 4 *ARLs*–Shewhart control chart (design parameters estimated), TEST 1: "3-σ" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	2111,72	x	346,65	568,74	1192,92
0,5	967,84	x	974,28	437,75	1142,76
0,3	540,38	x	730,8	487,09	799,15
0,1	473,11	490,15	560,42	508,98	586,61
0	**486,14**	**486,14**	**486,14**	**486,14**	**486,14**
-0,1	503,28	458,77	447,48	458,49	x
-0,3	589,39	x	537,13	512,63	x
-0,5	1242,9	x	861,8	593,1	x
-0,8	9966,6	x	6255,6	1253,5	x

Table 4 contains the results of the simulation experiment similar to those presented in Table 1, i.e. in presence of no shift in the level of a process. Somewhat unexpectedly these results are different not only with respect to the simulated values of the *ARL*. The dependence of the *ARL* upon the value of τ is also somewhat different than that represented in Table 1. The results of the experiment displayed in Table 4 show that the existence of dependence of any type results, in general, in increasing value of the *ARL*. Only in few cases, in presence of weak dependence, the values of *ARL* are slightly smaller than in the case of independence. When the strength of dependence is low, the values of the *ARL* are similar. Only for Gumbel's copula this value is visibly larger than in the case of independence. For moderate values of Kendall's τ practically acceptable worsening of the value of the *ARL* can be noticed only in the case of Frank's copula. In the case of strong dependence, both positive and negative, the values of the *ARL* are large enough to make the chart insensible to the process deterioration of that magnitude. An interesting, and difficult to explain, exception is the case of Clayton's copula where the large value of *ARL* for $\tau = 0,5$ decreases to a low value for $\tau = 0,8$. A phenomenon of a similar type is also seen in the case of Gumbel's copula.

Table 5 *ARLs*–Shewhart control chart (design parameters estimated), TEST 1: "3-σ" rule, shift of 1σ

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	352,88	x	55,77	80,64	206,43
0,5	84,7	x	102,46	55,72	205,43
0,3	59,61	x	57,34	51,55	92,76
0,1	54,88	51,4	50,51	49,12	58,54
0	**48,09**	**48,09**	**48,09**	**48,09**	**48,09**
-0,1	48,9	50,72	50,15	47,39	x
-0,3	56,1	x	56,07	53,08	x
-0,5	78,25	x	80,61	59,4	x
-0,8	769,18	x	433,52	89,26	x

When the magnitude of the process deterioration is large (i.e. when the shift in the process level is equal to 3σ) the picture is anew different. First of all, it can be noticed that in the case of small and moderate negative dependencies the value of the *ARL* may be smaller than in the case of independence. It means that negative dependence, unless it is not too strong, has a positive impact on the ability of the chart to detect deteriorations of large magnitude. In case of strong negative dependence the situation is different, and the value of the *ARL* usually becomes too large. In the case of the normal copula this value becomes completely unacceptable. In case of positive dependence good properties of the chart are observed for Frank's and Gumbel's copulas.

The results presented in Tables 4–6 show a very complicated situation. Only in the case of Frank's copula the performance of the Shewhart control chart is more or less robust to the existence of dependence between consecutive observations. In

Table 6 *ARLs*–Shewhart control chart (design parameters estimated), TEST 1: "3-σ" rule, shift of 3σ

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	7,35	x	11,95	4,80	2,00
0,5	3,27	x	3,94	2,68	2,54
0,3	2,54	x	2,70	2,26	2,55
0,1	2,15	2,05	2,18	2,15	2,19
0	**2,06**	**2,06**	**2,06**	**2,06**	**2,06**
-0,1	1,92	2,07	1,96	1,87	x
-0,3	1,85	x	1,95	2,00	x
-0,5	1,92	x	2,09	2,02	x
-0,8	21,47	x	4,22	2,45	x

all remaining cases one cannot observe situations which are difficult to describe and explain. Only in the case of the normal copula the dependence of the *ARL* on the strength of dependence can be described in a relatively simple way: the chart is completely insensible to process shifts only in the case of strong, both positive and negative, dependence.

ARL is the most frequently used statistical characteristic of control charts. Another characteristic which is often calculated is the variance of the run length. Specialist are fully aware of the fact that the run length is a highly skewed random variable, and these two characteristics are not sufficient for the comprehensive description of the statistical properties of control charts. The coefficient of skewness whose value equal to 2 is well known for the chart with known design parameters is rarely calculated for other cases. In Table 7 we present the values of the coefficient of skewness of the run length for the case of estimated design parameters and shift equal to 3σ.

