

**Raport Badawczy**

**RB/76/2011**

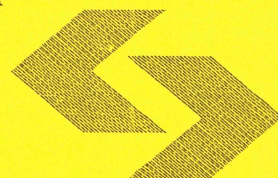
**Research Report**

**Karta Kendalla  
jako nieparametryczne  
narzędzie do wykrywania  
zależności pomiędzy kolejnymi  
pomiarami procesów**

**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 Zakładu zgłaszający pracę:  
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2011

# 1 Wprowadzenie

Celem statystycznego sterowania procesami (SPC) jest wykrywanie ich rozregulowań, co w konsekwencji powinno prowadzić do znajdowania i usuwania przyczyn rozregulowania, a więc do poprawy jakości procesu. Od kilkudziesięciu już lat stosowane są do tego celu proste metody statystyczne, a w tym niezwykle popularne karty kontrolne Shewharta oraz inne karty kontrolne jak np. CUSUM, MAV i EWMA. Te klasyczne i szeroko stosowane karty kontrolne konstruowane są przy założeniu, że obserwacje procesu są niezależne o jednakowym rozkładzie. W większości przypadków w praktyce założenia te są spełnione przynajmniej asymptotycznie. Jednakże istnieją procesy, dla których występowanie korelacji pomiędzy kolejnymi pomiarami (obserwacjami) jest charakterystyczne. Do takich procesów należą procesy chemiczne jak i różne ciągłe procesy produkcyjne (Wardell, Moskowitz, Plante, 1994), (Alwan i Roberts, 1995).

Występowanie zależności pomiędzy kolejnymi pomiarami procesu wpływa z reguły negatywnie na działanie klasycznych narzędzi statystycznych. Stąd też zjawisko występowania zależności pomiędzy kolejnymi pomiarami procesu warto wziąć pod uwagę przy konstrukcji narzędzi statystycznej kontroli jakości. Pierwsze prace na ten temat powstały w latach siedemdziesiątych (Johnson i Bagshaw, 1974), (Vasilopoulos i Stamboulis, 1978). Natomiast najwięcej prac pojawiło się pod koniec lat osiemdziesiątych i dziewięćdziesiątych.

Opracowanych zostało wiele konkretnych podejść do problemu występowania korelacji między pomiarami procesów. Najstarsze z nich to karty kontrolne z odpowiednio zmodyfikowanymi granicami (Lu, Reynolds, 1999a, 1999b, 2001), (Schmid, 1995, 1997), (VanBrackle III, Reynolds, 1997), (Vasilopoulos, Stamboulis, 1978), (Zhang, 1998). Drugie podejście dotyczy kart kontrolnych dla residuów. Zaproponowali je Alwan i Roberts (1988). Wśród prac poświęconych temu zagadnieniu warto wymienić następujące: (Lu, Reynolds, 1999a, 1999b, 2001), (Montgomery, Mastrangelo, 1991), (Schmid, 1997). Lu i Reynolds porównywali karty kontrolne ze zmodyfikowanymi odpowiednio granicami dla oryginalnych danych z kartami dla residuów. Pokazali, że najlepsze wyniki otrzymywane są w przypadku stosowania jednocześnie kart EWMA lub CUSUM dla oryginalnych danych oraz karty Shewharta dla residuów. Inne, trzecie podejście, wprowadzone zostało przez Yourstone i Montgomery (1989, 1991). Opiera się ono na monitorowaniu procesu na podstawie współczynnika autokorelacji, odzwierciedlając strukturę korelacji w monitorowanym procesie. Z kolei jeszcze inne, oryginalne, jedne z głównych podejść do rozważanego problemu to tzw. karta kontrolna ARMA (Jiang, Tsui, Woodall, 2000).

Procedury SPC dla skorelowanych danych są zazwyczaj proponowane w przypadku monitorowania pojedynczych pomiarów, których kontrolowanie jest charakterystyczne dla ciągłych procesów produkcji. Znacznie mniej prac dotyczy przypadku, gdy autokorelacja występuje wewnątrz próbki. Przegląd prac na ten temat można znaleźć w artykule Knoth, Schmid, Schone (2001).

Rozważając różne metody SPC dla skorelowanych danych należy wziąć pod uwagę nie tylko ich efektywność, ale i prostotę (Vasilopoulos, Stamboulis, 1978). Wyniki otrzymane w pracy Wardell, Moskowitz, Plante (1994) wskazują, iż działanie zaproponowanej przez Alwana i Roberta karty kontrolnej residuów w niektórych przypadkach dodatniej korelacji może nie być tak efektywne jak działanie prostej karty kontrolnej Shewharta. Jednakże w pozostałych przypadkach zastosowanie tych raczej skomplikowanych procedur (ich użycie wymaga ponadto odpowiedniego komputerowego oprogramowania) wydaje się jak najbardziej użyteczne. Innym przykładem są wyniki otrzymane przez Timmera (1998). W swojej pracy podaje on, iż w przypadku prostych modeli autoregresyjnych AR(1) zastosowanie wyłącznie prostych kart kontrolnych opartych na współczynniku autokorelacji jako narzędzi SPC może nie być efektywne.

Przy okazji omawiania narzędzi SPC dla skorelowanych danych warto wspomnieć, iż rozpatrując skorelowane dane nie możemy polegać, nawet w przypadku prostych kart kontrolnych, na metodach estymacji odpowiednich parametrów opartych na założeniu o niezależności obserwacji. Konieczne są tu pewne ich modyfikacje (Maragah, Woodall, 1992). Ponadto stosując procedury SPC dla skorelowanych

danych dobrze jest znać strukturę korelacji, bowiem większość owych procedur wymaga ich znajomości. Często dodatkowo wymagana jest też znajomość parametrów procesu. Lu i Reynolds (1999, 2001) podali, iż precyzyjnej estymacji takich parametrów można dokonać dysponując co najmniej paruset obserwacjami. Biorąc pod uwagę, że wszystkie obliczenia potrzebne do budowy i uruchomienia procedur SPC dla skorelowanych danych nie są prostym zabiegiem dla przeciętnej praktyka, powstaje pytanie: jak w prostszy sposób radzić sobie z problemem występowania zależności pomiędzy pomiarami procesu, a dokładniej jak weryfikować występowanie zależności wśród pomiarów procesu w prosty, uniwersalny sposób?

