F O L I A   O E C O N O M I C A   C R A C O V I E N S I A
Vol. LIV (2013) PL ISSN 0071-674X

STATYSTYCZNE STEROWANIE PROCESAMI O DANYCH STOCHASTYCZNIE ZALEŻNYCH — PUŁAPKI ROZWIĄZAŃ STANDARDOWYCH

OLGIERD HRYNIEWICZ
Instytut Badań Systemowych PAN
e-mail: hryniewi@ibspan.waw.pl

ABSTRACT

O. Hryniewicz. *Statistical process control for stochastically dependent data — pitfalls of using standard solution.* Folia Oeconomica Cracoviensia 2013, 54: 93–106.

Shewhart control charts are the most frequently used tools of statistical process control. In their standard form they are designed under the assumption that consecutive observations are statistically independent and described by the normal distribution. When these assumptions are not fulfilled statistical properties of the Shewhart control charts are different from those assumed for the design purposes. When consecutive observations are not independent the properties of some Shewhart control charts have been investigated only in the case of classic autoregression processes. Hryniewicz (2012) considered the influence of the type of dependence, described in terms of copulas, on the properties of the Shewhart charts for monitoring the mean value of the process. In this paper some results from Hryniewicz (2012) have been recalled. Some new results, obtained for the *R*-chart used for the control of the variability of a process, have been presented.

STRESZCZENIE

Karty kontrolne Shewharta są najczęściej stosowanym narzędziem statystycznego sterowania procesami. W swojej podstawowej postaci są one projektowane przy założeniu, że kolejne obserwacje procesu są statystycznie niezależne, i że są opisane rozkładem normalnym. Jeśli powyższe założenia nie są spełnione, to własności statystyczne kart kontrolnych Shewharta różnią się od tych, które zakłada się w procesie projektowania. Gdy kolejne obserwacje nie są niezależne, własności niektórych kart kontrolnych Shewharta zostały zbadane dla przypadku klasycznych procesów autoregresji. Hryniewicz (2012) rozpatrywał wpływ typu zależności pomiędzy obserwacjami, opisanego za pomocą pojęcia kopuli, na własności karty Shewharta służącej do monitorowania wartości średniej procesu. Niektóre z własności karty Shewharta omawiane w tamtej pracy zostały przypomniane w niniejszym opracowaniu, które zawiera ponadto nowe wyniki dotyczące analogicznego zagadnienia w odniesieniu do karty kontrolnej *R*, służącej do sterowania zmiennością monitorowanych procesów.

KEY WORDS — SŁOWA KLUCZOWE

Karty kontrolne Shewharta, obserwacje zależne, średnia długość przebiegu

Shewhart control charts, dependent observations, average run length

1. WPROWADZENIE

Statystyczne sterowanie procesami (*Statistical Process Control — SPC*) zostało zaproponowane przez Waltera Shewharta w latach dwudziestych ubiegłego wieku. Przez wiele lat stanowiło mało popularny dział statystycznego sterowania jakością (*Statistical Quality Control — SQC*), zwanego tradycyjnie w Polsce statystyczną kontrolą jakości (SKJ). Przez wiele lat procedury statystycznej kontroli odbiorczej stanowiły podstawowe narzędzia SQC, co — moim zdaniem — w pełni uzasadniało tradycyjną wersję polskiego tłumaczenia angielskiego terminu SQC. Dopiero wprowadzenie do repertuaru narzędzi SPC statystycznych metod sekwencyjnych takich jak karty kontrolne sum skumulowanych (CUSUM) spowodowało, że zagadnieniami SPC zaczęło zajmować się na świecie wielu badaczy. Jeżeli dodamy, że praktyka produkcyjna wymusiła zwiększenie wagi narzędzi SPC względem wagi narzędzi statystycznej kontroli odbiorczej, to stanie się jasne, że obecnie statystyczne sterowanie procesami (SPC) stanowi główny dział statystycznego sterowania jakością, a przez wielu praktyków jest z nim po prostu utożsamiane.

Statystyczna kontrola jakości, rozumiana w zasadzie jako statystyczna kontrola odbiorcza, była przedmiotem zainteresowania polskich naukowców (Obalski, Oderfeld, Steinhaus) od początku lat pięćdziesiątych ubiegłego wieku. W tym czasie powstało wiele polskich norm statystycznej kontroli odbiorczej, które wyróżniały się na świecie swoją oryginalnością. Natomiast problemami SPC zajmowało się w Polsce niewielu praktyków, nie wykraczając poza zastosowanie znanych na całym świecie (i stosowanych powszechnie do dzisiaj) kart kontrolnych Shewharta. Śp. Profesor Andrzej Iwasiewicz, pamięci którego poświęcone jest niniejsze wydawnictwo, był pierwszym polskim naukowcem, który prowadził badania naukowe nad nowymi narzędziami SPC, a także wprowadzał je do praktyki. Najważniejszym owocem jego prac w tym zakresie była wydana przez PWN oryginalna monografia *Statystyczna kontrola jakości w toku produkcji: systemy i procedury* — Iwasiewicz (1985). Można więc przyjąć, że był On w Polsce prekursorem badań nad nowoczesnymi metodami statystycznego sterowania procesami.