Table 7 Skewness of the run length - Shewhart control chart (design parameters estimated), shift of 3σ

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	9,53	x	7,18	5,89	2,1
0,5	3,7	x	4,05	4,11	7,16
0,3	3,02	x	7,47	3,63	3,98
0,1	2,7	2,54	3,3	2,7	2,94
0	**2,49**	**2,49**	**2,49**	**2,49**	**2,49**
-0,1	2,12	2,54	2,6	2,8	x
-0,3	3,11	x	3,75	2,32	x
-0,5	6,87	x	6,47	2,72	x
-0,8	27,96	x	10,59	6,08	x

The values given in Table 7 show that the times to alarm are highly skewed, especially in case of strong (both positive and negative) dependence. In practice it means that despite reasonable values of the *ARL* there is quite substantial possibility that even significant process deterioration may not be detected sufficiently quickly.

4 Tests based on runs in case of dependent data

Classical Shewhart control chart has been supported by additional decisions rules based on runs. Different rules have been proposed by many authors, but the most popular ones were proposed in the Western Electric handbook in 1956. They are also described in the international standard ISO 8258 and in the paper by Nelson [19]. These rules are designed with the aim to detect deteriorations of different type. Statistical properties of control charts with supporting run rules can be computed using the Markov chain approach. A general solution of this problem has been proposed in the paper by Champ and Woodall [3]. This methodology has been successfully implemented for the calculations made under the assumption of independence of observations, and full knowledge of the values of design parameters. However, in case of dependent observations, and for estimated values of design parameters such computations are very difficult or even hardly possible. Therefore, in our analysis we used the results of the Monte Carlo simulation experiments. The settings of these experiments are the same as in the cases described in the previous sections of this paper.

Classical Shewhart control chart with 3-sigma control limits in the set of rules proposed in the Western Electric handbook is called TEST 1. The next rule, named in the handbook TEST 2, triggers an alarm when 9 consecutive observations are plotted on the chart either above or below the center line. This rule should be sensitive to upwards or downwards shifts of the process level. The average run length to false alarm for this test are presented in Table 8. The estimated value of the *ARL* when data are independent is quite large (512,87). However it decreases rapidly when consecutive observations are positively dependent. In case of strong positive dependence the expected time to a false alarm is astonishingly small (about 13), and practically does not depend upon the type of dependence. On the other hand, in presence of negative dependence the values of ARL_0 are increasing to very large values for the case on normal and Frank copulas. It has to be noted, however, that in the case of dependence described by the Clayton copula the behaviour of ARL_0 as a function of Kendall's τ is not monotonic, and in case of negative dependence is significantly different than in the cases of normal and Franks copulas.

One of the most popular rule, known as TEST 3 or "6 increasing (decreasing) in a row", is used for the detection of harmful trends. The properties of this test do not depend upon the design parameters, and may be evaluated using recently published results of Ferguson *et al.* [4]. In Table 9 we present the values of the *ARL* for this particular test when the process is in the in-control state.

These results show that dependencies have detrimental impact on the properties of this test. Tn case of positive dependence the average time to a false alarm becomes unacceptably small. On the other hand, the negative dependence (especially

Table 8 *ARLs*–Shewhart control chart (design parameters known), TEST 2: "9 consecutive below(above) center" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	13,21	x	12,33	13,01	13,18
0,5	29,4	x	29,01	25,58	29,83
0,3	74,96	x	70,18	65,9	74,28
0,1	256,53	512,17	255,74	238,02	257,31
0	**512,87**	**512,87**	**512,87**	**512,87**	**512,87**
-0,1	1036,5	510,18	964,18	1151,15	x
-0,3	4183,7	x	2074,9	6554,6	x
-0,5	14756,3	x	1589,4	32137,8	x
-0,8	85377	x	630,3	>100000	x

Table 9 *ARLs*–Shewhart control chart (design parameters known), TEST 3: "6 in a row" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	24,8	x	27,3	26,1	24,3
0,5	30,1	x	32,4	30,9	30,26
0,3	48,2	x	47,4	48,4	46,1
0,1	97,1	147,3	95,3	96,1	92,4
0,05	119,2	146,4	117,1	118,5	115,0
0,01	140,4	147,1	140,1	141,1	138,0
0	**147,1**	**147,1**	**147,1**	**147,1**	**147,1**
-0,01	153,66	148,50	153,85	155,59	x
-0,05	183,2	147,1	182,6	186,2	x
-0,1	225,7	146,6	225,8	236,7	x
-0,3	539,14	x	507,63	665,26	x
-0,5	1228,96	x	1237,65	1831,29	x
-0,8	4723,89	x	13773,52	5347,5	x

the strong one) may decrease the ability of the test to detect trends in data. Similar results, which are not presented in this paper because of its limited volume, have been observed in preliminary experiments for the case of deteriorated processes.