Oczywiste jest, iż praktycy woleliby unikać używania skomplikowanych procedur SPC dla skorelowanych danych w przypadkach, w których ich stosowanie nie wydaje się niezbędne. Istotna wydaje się zatem umiejętność rozstrzygnięcia, czy w danej sytuacji konieczne jest stosowanie zaawansowanych metod statystycznej kontroli jakości dla skorelowanych danych. W takim przypadku ważne jest wykrywanie zależności pomiędzy obserwacjami w prosty i uniwersalny sposób.

Najprostszym rozwiązaniem owego problemu jest karta kontrolna dla współczynnika autokorelacji (Yourstone i Montgomery, 1989, 1991). Jednakże działa ona efektywnie przede wszystkim dla procesów gaussowskich. W przeciwnym przypadku zaleca się stosowanie nieparametrycznych testów statystycznych. Ciekawy przegląd takich testów przedstawili Hallin i Melard (1988). Niestety większość przedstawionych tam testów jest skomplikowana i nie znajduje zastosowania w statystycznej kontroli jakości. Przedstawiona jednakże niedawno w pracy Fergusona, Genesta, Hallina (2000) statystyka  $\tau$  Kendalla wykorzystywana do testowania zależności pomiędzy kolejnymi obserwacjami w próbie wydaje się ciekawym narzędziem do zastosowania w dziedzinie SPC w aspekcie omawianego problemu. W oparciu o tę statystykę poszukiwana była prosta metoda pozwalająca wykrywać zależności pomiędzy kolejnymi obserwacjami.

**W pracy zaproponowano tzw. kartę Kendalla jako proste i uniwersalne nieparametryczne narzędzie statystyczne służące do wykrywania zależności pomiędzy kolejnymi pomiarami procesu.**

Podstawowe wiadomości dotyczące wykorzystywanej statystyki  $\tau$  Kendalla, jej własności oraz konstrukcja proponowanej karty kontrolnej przedstawione zostały w punkcie drugim autoreferatu. Statystyczne własności karty Kendalla były badane na podstawie symulacji metodami Monte Carlo. Opis modeli wykorzystywanych do modelowania zależności pomiędzy kolejnymi pomiarami procesów zawarty jest w punkcie trzecim. Z kolei wyniki owych badań symulacyjnych przedstawia punkt piąty. W punkcie czwartym, korzystając z definicji odporności procedur statystycznych zaproponowanej przez Zielińskiego (1978), przedstawiono wstępne wyniki dotyczące odporności karty Kendalla wobec pewnych odstępstw od założeń przyjmowanych przy konstrukcji klasycznych narzędzi SPC. W punkcie szóstym opisana została postać karty Kendalla dla danych rozmytych. W zakończeniu przedstawiono podsumowujące pracę wnioski, a także wskazano kierunki dalszych badań.

## 2 Karta Kendalla

Powszechnie znany w literaturze statystycznej współczynnik Kendalla jako rangowa miara zależności dla próby dwuwymiarowej  $(X, Y)$  wprowadzony został w 1938 roku przez Kendalla. Do niedawna jego rozkład znany był wyłącznie w przypadku niezależności. Tymczasem w 2000 roku Ferguson, Genest i Hallin zbadali własności współczynnika Kendalla w przypadku, gdy kolejne obserwacje są skorelowane. Ich wyniki zostały wykorzystane w dalszej części pracy.

Niech  $Z_1, Z_2, \dots, Z_n$  będzie próbą losową złożoną z  $n$  kolejnych obserwacji kontrolowanego procesu oraz  $(X_i, Y_i)$ , gdzie  $X_i = Z_i, Y_i = Z_{i+1}$  dla  $i = 1, \dots, n-1$  będzie dwuwymiarowym wektorem losowym. Statystyka  $\tau$  Kendalla opisująca zależność zmiennych losowych  $X$  i  $Y$  ma postać:

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

gdzie

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

Równoważnie współczynnik Kendalla można zapisać używając terminologii par niezgodnych, czyli takich par  $(Z_i, Z_{i+1}), (Z_j, Z_{j+1})$ , dla których  $Z_i < Z_j$  i  $Z_{i+1} > Z_{j+1}$  bądź  $Z_i > Z_j$  i  $Z_{i+1} < Z_{j+1}$ . Wówczas

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

gdzie

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

zaś  $I(A)$  jest funkcją indykatorową na zbiorze  $A$ . W przypadku wzajemnie niezależnych par obserwacji  $(X_i, Y_i), i = 1, 2, \dots, n-1$  rozkład prawdopodobieństwa statystyki (1) jest znany. Jednakże w rozpatrywanym przez nas przypadku założenie to nie jest spełnione, nawet wtedy, gdy oryginalne obserwacje  $Z_1, Z_2, \dots, Z_n$  są wzajemnie niezależne. Dla takiego przypadku Ferguson, Genest i Hallin (2000) wyznaczyli wartości prawdopodobieństwa  $P_n(M \leq m)$ . Dla małych prób, tj. o liczności  $n = 3, \dots, 10$  obliczyli dokładne wartości tego prawdopodobieństwa, zaś dla większych prób ( $n > 10$ ) podali jego przybliżone wartości. W pracy wyznaczono prawdopodobieństwa  $P(\tau_n \geq \tau_{kryt})$ , gdzie  $\tau_{kryt}$  jest wartością krytyczną statystyki  $\tau$  Kendalla. W Tabeli 1 dla wybranych wartości  $n$  przedstawiono prawdopodobieństwa zaobserwowania wartości  $\tau_n \geq \tau_{kryt}$ , gdy kolejne obserwacje  $Z_1, Z_2, \dots, Z_n$  są wzajemnie niezależne.