Od okresu pionierskiego w zakresie badań nad procedurami SPC upłynęło już wiele lat. W międzyczasie powstało wiele prac naukowych napisanych przez wybitnych naukowców i poświęconych temu zagadnieniu. Dla potrzeb statystycznego sterowania procesami zaproponowano wykorzystanie zaawansowanych metod statystycznych, często powstałych w innych obszarach zastosowań statystyki matematycznej. Powstało też wiele prac, w których zaproponowano różne metody pozwalające odejść od podstawowego paradygmatu tradycyjnego SPC, jakim jest przyjęcie *niezależności* wyników pomiarów opisujących te procesy charakterystyk jakościowych. Zostały zaproponowane nowe statystyczne procedury kontrolne, które odznaczały się jednak dużym (w ocenie praktyków) stopniem skomplikowania. W pracach tych przyjmowano pewne popularne modele zależności pomiędzy wynikami kolejnych pomiarów, nie wykraczając przy tym poza opis procesu wykorzystujący pojęcie wielowymiarowego rozkładu normalnego.

W pionierskiej pracy — Hryniewicz i Szediw (2010) pokazano, że w przypadku występowania innych typów zależności stochastycznej oraz różniących się od normalnego rozkładów wartości charakterystyk jakościowych, stosowanie zaproponowanych dotychczas procedur może prowadzić do błędnych decyzji. Sytuacja staje się wręcz dramatyczna, gdy do sterowania jakością procesów o nietypowo zależnych obserwacjach wykorzystamy tradycyjne metody SPC, jakimi są karty kontrolne Shewharta. Po raz pierwszy zwrócono na to uwagę w pracy Hryniewicza (2012).

Niniejszy artykuł można traktować jako pewne uzupełnienie wymienionych powyżej prac. Pokazano w nim pułapki, z jakimi mogą spotkać się praktycy, którzy stosują powszechnie znane (standardowe) metody statystyczne do sterowania procesami o obserwacjach zależnych. W kolejnych sekcjach artykułu przedstawiony zostanie w dużym skrócie zarys metodologii SPC dla procesów o zależnych obserwacjach. Następnie zostaną omówione problemy z praktyczną oceną siły zależności pomiędzy takimi obserwacjami. Na zakończenie przedstawione zostaną przykłady pokazujące potencjalne niebezpieczeństwa związane z zastosowaniem klasycznych kart kontrolnych Shewharta w przypadku statystycznego sterowania procesami o danych zależnych.

2. STATYSTYCZNE STEROWANIE PROCESAMI (SPC) W PRZYPADKU DANYCH ZALEŻNYCH

Podstawowym założeniem, jakie przyjmuje się konstruując standardowe karty kontrolne (Shewharta, CUSUM, EWMA) jest *wzajemna niezależność* kolejnych obserwacji kontrolowanego procesu. W przypadku klasycznej karty Shewharta przekłada się to na niezależność kolejnych punktów wykreślanych na karcie kontrolnej, co oczywiście nie jest prawdą w przypadku bardziej skomplikowanych procedur takich jak karty CUSUM lub EWMA. Drugim, stosowanym np. w normach założeniem jest przyjęcie rozkładu normalnego jako probabilistycznego modelu opisującego wyniki pomiarów. Nieadekwatność pierwszego z tych założeń w przypadku wielu procesów (zwłaszcza procesów ciągłych, jak np. procesy chemiczne) wykazano w wielu pracach takich jak np. Wardell *i in.* (1994) oraz

Alwan i Roberts (1995). W tej sytuacji starano się zaproponować rozwiązania, które pozwoliłyby na korzystanie z metod SPC w przypadku sterowania procesami o zależnych obserwacjach.

Pierwsze prace poświęcone kartom kontrolnym (głównie kartom Shewharta) pojawiły się w latach siedemdziesiątych ubiegłego stulecia. Należą do nich prace: Johnson i Bagshaw (1974) oraz Vasilopoulos i Stamboulis (1978). Można je traktować jako prekursorów całego nurtu prac związanych z nazwiskami Lu, Reynoldsa Jr. oraz Schmida, w których proponowano różne modyfikacje klasycznych kart kontrolnych, uwzględniające określone rodzaje zależności pomiędzy obserwacjami procesów. Inny nurt został zainicjowany pracą — Alwan i Roberts (1988), w której do sterowania procesami o danych zależnych zaproponowano zastosowanie klasycznych kart kontrolnych, ale wykorzystanych do analizy residuów, a nie do oryginalnych pomiarów. Istnieje też trzeci nurt zapoczątkowany pracą: Yourstone i Montgomery (1989), w którym do monitorowania procesów o danych zależnych zaproponowano kartę kontrolną współczynnika autokorelacji. Szersze omówienie tych podejść można znaleźć w niepublikowanej rozprawie doktorskiej, zob. Olwert (2012).