In Table 10 we present the results of the simulation experiment for the TEST 4. In case of this test an alarm is set when 14 consecutive observations are alternating, i.e. the differences between consecutive values are alternately positive and negative. The role of this rule is to show that consecutive observations are in fact from two alternating sources, e.g. from two production shifts. The results of simulations are presented in Table 10. Positive dependence between consecutive observations increases, as expected, the average run length. Negative dependence shortens the time to a false alarm significantly, and this behaviour is hardly unexpected as the negative dependence forces consecutive observation values to alternate.

Another popular additional decision rule, known as TEST 5 or "2 out of 3 in a row observation in an outer zone", is used to improve the ability to detect small

Table 10 *ARL*s–Shewhart control chart (design parameters known), TEST 4: "14 consecutive alternating" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	9920,6	x	7878,9	7699,3	10124,1
0,5	5541,4	x	4285,5	4982,4	5341,5
0,3	2841,6	x	2483,2	2827,4	2816,3
0,1	1239,2	803,9	1238,5	1274,4	1255,6
0	**796,5**	**796,5**	**796,5**	**796,5**	**796,5**
-0,1	494,8	800,7	484,2	480,1	x
-0,3	172,1	x	160,9	164,5	x
-0,5	59,0	x	60,5	53,9	x
-0,8	21,2	x	19,2	22,65	x

shifts of the process level. In Table 11 we present the values of the *ARL* for the case of known parameters of the process.

Table 11 *ARL*s–Shewhart control chart (design parameters known), TEST 5: "2 out of 3 in the outer zone" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	165,6	x	206,3	102,1	169,8
0,5	95,9	x	184,5	159,4	111,9
0,3	143,0	x	155,5	249,2	130,9
0,1	323,3	510,4	245,8	413,6	222,9
0	**510,1**	**510,1**	**510,1**	**510,1**	**510,1**
-0,1	321,9	510,0	620,6	598,5	x
-0,3	401,6	x	335,4	567,6	x
-0,5	193,2	x	165,4	379,9	x
-0,8	209,7	x	134,8	201,1	x

In the case of TEST 5 we observe non-monotonic dependence between the *ARL* and the strength of dependence. In case of positive dependence the values of *ARL* are decreasing with the increasing strength of dependence, but for the very strong dependence ($\tau = 0,8$) they begin to decrease. In case of negative dependence the dependence of the value of the *ARL* and the value of τ cannot be easily explained at the current stage of our research. For example, in case of the normal copula this dependence is highly non-monotonic. On the other hand, in case of Frank's copula the largest value of the *ARL* is observed for small negative dependence, and then the value of the *ARL* decreases with increasing (decreasing) values of τ. Interesting is the case of the FGM copula where in contrast to other considered cases the existing weak dependence does not influence the value of the *ARL*.

TEST 6 has a similar decision rule, and similar application. The alarm is triggered when 4 out of 5 consecutive observations are beyond the inner zone on the same side of the center line. The results of the simulation experiment are in this case given in Table 12.

Table 12 *ARLs*–Shewhart control chart (design parameters known), TEST 6: "4 out of 5 beyond the inner zone" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	30,7	x	32,8	24,9	31,0
0,5	31,5	x	39,1	36,1	33,0
0,3	58,3	x	56,4	68,7	54,9
0,1	193,4	290,5	140,1	171,4	138,5
0	**291,7**	**291,7**	**291,7**	**291,7**	**291,7**
-0,1	598,0	292,1	611,1	540,5	x
-0,3	4561,2	x	1170,8	2823,3	x
-0,5	>200000	x	1830,8	>50000	x
-0,8	>500000	x	9275,9	>300000	x

Despite the similar aim and type of construction the behaviour of TEST 6 is completely different than that of TEST 5. In case of positive dependence the values of *ARL* are decreasing rapidly withe the increasing strength of dependence. In case of negative dependence the type of behaviour is reversed; the values of *ARL* are increasing to very large numbers for the case of $\tau = -0,8$.

TEST 7, with the rule "15 consecutive observations in the inner zone", is used for the detection of faulty (too wide) decision rules. The values of the *ARL* in this case are given in Table 13. The values of *ARL*'s are decreasing with increasing values of the absolute value of τ. In the case of this rule the type of dependence described by different copulas does not seem to influence the results of the experiment.