Analizując wyniki zamieszczone w Tabeli 1, łatwo zauważyć, że poza przypadkiem, dla którego  $n = 6$  i  $\tau_{kryt} = 1$  nie jest możliwe skonstruowanie testu niezależności wobec dodatniej zależności z takim samym prawdopodobieństwem fałszywego alarmu jak w przypadku prostej karty kontrolnej Shewharta. Dzieje się tak również dla dwustronnych testów niezależności, jak i prób o większej liczności  $n$  (Ferguson, Genest, Hallin, 2000). Fakt ten wynika z dyskretnego charakteru statystyki  $\tau$  Kendalla.

Ważną z praktycznego punktu widzenia charakterystyką jest oczekiwana liczba pomiarów (ARL) do pojawienia się takiej sekwencji obserwacji, dla których spełniona będzie nierówność  $\tau_n \geq \tau_{kryt}$ . W przypadku niezależności kolejnych obserwacji pojawienie się takiej sekwencji będzie fałszywym alarmem sugerującym rozregulowanie procesu. Niestety, wyznaczenie wartości ARL w sposób analityczny jest niezwykle trudne ze względu na wzajemną zależność kolejnych wartości statystyki  $\tau_n$ . W rezultacie wartości tej charakterystyki dają się wyznaczyć wyłącznie metodami symulacyjnymi. Tabela 2 zawiera

Tablica 1: Wartości krytyczne statystyki  $\tau$  Kendalla

n=6		n=7		n=8		n=9	
$\tau_{kryt}$	$P(\tau \geq \tau_{kryt})$	$\tau_{kryt}$	$P(\tau \geq \tau_{kryt})$	$\tau_{kryt}$	$P(\tau \geq \tau_{kryt})$	$\tau_{kryt}$	$P(\tau \geq \tau_{kryt})$
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

Tablica 2: Wartości ARL dla niezależnych obserwacji

n=6		n=7		n=8		n=9	
$\tau_{kryt}$	ARL	$\tau_{kryt}$	ARL	$\tau_{kryt}$	ARL	$\tau_{kryt}$	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

wartości ARL wyznaczone w podany sposób dla obserwacji wzajemnie niezależnych z rozkładu normalnego.

Otrzymane wyniki potwierdzają wniosek, iż niemożliwe jest skonstruowanie testu statystycznego o własnościach bliskich karcie kontrolnej Shewharta. Wartości ARL są zdecydowanie większe lub wyraźnie mniejsze od oczekiwanych, odpowiadających prawdopodobieństwu fałszywego alarmu 0,0027. Duże wartości ARL są oczywiście korzystniejsze, ale testy o takich własnościach mają jednocześnie duże wartości oczekiwanego czasu do wykrycia prawdziwych alarmów, co jest własnością zdecydowanie niekorzystną. Z kolei bardzo małe wartości ARL świadczą o wysokim prawdopodobieństwie fałszywych alarmów.

Mając na uwadze prostotę i uniwersalność poszukiwanego narzędzia do wykrywania zależności pomiędzy kolejnymi pomiarami procesów, proponujemy skonstruowanie na podstawie statystyki (1) następującej karty kontrolnej

$$UCL = \min(E(\tau_n) + k\sigma(\tau_n), 1) \quad (5)$$

$$LCL = \max(E(\tau_n) - k\sigma(\tau_n), -1), \quad (6)$$

gdzie  $UCL$  i  $LCL$  są granicami kontrolnymi, odpowiednio, górną i dolną. Do wyznaczenia granic kontrolnych wykorzystujemy wzory na wartość oczekiwaną  $E(\tau_n)$  i wariancję  $\sigma^2(\tau_n)$  podane przez Fergusona, Genesta i Hallina (2000), a mianowicie

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

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

Zaproponowana powyżej karta kontrolna w dalszej części pracy nazywana jest **kartą Kendalla**.

Warto zauważyć, że dla małych wielkości  $n$  rozkład  $\tau_n$  nie jest symetryczny. W związku z tym własności karty Kendalla mogłyby okazać się lepsze w przypadku zastosowania granic kontrolnych niesymetrycznych wokół wartości oczekiwanej  $E(\tau_n)$ . Niemniej ze względu na prostotę rozwiązania możliwość ta nie jest w pracy rozważana. Własności zaproponowanej karty Kendalla przedstawione są w piątym punkcie autoreferatu.

### 3 Modelowanie zależności

W dalszym ciągu przyjmuje się, że  $X_i = Z_i$ ,  $Y_i = Z_{i+1}$ ,  $i = 1, \dots, n-1$  są zmiennymi losowymi opisującymi dwie kolejne obserwacje procesu w próbie  $Z_1, \dots, Z_n$ . W celu zbadania własności zaproponowanej karty Kendalla konieczne jest zmodelowanie zależności stochastycznej między zmiennymi  $X$  i  $Y$ .

Metody badania zależności są powszechnie znane. Zwykle zależność mierzy się za pomocą różnych współczynników, np. współczynnika Spearmana, Kendalla. Istnieje też wiele dwuwymiarowych modeli opisujących zależności występujące w próbie losowej (Mardia, 1970).

W niniejszej pracy karta Kendalla badana jest przede wszystkim na podstawie porównania ze znaną kartą autokorelacji. Oczywiście jest, iż karta autokorelacji powinna działać znacznie efektywniej niż karta Kendalla w przypadku modelu autoregresyjnego. Jednakże można przypuszczać, że w sytuacji, gdy zależność między  $X$  a  $Y$  jest inna niż w procesie autoregresji, to proponowana karta może działać znacznie lepiej niż karta autokorelacji. Aby się o tym przekonać, należy w odpowiedni sposób opisać ową zależność. Pomocne w tym celu okazują się tzw. kopuły.