3. BADANIE ZALEŻNOŚCI W PROCESACH

Zagadnienie badania zależności pomiędzy kolejnymi obserwacjami procesu jest szczególnym przypadkiem jednego z najważniejszych zadań statystyki, jakim jest badanie zależności pomiędzy zmiennymi losowymi. Zagadnieniem tym statystycy zajmują się co najmniej od XIX wieku, a poświęcona temu literatura jest niezwykle bogata. W praktyce najczęściej spotykaną miarą zależności pomiędzy obserwacjami w procesie jest współczynnik autokorelacji (rzędu k) będący współczynnikiem korelacji liniowej Pearsona pomiędzy ciągiem obserwacji a tym samym ciągiem przesuniętym o k jednostek i w wersji populacyjnej zdefiniowany dla procesów stacjonarnych jako:

$$R_k = \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{\sigma^2}, \tag{1}$$

gdzie μ jest wartością oczekiwaną zmiennej losowej X, a σ^2 jej wariancją, przy czym założona stacjonarność obserwowanego procesu pociąga za sobą stałość tych dwu parametrów w czasie.

Współczynnik autokorelacji R_k jest, jako szczególny przypadek współczynnika korelacji Pearsona, miernikiem zależności liniowej. O jego ograniczonej przydatności do oceny zależności w przypadkach nieliniowych, np. nieliniowych zależnościach regresyjnych, można przeczytać w każdym podręczniku statystyki. Jest jednak rzeczą znaną mniej powszechnie, że jest on dobrą miarą zależności pomiędzy dwiema zmiennymi losowymi wyłącznie wtedy, gdy ich łączny rozkład należy do rodziny tzw. rozkładów elipsoidalnych, przy czym wniosek odwrotny nie jest już uprawniony. W praktyce statystycznej oznacza to, że współczynnik korelacji liniowej Pearsona jest dobrą miarą zależności tylko w przypadku, gdy badane zmienne losowe opisane są łącznym dwuwymiarowym rozkładem normalnym.

W celu poszukania bardziej uniwersalnej miary zależności pomiędzy zmiennymi losowymi rozpatrzmy przypadek najbardziej ogólny. Z fundamentalnego twierdzenia Sklara (1959) wynika, że łączny rozkład prawdopodobieństwa $F(x,y)$ dowolnych zmiennych losowych X i Y jest w sposób jednoznaczny opisany funkcją C zwaną *kopulą*

$$F(x, y) = C\big(F(x), G(y)\big), \tag{2}$$

przy czym funkcje $F(x)$ oraz $G(y)$ są, odpowiednio, dystrybuantami rozkładów brzegowych zmiennych losowych X oraz Y, a funkcję C definiuje się jako:

$$C: [0,1]^2 \to [0,1], \tag{3}$$

spełniającą warunki:

$$\forall u, v \in \langle 0,1\rangle \; C(0, v) = 0, C(u, 0) = 0,$$

$$\forall u, v \in \langle 0,1\rangle \; C(1, v) = v, C(u, 1) = u,$$

$$\forall u_1, u_2, v_1, v_2 \in \langle 0,1\rangle, u_1 < u_2, v_1 < v_2,$$

$$C(u_1, v_2) + C(u_2, v_1) \leq C(u_1, v_1) + C(u_2, v_2).$$

Jak łatwo zauważyć, miary zależności pomiędzy zmiennymi losowymi, których łączny rozkład opisany jest kopulą, powinny zależeć wyłącznie od typu kopuli i jej parametrów. Tymczasem, współczynnik korelacji Pearsona zależy (o czym wiadomo od bardzo dawna) od postaci rozkładów brzegowych.

Dla dowolnej kopuli dwuwymiarowej słuszne są nierówności znane w teorii prawdopodobieństwa (i wyznaczone przed wprowadzeniem pojęcia kopuli) jako nierówności Frechéta–Hoeffdinga

$$max(0, u + v - 1) \leq C(u, v) \leq min(u, v). \tag{4}$$

Granice Frechéta–Hoeffdinga odpowiadają przypadkom minimalnej i maksymalnej sile zależności pomiędzy zmiennymi losowymi. W pracy Hryniewicza i Karpińskiego (2012) wykorzystano je w celu wykazania istnienia następujących własności współczynnika korelacji liniowej Pearsona.

Własność 1: Jeśli rozkłady prawdopodobieństwa zmiennych losowych X oraz Y mają ten sam kształt, wówczas maksymalna wartość współczynnika korelacji Pearsona jest równa 1 ($r_{max} = 1$).