Table 13 *ARLs*–Shewhart control chart (design parameters known), TEST 7: "15 consecutive in the inner zone" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	41,7	x	50,1	56,6	42,2
0,5	186,7	x	180,4	236,1	163,8
0,3	513,0	x	435,8	577,3	403,2
0,1	937,1	963,9	810,4	909,2	772,0
0	**963,2**	**963,2**	**963,2**	**963,2**	**963,2**
-0,1	898,5	969,3	991,9	905,2	x
-0,3	506,3	x	456,7	577,5	x
-0,5	187,7	x	144,1	237,5	x
-0,8	41,6	x	41,4	56,9	x

TEST 8 is designed to to detect stratification. Its decision rule "15 consecutive observations beyond the inner zone" let us detect inhomogeneous data coming from two sources with significantly different average values. In Table 14 we present the values of the *ARL* for false alarms triggered by this rule. The behaviour of the values of *ARL* requires further investigations. The values of *ARL* are increasing rapidly when data are weakly (both positively and negatively) dependent, and then begin to decrease rapidly with the increasing strength of dependence.

Table 14 *ARLs*–Shewhart control chart (design parameters known), TEST 8: "15 consecutive beyond the inner zone" rule, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	70,2	x	136,1	103,6	74,9
0,5	319,3	x	186,7	785,8	220,2
0,3	2648,2	x	533,4	4230,8	854,0
0,1	13281,9	14086,9	7385,3	12320,1	6895,9
0	**963,2**	**963,2**	**963,2**	**963,2**	**963,2**
-0,1	11797,8	14280,0	14423,3	12374,4	x
-0,3	2636,7	x	1938,1	4249,1	x
-0,5	318,8	x	273,7	778,3	x
-0,8	70,2	x	62,4	103,1	x

In our experiments we have also calculated the properties of chart with combined decision rules. When TEST 1 is combined with TEST 5 the *ARL* in case of estimated design parameters and independence has been evaluated as equal to 293,11. The exact calculations performed for this case, but for known values of design parameters, by Champ and Woodall [3] gave the value of the *ARL* equal to 225,44.

Table 15 *ARLs*–Shewhart control chart (design parameters known), TEST 1+TEST 5, no shift

Kendall's τ	Normal	FGM	Clayton	Frank	Gumbel
0,8	165,44	x	180,76	91,68	168,74
0,5	93,46	x	151,39	122,71	105,92
0,3	120,51	x	130,32	161,28	115,87
0,1	187,27	226,65	169,16	207,24	165,62
0	**225,45**	**225,45**	**225,45**	**225,45**	**225,45**
-0,1	250,41	226,58	244,68	240,57	x
-0,3	222,3	x	201,0	235,94	x
-0,5	168,26	x	145,03	202,07	x
-0,8	209,72	x	131,67	149,55	x

The results presented in Table 15 confirm this value, and additionally show how the *ARL* in the case of this combination of tests depends on the type and the strength of dependence. It can be noted that the dependence of the *AL* on the value of τ is similar to that of TEST 5.

5 Conclusions

The results presented in this paper confirm without any doubts the findings of many authors who considered the bahaviour of Shewhart control charts in case of dependent data described by autoregressive stochastic processes. What seems to be new is the demonstration that that the type of dependence, encapsulated in the type of respective copula, plays important role. Moreover, it becomes very clear that the knowledge of the strength of dependence, measured using popular statistical measures of dependence such as Kendall's τ is not sufficient for the evaluation of the properties of the Shewhart control chart.

From the results presented in this paper one can derive the following recommendations. First, it is necessary to detect the existence of dependence in data. This can be done using the Kendall control chart proposed by Hryniewicz and Szediw [8]. Then, it is necessary to indicate the copula which fits to the observed data. Unfortunately, the appropriate tests, such as presented e.g. in the paper by Fermanian [5], seem to be not simple enough to be used by quality control practitioners. Therefore, a lot has to be done in order to propose even approximate but simple methods for the identification of an actual copula. Then, the future investigations should be concentrated on finding appropriate corrections to clasical procedures, similar in spirit to those that has been proposed in case of dependencies described by the normal copula.