Podstawowe twierdzenie w teorii kopuł - twierdzenie Sklara (1959) przyczyniło się do wzrostu zastosowań kopuł zwłaszcza w dziedzinie modelowania wielowymiarowych rozkładów oraz ich symulacji. Kopuły stały się popularne w analizie szeregów czasowych, szczególnie w zagadnieniach finansowych (Embrechts, Lindskog, McNeil, 2003).

W rozważanym problemie modelowania zależności stochastycznej różnej od zależności właściwej dla modeli autoregresyjnych wygodne jest zastosowanie kopuł. W związku z tym do opisu zależności pomiędzy kolejnymi pomiarami procesu wykorzystywane są w pracy, poza modelem autoregresyjnym, kopuły Farliego-Gumbela-Morgensterna (FGM), Placketta i Franka. Szczegółowe informacje na temat tych kopuł można znaleźć w pracy Nelsena (2006). W autoreferacie zostaną one tylko pokrótce scharakteryzowane.

Ogólnie kopułę można określić jako dystrybuantę rozkładu prawdopodobieństwa na kostce jednostkowej o równomiernych rozkładach brzegowych. W przypadku dwuwymiarowym kopułę można zapisać jako funkcję:  $C : [0, 1]^2 \rightarrow [0, 1]$ . Zgodnie z twierdzeniem Sklara dystrybuantę łączną  $F_{12}$  wektora  $(X, Y)$ , gdzie  $X, Y$  są ciągłymi zmiennymi losowymi można przedstawić jednoznacznie jako  $F_{12}(x, y) = C(F_1(x), F_2(y))$ , gdzie  $F_1(x), F_2(y)$  są dystrybuantami brzegowymi, odpowiednio, zmiennych  $X$  i  $Y$ . Używając przekształcenia  $u_1 = F_1(x)$ ,  $u_2 = F_2(y)$ , gdzie  $u_1, u_2$  pochodzą z rozkładu równomiernego na przedziale  $[0, 1]$ , kopuły FGM ( $C_{FGM}$ ), Placketta ( $C_P$ ) i Franka ( $C_F$ ) definiowane są, odpowiednio, następującymi wzorami

$$C_{FGM}(u_1, u_2) = u_1 u_2 \{1 + \alpha(1 - u_1)(1 - u_2)\}, \quad \alpha \in [-1, 1], \quad (9)$$

$$C_P(u_1, u_2) = \begin{cases} u_1 u_2, & \alpha = 1, \\ \frac{[1 + (\alpha - 1)(u_1 + u_2)] - \sqrt{[1 + (\alpha - 1)(u_1 + u_2)]^2 - 4u_1 u_2 \alpha (\alpha - 1)}}{2(\alpha - 1)}, & \alpha \in \mathbb{R}_+ \setminus \{1\}, \end{cases} \quad (10)$$

$$C_F(u_1, u_2) = -\frac{1}{\alpha} \ln \left( 1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1} \right), \quad \alpha \in \mathbb{R} \setminus 0. \quad (11)$$

Parametr  $\alpha$  występujący w powyższych wzorach charakteryzuje związek między zmiennymi  $X$  i  $Y$ . Zmienne są niezależne wtedy i tylko wtedy, gdy  $\alpha = 0$ ,  $\alpha = 1$ , bądź  $\alpha \cong 0$ , odpowiednio, dla ko-

puł  $C_{FGM}$ ,  $C_P$  i  $C_F$ . Dla parametrów  $\alpha$  większych bądź mniejszych od tych wartości, siła zależności wzrasta w kierunku, odpowiednio, dodatnim lub ujemnym.

Jak wcześniej wspomniano, do opisu zależności pomiędzy zmiennymi  $X$  i  $Y$  użyty został również dwuwymiarowy rozkład normalny o łącznej gęstości wektora losowego  $(X, Y)$  postaci

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp(-Q), \quad (12)$$

gdzie

$$Q = \frac{1}{1-\rho^2} \left\{ \frac{(x-m_X)^2}{\sigma_X^2} - 2\rho \frac{(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2} \right\}, \quad (13)$$

zaś paramter  $\rho$  jest współczynnikiem korelacji. Jeśli zmienne  $X$  i  $Y$  są niezależne, wówczas  $\rho = 0$ . Rozkład warunkowy zmiennej  $Y$  dla danego  $X = x$  jest rozkładem normalnym ze średnią  $m_Y + \rho(\sigma_X/\sigma_Y)(x-m_X)$  i wariancją  $\sigma_Y\sqrt{1-\rho^2}$ . W szczególności, gdy  $m_X = m_Y = 0$  oraz  $\sigma_X = \sigma_Y = 1$  jest to rozkład normalny ze średnią  $x\rho$  i wariancją  $\sqrt{1-\rho^2}$ . Proponowany model jest zatem dobrze znanym modelem autoregresji.

Wykorzystując wymienione modele zależności, zbadano symulacyjnie własności karty Kendalla. Wyniki tych badań zostały zawarte w punkcie 5.

## 4 Odporność karty Kendalla

Pojęcie odporności stosowane jest w odniesieniu do różnych procedur statystycznych. W literaturze istnieje wiele interpretacji podanej pierwotnie w sposób opisowy definicji odporności (Box, Andersen, 1955). Wiele prac badawczych na temat odporności poświęconych jest odpornej estymacji parametrów modelu. Są wśród nich prace Hubera (1964) i Hampla (1974), głównych twórców teorii odpornego wnioskowania statystycznego. Choć przedstawione przez nich ujęcia odporności są różne, oba miały na celu sformalizowanie idei odporności. Odminną, w stosunku do Hubera i Hampla, koncepcję odporności podała Zieliński (1978). Określił on odporność jako funkcję rzeczywistą na przestrzeni parametrów danego problemu statystycznego mierzącą wrażliwość zmian wartości badanej statystyki lub związanej z nią wielkości. Koncepcja Zielińskiego poprzez swą uniwersalność pozwala na badanie odporności różnych procedur statystycznych wobec dowolnego rodzaju zaburzenia. Między innymi z tego względu wykorzystywana jest w niniejszej pracy do badania odporności karty Kendalla.