Własność 2: Jeśli rozkłady prawdopodobieństwa zmiennych losowych X oraz Y są symetryczne względem zera i mają ten sam kształt, wówczas minimalna wartość współczynnika korelacji Pearsona jest równa -1 ($r_{min} = -1$).

Własność 3: Jeśli przynajmniej jedna ze zmiennych losowych X oraz Y ma rozkład symetryczny, wówczas $r_{\min} = -r_{\max}$.

Przykład: Niech zmienne losowe X oraz Y mają ten sam wykładniczy rozkład prawdopodobieństwa o wartości oczekiwanej równej 1. Wówczas mamy:

$$r_{\min} = Cov_{neg}(X,Y) = \int_0^\infty (x-1)\left(\left\{-\ln\left[1-e^{-x}\right]\right\}-1\right)e^{x}dx = 1-\frac{\pi^2}{6} = -0{,}644934. \quad (5)$$

Całka w (5) została wyznaczona z wykorzystaniem obliczeń symbolicznych i numerycznych wykonanych za pomocą pakietu matematycznego MATHEMATICA™.

Jak widać, zastosowanie współczynnika korelacji Pearsona do oceny siły zależności pomiędzy zmiennymi losowymi może prowadzić do zupełnie błędnych wniosków, zwłaszcza, gdy zmienne te mają niesymetryczne rozkłady prawdopodobieństwa i odznaczają się ujemną zależnością. W związku z tym, do oceny siły zależności pomiędzy wartościami kolejnych obserwacji procesu lepiej wykorzystywać nieparametryczne miary zależności nie mające niepożądanych właściwości współczynnika korelacji Pearsona. Jedną z nich jest współczynnik korelacji rangowej τ Kendalla.

Współczynnik korelacji rangowej τ Kendalla został zaproponowany w roku 1938 i wykorzystuje pojęcie *zgodnych* i *niezgodnych* par obserwacji. Para wektorowych obserwacji (x_i,y_i) (x_j,y_j) ciągłych zmiennych losowych (X,Y) jest *zgodna*, jeśli odpowiadające sobie rangi składowych obu wektorów są zgodne, tzn. albo $R_i > R_j$ oraz $S_i > S_j$ albo $R_i < R_j$ oraz $S_i < S_j$. W przeciwnym wypadku taką parę nazywamy *niezgodną*. Próbkowy współczynnik korelacji Kendalla τ jest definiowany jako

$$\tau_{xy} = 2\frac{\text{liczba par zgodnych} - \text{liczba par niezgodnych}}{n(n-1)}. \quad (6)$$

Wygodna reprezentacja τ została zaproponowana w pracy Genesta i Rivesta (1993) w następującej postaci:

$$\tau_{xy} = \frac{4}{n}\sum_{i=1}^{n} V_i - 1, \quad (7)$$

gdzie

$$V_i = card\{(X_j,Y_j): X_j < X_i, Y_j < Y_i\}/(n-1), i = 1,\ldots,n. \quad (8)$$

Populacyjna wersja τ Kendalla została podana w monografii, zob. Nelsen (2006). Dla dowolnej kopuli wartość τ wyznaczana jest z zależności:

$$\tau(X,Y) = 4\iint_{[0,1]^2} C(u,v)\,dC(u,v) - 1. \tag{9}$$

Warto zauważyć, że w przypadku zmiennych opisanych dwuwymiarowym rozkładem normalnym istnieje następująca zależność pomiędzy współczynnikiem korelacji r Pearsona i współczynnikiem korelacji τ Kendalla:

$$\tau = \arcsin(r)/(\pi/2). \tag{10}$$

Dla innych rozkładów wielowymiarowych takiej zależności nie da się prosto wyznaczyć.

Powszechnie znany współczynnik korelacji rangowej ρ Spearmana może być również wyznaczony dla dowolnej kopuli, jednakże dla najczęściej stosowanych kopul nie istnieją jawne wzory pozwalające przeprowadzać odpowiednie obliczenia.

W przypadku gdy n kolejnych obserwacji procesu tworzy pewien szereg czasowy $Z_1, Z_2, \ldots, Zn$, do analizy ich wzajemnej zależności można wykorzystać współczynnik korelacji τ Kendalla przyjmując, że składowe wektora (X_i, Y_i) wyznacza się z zależności: $X_i = Z_i$ oraz $Y_i = Z_{i+1}$ dla $i = 1,\ldots,n-1$. Oznaczmy przez M losową liczbę niezgodności, tzn. liczbę par wektorów (Z_i, Z_{i+1}) i (Z_j, Z_{j+1}), w przypadku których mamy albo $Z_i < Z_j$ i $Z_{i+1} > Z_{j+1}$, albo $Z_i > Z_j$ i $Z_{i+1} < Z_{j+1}$. Współczynnik korelacji (a w zasadzie autokorelacji) τ Kendalla wyznaczany jest teraz ze wzoru:

$$\tau_n = 1 - \frac{4M}{(n-1)(n-2)}, \tag{11}$$

gdzie

$$M = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} I\left(Z_i < Z_j, Z_{i+1} > Z_{j+1}\right), \tag{12}$$

a $I(A)$ jest funkcją wskaźnikową zbioru A.