References

1. Albers W., Kallenbeg W.C.M (2004), Estimation in Shewhart control charts: effects and corrections. Metrika 59: 207–234
2. Alwan L.C., Roberts H.V. (1988), Time-Series Modeling for Statistical Process Control, Journal of Business & Economic Statistics 6: 87–95
3. Champ C.W., Woodall W.H. (1987), Exact Results for Shewhart Control Charts With Supplementary Runs Rules, Technometrics 29: 393–399
4. Ferguson T.S., Genest C., Hallin, M. (2000), Kendall's tau for Serial Dependence, The Canadian Journal of Statistics 28: 587–604
5. Fermanian J-D. (2005), Goodness-of-fit tests for copulas, Journal of Multivariate Analysis, 95: 119–152
6. Genest C., McKay R.J. (1986), The joy of copulas: Bivariate distributions with uniform marginals, American Statistician 88:1034–1043
7. Genest C., Rivest L-P. (1993), Statistical Inference Procedures for Bivariate Archimedean Copulas, Journal of the American Statistical Association 88:1034–1043
8. Hryniewicz O., Szediw A. (2010), Sequential Signals on a Control Chart Based on Nonparametric Statistical Tests. In: H.-J. Lenz, P.-Th. Wilrich, W. Schmid (eds) Frontiers in Statistical Quality Control 9, Physica-Verlag, Heidelberg, 99–118
9. Jiang W., Tsui K-L., Woodall H.W. (20000, A New SPC Monitoring Method: The ARMA Chart, Technometrics 42: 399–410
10. Johnson R.A., Bagshaw M. (1974), The Effect of Serial Correlation on the Performance of CUSUM Tests, Technometrics 16:103–112

11. Knoth S., Schmid, W., Schone, A. (2001), Simultaneous Shewhart-Type Charts for the Mean and the Variance of a Time Series. In: Lenz, H.J. and Wilrich, P.T. (eds) Frontiers in Statistical Quality Control 6, Physica-Verlag, Heidelberg, 61–79
12. Lu C.W., Reynolds Jr. M.R. (1999), EWMA Control Charts for Monitoring the mean of Autocorrelated Processes, Journal of Quality Technology 31: 166–188
13. Lu C.W., Reynolds Jr. M.R. (1999), Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes, Journal of Quality Technology 31: 259–274
14. Lu C.W., Reynolds Jr. M.R. (2001), CUSUM Charts for Monitoring an Autocorrelated Process, Journal of Quality Technology 33: 316–334
15. Maragah, H.D., Woodall, W.H. (1992), The Effect of Autocorrelation on the Retrospective X-Chart, Journal of Statistical Computation and Simulation 40: 29–42
16. Montgomery D.C., Mastrangelo C.M. (1991), Some Statistical Process Control Methods for Autocorrelated Data, Journal of Quality Technology 23: 179–193
17. Nelsen R.B. (2006), An Introduction to Copulas, Springer-Verlag, New York
18. Nelsen R.B., Quesada-Molina J.J. Rodriguez-Lallena J.A., Úbeda-Flores M. (2009) Kendall distribution functions and associative copulas, Fuzzy Sets & Systems 160:52–57
19. Nelson L.S. (1984), The Shewhart Control Chart - Tests for Special Causes, Journal of Quality Technology 16: 237–239
20. Quesenberry C.P. (1993) The effect of sample size on estimated limits of $\bar{X}$ and X control charts. Journal of Quality Technology 25: 237–247
21. Schmid W. (1995), On the run length of a Shewhart control chart for correlated data, Statistical Papers 36: 111–130
22. Schmid W. (1997), On EWMA charts for time series, In: Lenz, H.J. and Wilrich, P.T. (eds) Frontiers in Statistical Quality Control 5, Physica-Verlag, Heidelberg, 115–137
23. Sklar A. (1959), Fonctions de répartition á n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8:229–231
24. VanBrackle III L.N., Reynolds Jr. M.R. (1997), EWMA and CUSUM Control Charts in Presence of Correlations, Communications in Statistics Simulation and Computation 26: 979–1008
25. Vasilopoulos A.V., Stamboulis A.P. (1978), Modification of Control Limits in the Presence of Correlation, Journal of Quality Technology 10: 20–30
26. Wardell D.G., Moskowitz, H., Plante, R.D. (1994), Run-Length Distributions of Special-Cause Control Charts for Correlated Processes, Technometrics 36: 3–17
27. Woodall W.H., Montgomery D.C. (1999), Research issues and ideas in statistical process control, Journal of Quality Technology, 25: 188–198
28. Yourstone S.A., Montgomery D.C. (1991), Detection of Process Upsets Sample Autocorrelation Control Chart and Group Autocorrelation Control Chart Applications, Quality and Reliability Engineering International 7: 133–140
29. Zhang N.F. (1998), Statistical Control Chart for Stationary Process Data, Technometrics 40: 24–38