Badając odporność według kryteriów Zielińskiego należy określić model standardowy i rozszerzony opisujący odchylenia od przyjętego modelu standardowego. Oczywiście model standardowy jest szczególnym przypadkiem modelu rozszerzonego. Używając aparatu matematycznego oba modele można zapisać w następujący sposób.

Rozważmy model statystyczny  $M_0 = (\mathcal{X}, \mathcal{A}, \mathcal{P}_0)$ , gdzie  $\mathcal{X}$  jest przestrzenią prób,  $\mathcal{A}$  –  $\sigma$ -algebrą zdarzeń losowych w rozważanym problemie statystycznym, a  $\mathcal{P}_0$  – pewną podrodziną rodziny  $\mathcal{P}$  wszystkich rozkładów prawdopodobieństwa na  $\mathcal{A}$ . Niech  $2^{\mathcal{P}}$  oznacza rodzinę wszystkich podzbiorów rodziny  $\mathcal{P}$ , zaś  $\pi$  będzie następującym odwzorowaniem

$$\pi: \mathcal{P}_0 \rightarrow 2^{\mathcal{P}}.$$

Zakłada się, że  $P \in \pi(P)$ . Niech  $\mathcal{P}_1 = \bigcup_{P \in \mathcal{P}_0} \pi(P)$ . Wtedy model  $M_1 = (\mathcal{X}, \mathcal{A}, \mathcal{P}_1)$  nosi nazwę modelu rozszerzonego (rozszerzenia modelu  $M_0$ ).

Odwzorowanie  $\pi$  interpretowane jest jako odzwierciedlenie odchyłeń od przyjętego modelu (tzw. modelu standardowego). Zamiast danego rozkładu  $P$  próby  $X_1, \dots, X_n$ , mamy rozkład  $\pi(P)$  określający model rozszerzony, tzn. model włączający odstępstwa od założeń przyjętych w modelu standardowym. Zakładając, że  $\rho$  oznacza pewną własność statystyki  $T = T(X_1, \dots, X_n)$ , odporność statystyki  $T$  definiowana jest w następujący sposób

$$r_\rho(T) = \sup \{ \rho(T) : P \in \pi(P) \} - \inf \{ \rho(T) : P \in \pi(P) \}. \quad (14)$$



Funkcja  $r_\rho(T)$  określa miarę stabilności (wrażliwości) własności  $\rho$  statystyki  $T$  wobec odchyień od modelu standardowego. Jest funkcją rzeczywistą, co pozwala na porównywanie różnych procedur statystycznych pod względem ich odporności. Statystyka  $T$  jest więc tym bardziej odporna, im bardziej stabilna jest własność  $\rho$ , a więc im bardziej jest bliska zeru funkcja  $r_\rho(T)$ .

W badaniu odporności karty Kendalla jako model standardowy przyjęto model, w którym kolejne obserwacje procesu  $Z_1, \dots, Z_n$  są niezależne o jednakowym rozkładzie normalnym. Odporność karty można rozważać wobec dwojakiego rodzaju zaburzeń. W pierwszej kolejności jako odstępstwo od modelu standardowego przyjęto występowanie zależności między kolejnymi pomiarami procesu. W takim przypadku modelem rozszerzonym jest model, w którym kolejne obserwacje procesu są skorelowane, a do opisu zależności, podobnie jak proponowano wcześniej, używane są kopuły. Natomiast jako drugi oddzielnie rozważany typ zaburzenia przyjęte zostało odstępstwo od normalności rozkładu. W literaturze (Borror, Montgomery, Runger, 1999), (Chan, Hapuarachchi, Macpherson, 1988) najczęściej do opisu zaburzenia normalności wykorzystuje się rodzinę rozkładów Gamma, rozkładów t-Studenta, bądź rodzinę Tukeya. W niniejszej pracy jako drugi model rozszerzony potraktowano rodzinę rozkładów Gamma z parametrem skali równym 1.

Oprócz określenia wobec czego, należy jeszcze rozważyć ze względu na co badana jest odporność karty. Najczęściej w przypadku badania odporności kart kontrolnych bada się prawdopodobieństwo fałszywego alarmu bądź ARL (Borror, Montgomery, Runger, 1999), (Chan, Hapuarachchi, Macpherson, 1988). W niniejszej pracy rozpatrywana jest odporność proponowanej karty Kendalla ze względu na ARL wobec dwojakiego rodzaju zaburzeń. Wyniki badań przedstawione są w następnym punkcie.

## 5 Wyniki badań symulacyjnych

W celu zbadania efektywności proponowanej metody wykrywania zależności pomiędzy kolejnymi obserwacjami kontrolowanego procesu przeprowadzono eksperymenty symulacyjne. Jako miarę efektywności karty przyjęto ARL, czyli oczekiwaną liczbę pomiarów do momentu wystąpienia sygnału alarmowego, tzn. oczekiwaną liczbę pomiarów (ARL) potrzebnych do wykrycia rozregulowania, gdy ono istnieje lub oczekiwaną liczbę pomiarów do wystąpienia fałszywego alarmu, gdy kolejne obserwacje są wzajemnie niezależne.

Badania symulacyjne, niezależnie od przyjętego modelu, przeprowadzono w następujący sposób. Dla ustalonej wielkości próby  $n$  i przy założeniu, że kolejne dwie obserwacje procesu  $Z_i, Z_{i+1}$ ,  $i = 1, \dots, n - 1$  w tej próbie są zależne, wygenerowano próbę  $z_1, \dots, z_n$  i obliczono dla niej wartość statystyki  $\tau$  Kendalla. Jeśli wartość ta nie przekraczała granic kontrolnych karty Kendalla, to generowano kolejną obserwację  $z_{n+1}$  i wyznaczano wartość statystyki  $\tau$  Kendalla na podstawie próby  $z'_1 = z_2, \dots, z'_n = z_{n+1}$ . Krok ten był powtarzany do momentu, gdy wartość statystyki  $\tau$  Kendalla znalazła się poza granicami kontrolnymi karty. Liczba wygenerowanych obserwacji  $z_i$  włącznie z obserwacją, dla której przekroczona została linia kontrolna jest charakterystyką ARL.