W przypadku klasycznej statystyki τ Kendalla danej wzorem (7), jej rozkład w przypadku wzajemnej niezależności zmiennych X oraz Y znany jest od dawna. Jeżeli jednak mamy do czynienia z kolejnymi obserwacjami procesu, to nawet przy niezależnych obserwacjach Z_i, Z_{i+1}, odpowiednie pary wektorów definiujących τ Kendalla nie są niezależne. Rozkład statystyki τ Kendalla został w tym przypadku wyznaczony dopiero w pracy Fergusona i in., (2000). W pracy Hryniewicza i Szediw (2010) skorzystano z tego wyniku do konstrukcji karty kontrolnej pozwalającej monitorować występowanie zależności pomiędzy kolejnymi obserwacjami procesu. Karta ta ma o wiele lepsze właściwości w porównaniu do wspomnianej już karty kontrolnej współczynnika autokorelacji zaproponowanej w pracy: Yourstone i Montgomery (1989).

4. WPŁYW SIŁY I TYPU ZALEŻNOŚCI NA WŁASNOŚCI KARTY KONTROLNEJ SHEWHARTA

Jak już wspomniano we Wstępie, własności kart kontrolnych Shewharta dla przypadku, gdy zależności pomiędzy kolejnymi obserwacjami procesu mają klasyczny charakter regresyjny (autoregresyjny), były badane przez wielu autorów. Przyjmowane w tych pracach założenia, rozpatrywane w terminach teorii kopul, sprowadzają się do przyjęcia jako modelu zależności kopuli gaussowskiej (normalnej) danej zależnością:

$$C_N(u,v) = \Phi_N\left(\Phi^{-1}(u), \Phi^{-1}(v); r\right), \tag{13}$$

gdzie $\Phi_N(x,y;r)$ jest dystrybuantą dwuwymiarowego standaryzowanego rozkładu normalnego o współczynniku korelacji r, zaś $\Phi^{-1}(x)$ jest funkcją odwrotną do dystrybuanty standaryzowanego rozkładu normalnego (funkcją kwantylową). W pracy Hryniewicza (2012) postawiono pytanie o własności karty Shewharta, gdy zależność pomiędzy kolejnymi obserwacjami może być opisana inną kopulą, na przykład:
a) kopulą Claytona, zob. Clayton (1978)

$$C(x,y) = \left[F^{-\theta}(x) + G^{-\theta} - 1\right]^{-1/\theta}, \theta \in \{(-1,\infty) \setminus \{0\}\}, \tag{14}$$

b) kopulą Franka, zob. Frank (1979)

$$C(x,y) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta F(x)} - 1\right)\left(e^{-\theta G(y)} - 1\right)}{e^{-\theta} - 1}\right), \theta \in \{(-\infty,\infty) \setminus \{0\}\}, \tag{15}$$

c) oraz kopulą Gumbela, zob. Gumbel (1960)

$$C(x,y) = \exp\left(-\left[(-\ln F(x))^{\theta} + (-\ln G(y))^{\theta}\right]^{1/\theta}\right), \theta > 0 . \tag{16}$$

Próba odpowiedzi na to pytanie została dokonana z wykorzystaniem metod symulacyjnych (Monte Carlo). Skorzystano tu ze znanych związków pomiędzy postaciami w/w kopul a współczynnikiem τ Kendalla, a także znanych metod generowania zmiennych wektorów losowych opisanych wymienionymi powyżej kopulami. W symulacjach opisanych w pracy Hryniewicza (2012) przyjęto założenie, że rozkłady brzegowe we wszystkich rozpatrywanych przypadkach są rozkładami normalnymi.

Podstawową charakterystyką każdej karty kontrolnej jest oczekiwana liczba kontroli pomiędzy kolejnymi alarmami ARL (*Average Run Length*). Jeżeli proces znajduje się w stanie ustabilizowanym, wartość ARL określana wówczas jako ARL0 oznacza oczekiwany czas pomiędzy fałszywymi alarmami (oczekiwany czas do wystąpienia fałszywego alarmu). W Tablicy 1, opracowanej na podstawie wyników opublikowanych w pracy Hryniewicza (2012) podano, jak zmienia się wartość ARL0 dla karty X–średnie Shewharta w zależności od mierzonej współczynnikiem korelacji Kendalla τ siły zależności pomiędzy kolejnymi obserwacjami procesu oraz od typu kopuli opisującej taką zależność.