Opisany proces symulacji był powtarzany  $10^5$  razy dla prób o licznosci 10, 20, 50 oraz  $10^6$  razy w przypadku prób o mniejszej licznosci. Wartości ARL dla poszczególnych rozmiarów prób zostały uśrednione.

Z uwagi na porównywanie zachowania się karty Kendalla z działaniem prostej karty autokorelacji, analogiczne badania symulacyjne wykonano również dla tej ostatniej.

Karta autokorelacji została tu zdefiniowana jako  $E(\rho_n) \pm k\sigma(\rho_n)$ , gdzie  $E(\rho_n)$  i  $\sigma(\rho_n)$  oznaczają wartość oczekiwaną i odchylenie standardowe współczynnika autokorelacji  $\rho_n$ . Na potrzeby konstrukcji karty przyjęto, iż wartość oczekiwana i wariancja współczynnika autokorelacji przyjmują swoje asymptotyczne wartości, tj.  $E(\rho_n) = 0$ , a  $Var(\rho_n) = \frac{n-1}{n(n+2)}$  (Moran, 1947).

Badania karty Kendalla i równoległe karty autokorelacji zostały wykonane według następującego scenariusza:

- przy założeniu, że zależność dwóch kolejnych obserwacji procesu opisywana jest przez dwuwymiarowy rozkład normalny
  - wyznaczono ARL dla karty Kendalla z parametrem  $k = 3$ ,
  - wyznaczono ARL dla karty autokorelacji z parametrem  $k = 3$ ,
  - porównano obie karty dla prób o licznosciach  $n = 10$  i  $n = 50$ ,
- zakładając, że zależność dwóch kolejnych obserwacji procesu opisywana jest za pomocą kopuły FGM, Placketta i Franka
  - obliczono ARL dla karty Kendalla z parametrem  $k = 3$ ,
  - obliczono ARL dla karty autokorelacji z parametrem  $k = 3$ ,
  - porównano obie karty dla prób o licznosciach  $n = 10$  i  $n = 50$  w przypadku różnych rozkładów brzegowych.

Ponadto zbadano odporność karty Kendalla i porównano ją z odpornością karty autokorelacji. W dalszej części tego punktu zostały przedstawione w postaci tabel bądź wykresów wyniki badań symulacyjnych.

Tabele 3 i 4 prezentują otrzymane wartości ARL dla obu kart z parametrem  $k=3$  w przypadku modelu autoregresji opisującego zależność między kolejnymi obserwacjami. Na podstawie wyników z Tabeli 3 nie trudno zauważyć, iż karta Kendalla z prostą do zapamiętania trzysigmową regułą nie jest dobrym narzędziem do wykrywania zależności między kolejnymi pomiarami procesu. Oczywiście wartość parametru  $k=3$  nie powinna być przyjmowana dla każdego rozmiaru próby  $n$ . Dopiero dla prób o licznosci co najmniej 10 otrzymywane wartości ARL układają się symetrycznie. Jednakże nawet dla większych prób wyniki nie są zadowalające.

Z kolei wyniki w Tabeli 4 zdają się być zaskakujące. Biorąc pod uwagę rodzaj przyjętego modelu można było się spodziewać, iż zachowanie karty autokorelacji będzie znacznie lepsze niż karty Kendalla. Dla małych prób obciążenie współczynnika autokorelacji oraz zastosowane przybliżenie Morana może mieć oczywiście negatywny wpływ na wyniki. Ogólnie jednak nie obserwuje się tu znaczącej przewagi działania karty autokorelacji.

W celu porównania obu kart należy przyjąć takie wartości parametru  $k$ , by w przypadku niezależności wszystkie wartości ARL były sobie bliskie. Porównania obu kart dokonano dla prób o licznosci  $n = 10$  i  $n = 50$ , przyjmując w pierwszym przypadku  $k = 2, 7$  oraz  $k = 2, 65$ , zaś w drugim  $k = 2, 2$  oraz  $k = 2, 16$ , odpowiednio, dla kart Kendalla i autokorelacji. Wyniki przedstawia Tabela 5 i Rysunek 1.

Wyniki w Tabeli 5 potwierdzają wcześniejsze wnioski na temat zachowania się obu kart. Dla prób 10-elementowych trudno jest określić, która z kart działa lepiej. W przypadku ujemnej zależności karta Kendalla wykrywa występowanie zależności szybciej niż karta autokorelacji, lecz w przypadku zależności dodatniej sytuacja jest odwrotna. Dla prób 50-elementowych działanie obu kart jest właściwie tak samo efektywne.

W kolejnym kroku wykonano analogiczne badania, tym razem rozpatrując kopuły jako model do opisu zależności między kolejnymi obserwacjami procesu. Jako rozkłady brzegowe zmiennych  $X$  i  $Y$  przyjęto rozkład normalny. Wyniki przedstawione są w Tabelach 6 i 7 (dla kopuły FGM) oraz na Rysunkach 2–4 (dla kopuły FGM, Placketta, Franka).

Porównując wyniki w Tabelach 3 i 6 oraz 4 i 7, można spostrzec iż w przypadku karty Kendalla wyniki otrzymane dla obu modeli nie różnią się istotnie. Niewielkie różnice wynikają raczej z zastosowania dwóch różnych generatorów. Fakt ten wynika z nieparametrycznego charakteru statystyki  $\tau$  Kendalla.

Tablica 3: Wartości ARL dla karty Kendalla z parametrem  $k=3$  dla modelu autoregresji

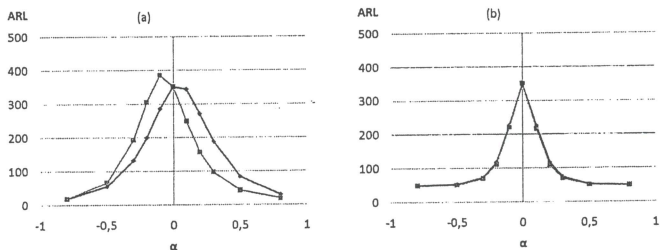
$\rho$	$n$						
	6	7	8	9	10	20	50
0.8	47.36	63.71	67.81	50.56	44.96	26.67	50.11
0.5	90.94	165.44	221.86	165.35	154.87	66.41	59.94
0.2	140.82	416.60	702.13	610.14	689.44	473.52	328.82
0.1	146.23	520.36	1156.77	859.95	1593.03	1026.06	1058.76
0	141.90	595.74	1623.15	1043.36	1497.27	1502.60	2597.57
-0.1	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

Tablica 4: Wartości ARL dla karty autokorelacji z parametrem  $k=3$  dla modelu autoregresji

$\rho$	$n$						
	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

Tablica 5: Porównanie ARL dla kart Kendalla i autokorelacji dla modelu autoregresji

$\rho$	$n=10$		$n=50$	
	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



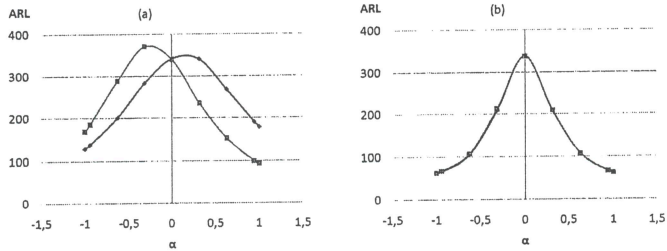
Rysunek 1: Porównanie ARL dla karty Kendalla (czarna linia) i karty autokorelacji (szara linia) dla modelu autoregresji (a)  $n = 10$ , (b)  $n = 50$

Tablica 6: Wartości ARL dla karty Kendalla z parametrem  $k = 3$  dla kopuły FGM z brzegowymi rozkładami normalnymi

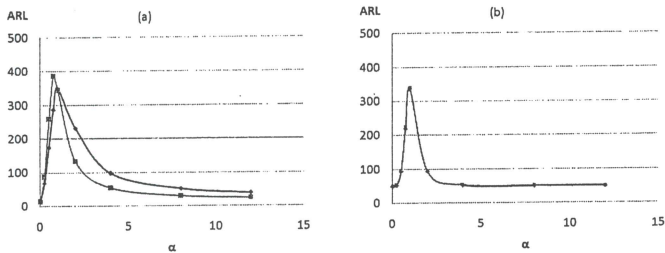
		$n$						
$\alpha$	$\rho$	6	7	8	9	10	20	50
1	0.318	128.83	297.60	470.94	371.06	389.99	191.19	112.55
0.942	0.3	129.50	315.19	510.86	395.31	420.79	216.42	127.04
0.628	0.2	136.43	402.49	729.07	578.95	651.87	443.14	288.60
0.314	0.1	145.80	490.94	1049.37	808.29	1016.84	969.05	955.37
0	0	144.74	578.02	1506.35	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

Tablica 7: Wartości ARL dla karty autokorelacji z parametrem  $k = 3$  dla kopuły FGM z brzegowymi rozkładami normalnymi

		$n$						
$\alpha$	$\rho$	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



Rysunek 2: Porównanie ARL dla karty Kendalla (linia czarna) i karty autokorelacji (linia szara) dla kopuły FGM o brzegowych rozkładach normalnych (a)  $n = 10$ , (b)  $n = 50$



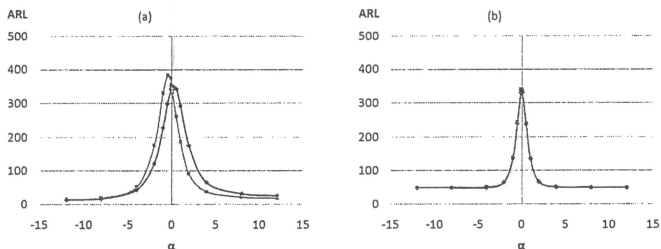
Rysunek 3: Porównanie ARL dla karty Kendalla (linia czarna) i karty autokorelacji (linia szara) dla kopuły Placketta o brzegowych rozkładach normalnych (a)  $n = 10$ , (b)  $n = 50$

Wyniki te potwierdzają przypuszczenie, że zachowanie się karty Kendalla skonstruowanej na podstawie nieparametrycznej statystyki nie zależy od rodzaju zależności występujących wśród obserwacji procesu. Przypuszczenia takiego nie można wysnuć dla karty autokorelacji. Wartości ARL otrzymane dla karty autokorelacji dla modelu autoregresji i kopuły FGM różnią się w znaczący sposób.

Podobnie jak dla poprzedniego modelu, dokonano porównania wyników dla obu kart przy ustalonych licznosciach próbek  $n = 10$  i  $n = 50$  w przypadku zastosowania różnych kopuł z brzegowymi rozkładami normalnymi. Dla prób 10-elementowych przyjęto  $k = 2, 71$  i  $k = 2, 73$ , odpowiednio, dla kart Kendalla i autokorelacji. Dla prób o licznosci  $n = 50$  dla obu kart przyjęto  $k = 2, 17$ .

Wnioski są takie same jak w przypadku modelu autoregresji. Dla prób 10-elementowych wartości ARL są mniejsze dla karty Kendalla, gdy zależność jest ujemna, zaś większe, gdy zależność jest dodatnia. Dla prób 50-elementowych różnice w działaniu obu kart nie są widoczne.