Z danych przedstawionych w Tablicy 1 wynika, że występowanie zależności pomiędzy kolejnymi obserwacjami procesu wydłuża oczekiwany czas do wystąpienia fałszywego alarmu, co jest cechą pożądaną. Wielkość takiego wydłużenia zależy jednak zdecydowanie od typu zależności. W przypadku kopuli gaussowskiej (klasyczny proces autoregresji) wzrost jest z początku wolny, ale w przypadku występowania silnych zależności, wartości ARL0 stają się bardzo duże. Można też zauważyć symetrię funkcji ARL0(τ) względem zera. Taka sama symetria występuje w przypadku kopuli Franka, ale zwiększenie wartości ARL0 jest tu zdecydowanie mniejsze. W przypadku kopuli Claytona funkcja ARL0(τ) jest wyraźnie niesymetryczna względem zera. Z kolei, w przypadku kopuli Gumbela, dla której zależności mogą być tylko dodatnie, już dla małych wartości τ występuje bardzo silny wzrost wartości ARL0.

Tablica 1

Wartości ARL0 dla karty X — średnie Shewharta

τ Kendalla	Gauss	Clayton	Frank	Gumbel
0,8	1391,0	759,3	385,4	1301,8
0,5	466,8	622,4	373,1	633,6
0,3	389,0	496,0	373,4	528,6
0,1	371,1	384,4	370,5	456,4
0	370,5	370,5	370,5	370,5
–0,1	371,3	370,6	371,2	X
–0,3	390,0	384,3	374,1	X
–0,5	468,4	433,9	375,0	X
–0,8	1379,4	898,6	384,0	X

Tablica 2

Wartości ARL dla karty X — średnie Shewharta — przesunięcie o 1 σ

τ Kendalla	Gauss	Clayton	Frank	Gumbel
0,8	273,2	90,3	86,7	653,0
0,5	71,9	48,6	54,2	126,7
0,3	52,3	45,2	47,7	73,1
0,1	45,1	43,7	45,4	50,4
0	43,8	43,8	43,8	43,8
–0,1	43,6	43,7	44,8	X
–0,3	44,4	44,6	44,5	X
–0,5	51,0	50,2	45,5	X
–0,8	135,2	101,0	52,4	X

Oznaczmy przez σ odchylenie standardowe statystyki wykreślanej na karcie Shewharta. Zobaczmy teraz jak reaguje karta Shewharta na przesunięcie wartości oczekiwanej o wartość $k\sigma$. W Tablicy 2 na podstawie wyników opublikowanych w pracy Hryniewicza (2012) podano wartości ARL w przypadku, gdy $k = 1$, co odpowiada przypadkowi stosunkowo niewielkiego rozregulowania kontrolowanego procesu. Z kolei, w Tablicy 3 podano wartości ARL w przypadku, gdy $k = 3$, co odpowiada przypadkowi bardzo silnego rozregulowania kontrolowanego procesu.

Tablica 3

Wartości ARL dla karty X — średnie Shewharta — przesunięcie o 3 σ

τ Kendalla	Gauss	Clayton	Frank	Gumbel
0,8	10,6	24,9	9,8	9,1
0,5	3,2	4,1	4,5	3,2
0,3	2,5	2,6	3,6	2,4
0,1	2,1	2,1	3,1	2,1
0	2,0	2,0	2,0	2,0
–0,1	1,9	1,9	2,9	X
–0,3	1,8	1,8	2,8	X
–0,5	1,7	1,8	2,7	X
–0,8	1,6	1,7	2,6	X

Z danych przedstawionych w Tablicach 2 oraz 3 wynika, że występowanie dodatnich zależności pomiędzy kolejnymi obserwacjami procesu w ujemny sposób wpływa na zdolność karty *X*–średnie Shewharta do wykrywania rozregulowania procesu. Dla zależności o małej lub umiarkowanej sile pogorszenie to jest stosunkowo niewielkie (z wyjątkiem przypadku kopuli Franka dla przypadku $k = 3$). Jednakże w przypadku silnych dodatnich zależności, własności detekcyjne karty stają się bardzo słabe. Dotyczy to zwłaszcza przypadku kopul Gaussa i Gumbela dla $k = 1$ oraz kopuli Claytona dla $k = 3$. Podobne pogorszenie własności detekcyjnych karty obserwowane jest również w przypadku zależności ujemnych, ale wyłącznie dla słabych rozregulowań procesu. Dla silnych rozregulowań ($k = 3$) wartości detekcyjne ulegają poprawie (z wyjątkiem przypadku kopuli Franka).

Podsumowując, można stwierdzić, że zarówno siła i kierunek zależności (czego można było się spodziewać), jak i opisany kopulą rodzaj zależności, w sposób istotny wpływają na własności karty *X*–średnie Shewharta. Okazuje się, że stwierdzenie takiego samego rozkładu obserwowanych wartości procesu oraz takiej samej siły zależności pomiędzy kolejnymi obserwacjami procesu nie wystarcza, by przewidzieć zdolności detekcyjne karty kontrolnej. Należy dodatkowo uwzględnić rodzaj zależności, co do momentu opublikowania pracy nie było zauważone, zob. Hryniewicz (2012).

Wpływ wzajemnej zależności pomiędzy kolejnymi obserwacjami kontrolowanego procesu na własności kart służących do kontroli wartości średniej procesu był — jak wspomniano powyżej — przedmiotem badań od wielu lat. W praktyce SPC stosuje się także karty kontrolne służące do kontroli zmienności procesu, takie jak np. karta odchylenia standardowego (karta *s*) Shewharta lub karta rozstępu (karta *R*) Shewharta. Ich własności zostały dobrze zbadane w przypadku, gdy kolejne obserwacje opisane są niezależnymi zmiennymi losowymi o rozkładach normalnych. Odejście od tych założeń powoduje, że analiza własności tych kart kontrolnych napotyka na olbrzymie trudności. W przeciwieństwie do kart służących do kontroli wartości średnich zmiennych o rozkładzie normalnym, rozkłady prawdopodobieństwa mierzonych zmiennych losowych oraz rozkłady prawdopodobieństwa wyznaczanych statystyk mają różną postać. Jeśli, na dodatek, chcielibyśmy uwzględnić wzajemne zależności pomiędzy wykreślanymi na karcie statystykami, powstają nowe problemy utrudniające określenie własności kart kontrolnych nawet metodami symulacyjnymi. Z tego też powodu prace poświęcone wpływowi zależności w kontrolowanych procesach na własności kart kontrolnych służących do kontrolowania zmienności procesu praktycznie nie istnieją.

W celu zbadania własności przykładowej karty służącej do kontroli zmienności procesu, jaką jest karta rozstępu *R* załóżmy, że obserwowane wartości rozpatrywanej charakterystyki jakościowej opisane są rozkładem normalnym, a ponadto ich wzajemna zależność opisana jest jedną z wymienionych już kopul.

Przyjmujemy też dodatkowe założenie upraszczające, zgodnie z którym zależność pomiędzy ostatnią obserwacją z danej próbki losowej a pierwszą obserwacją kolejnej próbki losowej jest taka sama jak zależność pomiędzy kolejnymi obserwacjami w próbce (traktujemy więc kolejne próbki jako kolejne segmenty kontrolowanego procesu). Dla tak opisanego procesu przeanalizowaliśmy wpływ siły i rodzaju istniejących zależności na występowanie fałszywych alarmów na karcie kontrolnej *R*.

Tablica 4

Wartości ARL0 dla karty *R* Shewharta

τ Kendalla	Gauss	Clayton	Frank	Gumbel
0,8	1180000	>1500000	401824	>10000000
0,5	11670	9770	1724	92669
0,3	1056	1474	501	4316
0,1	344	332	260	984
0	218	218	218	218
–0,1	151	147	180	X
–0,3	76	77	137	X
–0,5	53	55	96	X
–0,8	105	71	51	X

Wartości przedstawione w Tablicy 4 zostały wyznaczone metodami symulacyjnymi przy stosunkowo małych liczbach wykonanych przebiegów procesu symulacji. Nie są to więc oszacowania dokładne, choć ich analiza pozwala wysnuć kilka ważnych wniosków o charakterze praktycznym. Po pierwsze, widać bardzo szybki wzrost wartości parametru ARL0 wraz ze wzrostem siły dodatniej zależności pomiędzy kolejnymi obserwacjami kontrolowanego procesu. ARL0 rośnie najszybciej, gdy zależność opisana jest kopulą Gumbela, a najwolniej, gdy zależność opisana jest kopulą Franka. Ponieważ wartość ARL0 jest miarą oczekiwanego czasu do wystąpienia „fałszywego alarmu", tak duże wartości mogłyby wydawać się korzystne. Należy jednak pamiętać, że są one konsekwencją zbyt szerokich (w stosunku do zmienności kontrolowanego procesu) przedziałów pomiędzy granicami kontrolnymi. W rezultacie znacznemu wzrostowi ulegną również oczekiwane czasy do wystąpienia „prawdziwych alarmów", co jest zjawiskiem zdecydowanie niekorzystnym. W przypadku występowania pomiędzy kolejnymi obserwacjami zależności ujemnej, sytuacja jest niejako odwrotna. Wraz ze wzrostem siły ujemnej zależności wartość ARL0 najpierw maleje, a później rośnie (dla kopuli Franka dla wartości τ mniejszych od -0,8). Oznacza to częste występowanie „fałszywych alarmów", co jest konsekwencją zbyt wąskich (w stosunku do zmienności kontrolowanego procesu) przedziałów pomiędzy granicami kontrolnymi. Nie jest to sytuacja korzystna. Z drugiej jednak strony oznacza to, że w przypadku rozregulowania kontrolowanego procesu, które ma charakter wzrostu jego zmienności, alarmy pojawią się szybciej niż w przypadku procesów o niezależnych obserwacjach.

5. PODSUMOWANIE

Przedstawione w niniejszej pracy wyniki badań pokazują w sposób całkowicie jednoznaczny, że występowanie zależności pomiędzy kolejnymi obserwacjami kontrolowanego procesu w sposób istotny zmienia charakterystyki podstawowych narzędzi statystycznego sterowania procesami, jakimi są karty kontrolne Shewharta. Wniosek ten nie jest nowy i znalazł potwierdzenie w wielu pracach poświęconych analizie procesów opisanych klasycznymi (normalnymi) procesami autoregresji. Nowością, zwłaszcza w przypadku zastosowań karty kontrolnej R, służącej do monitorowania zmienności procesu, jest zauważenie istotnego wpływu rodzaju zależności pomiędzy kolejnymi obserwacjami na charakterystyki kart kontrolnych. Okazuje się, że w przypadku takich samych (normalnych) rozkładów obserwowanych wartości i takiej samej, mierzonej za pomocą współczynnika τ Kendalla siły zależności pomiędzy kolejnymi obserwacjami procesu, statystyczne charakterystyki kart kontrolnych Shewharta mogą być zasadniczo różne, w zależności od typu kopuli opisującej zależność kolejnych obserwacji.

BIBLIOGRAFIA

Alwan L.C., Roberts H.V. (1988), *Time–series modeling for statistical process control*, Journal of Business & Economic Statistics, 6, 87–95.

Alwan, L.C., Roberts H.V. (1995), *The problem of misplaced control limits*, *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 44, 269–306.

Bagshaw M., Johnson R.A. (1975), *The Effect of Serial Correlation on the Performance of CUSUM Tests II*, Technometrics, 17, 73–80.

Clayton G. G. (1978), *A model for Association in Bivariate Life Tables and its Applications in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence*, Biometrika, 65, 141–151.

Ferguson T.S., Genest C., Hallin M. (2000), *Kendall's tau for Serial Dependence*, The Canadian Journal of Statistics, 28, 587–604.

Frank M.J. (1979), *On the Simultaneous Associativity of F(x,y) and x+y-F(x,y)*, Æquationes Mathematicæ, 19, 194–226.

Genest C., Rivest L.P. (1993), *Statistical Inference Procedures for Bivariate Archimedean Copulas*, Journal of the American Statistical Association, 88, 1034–1043.

Gumbel E.J. (1960), *Distributions des valeurs extremês en plusieurs dimensions*, Publications de l'Institut statstique de l'Universitède Paris, 9, 171–173.

Hryniewicz O., Karpiński J. (2012), *Prediction of reliability — pitfalls of using Pearson's correlation*, Raport IBS PAN, RB/19/2012.

Hryniewicz O., Szediw A. (2010), *Sequential Signals on a Control Chart Based on Nonparametric Statistical Tests*, w: *Frontiers in Statistical Quality Control — 9* (red. H.-J. Lenz, P.-Th. Wilrich, W. Schmid), Physica–Verlag, Heidelberg, 99–117.

Hryniewicz O. (2012), *On the Robustness of the Shewhart Control Chart to Different Types of Dependencies in Data*, w: *Frontiers in Statistical Quality Control — 10* (red. H.-J. Lenz, W. Schmid, P.-Th. Wilrich), Physica–Verlag, Heidelberg, 19–33.

Iwasiewicz A. (1985), *Statystyczna kontrola jakości w toku produkcji: systemy i procedury*, PWN, Warszawa.

Nelsen R.B. (2006), *Introduction to Copulas*, Springer, New York.

Olwert A. (2012), *Karta Kendalla jako narzędzie do wykrywania zależności pomiędzy kolejnymi pomiarami procesów*, Rozpr. dokt. IBS PAN, Warszawa.

Sklar A. (1959), *Fonctions de répartition á n dimensions et leurs marges*, Publications de l'Institut de Statistique de l'Universite de Paris, 8, 229–231.

Vasilopoulos A.V., Stamboulis A.P. (1978), *Modification of Control Limits in the Presence of Correlation*, Journal of Quality Technology, 10, 20–30.

Wardell D.G., Moskowitz H., Plante R.D. (1994), *Run–Length Distributions of Special–Cause Control Charts for Correlated Processes*, Technometrics, 36, 3–17.

Yourstone S.A., Montgomery D.C. (1989), *A Time–Series Approach to Discrete Real–Time Process Quality Control*, Quality and Reliability Engineering International, 5, 309–317.