Wyciągnięte dotychczas wnioski przestają być prawdziwe, gdy jako rozkłady brzegowe przyjmiemy rozkłady inne od normalnego. Wykonano badania symulacyjne biorąc pod uwagę wszystkie rozważane kopuły z brzegowymi rozkładami wykładniczymi i równomiernymi. Rozpatrzono przy tym dwie sytuacje: pierwsza, gdy nie wiemy, iż rozkład danych jest różny od rozkładu normalnego; druga, gdy znamy postać rozkładu obserwacji. Rysunki 5–7 ilustrują wyniki otrzymane w przypadku wykładniczych rozkładów brzegowych. Rysunki 8–10 przedstawiają zaś wyniki dla modeli z jednostajnymi rozkładami brzegowymi. Na rysunkach linie czarna i szara opisują wartości ARL dla kart, odpowiednio, Kendalla



Rysunek 4: Porównanie ARL dla karty Kendalla (linia czarna) i karty autokorelacji (linia szara) dla kopuły Franka o brzegowych rozkładach normalnych (a)  $n = 10$ , (b)  $n = 50$

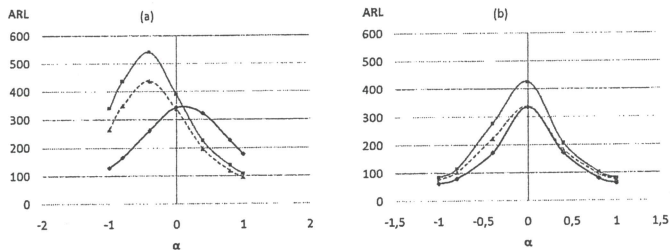
i autokorelacji w przypadku pierwszym, tj. gdy nie wiadomo, iż rozkład brzegowy nie jest normalny. W sytuacji, gdy znana jest postać rozkładu brzegowego można odpowiednio zmodyfikować kartę sterując wielkością parametru  $k$ . Modyfikacja taka jest konieczna dla karty autokorelacji. Na rysunkach przerywaną linią szarą oznaczono wartości ARL dla zmodyfikowanej karty autokorelacji.

Analizując wyniki zgromadzone na Rysunkach 5–10, otrzymujemy ciekawe wnioski. Mianowicie, w wielu rozpatrywanych przypadkach wyraźnie widoczna jest przewaga karty Kendalla. Dla modeli z wykładniczymi rozkładami brzegowymi obserwuje się ją już dla prób 10-elementowych. Tymczasem dla małych prób w przypadku modeli z jednostajnymi rozkładami brzegowymi karta Kendalla na pewno nie stanowi konkurencji dla prostej karty autokorelacji. Niezależnie jednak od charakteru rozkładów brzegowych karta Kendalla jest zdecydowanie lepszym rozwiązaniem od karty autokorelacji dla prób 50-elementowych. Szczególną jej przewagę obserwuje się dla modeli z wykładniczymi rozkładami brzegowymi. Nawet w sytuacji, gdy znana jest dokładana postać rozkładu brzegowego i możliwe jest odpowiednio ustawienie wielkości parametru  $k$ , karta Kendalla nie działa gorzej od zmodyfikowanej karty autokorelacji.

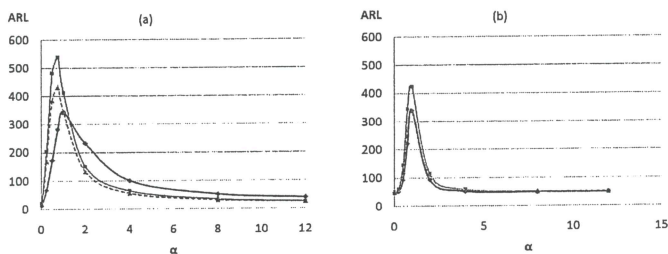
Powyższe wyniki pozwalają zaobserwować jak wrażliwa jest karta autokorelacji wobec postaci rozkładu brzegowego. Do badania zaproponowano dwa bardzo różne rozkłady: wykładniczy i jednostajny, by odzwierciedlić zachowanie się kart w skrajnych przypadkach (ciężkie ogony, skośność). Różnice widoczne w zachowaniu się obu kart w przypadku różnych od normalnego rozkładów brzegowych mają istotne znaczenie z praktycznego punktu widzenia. Podsumowując, kartę autokorelacji należy stosować z ostrożnością, jeśli brak jest pewności co do normalności rozkładu brzegowego. Z kolei okazuje się, że nawet w sytuacji, gdy znana jest postać rozkładu, a zatem można przyjąć odpowiednią wielkość parametru  $k$ , karta Kendalla nie zachowuje się gorzej od zmodyfikowanej karty autokorelacji.

Jak wcześniej wspomniano, czasami trudno jest dokładnie porównać efektywność kart Kendalla i autokorelacji. Dzieje się tak np. w przypadku, gdy rozkłady brzegowe są normalne. W takiej sytuacji wygodniej jest ocenić działanie obu kart na podstawie ich odporności wobec konkretnych zaburzeń. W punkcie 4 opisano stosowaną w niniejszej pracy metodę badania odporności kart kontrolnych. Zgodnie z nią przeprowadzono badania symulacyjne i otrzymano wyniki odporności obu kart wobec dwóch różnych zaburzeń.

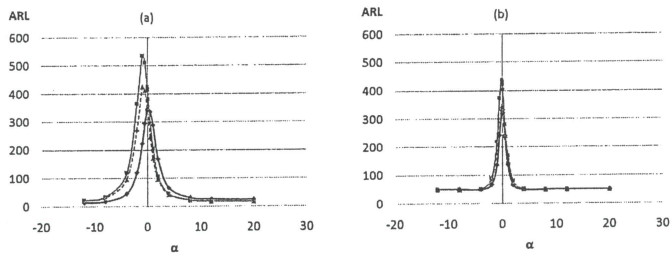
W pierwszej kolejności, zbadano odporność kart wobec odstępstwa od założenia o niezależności obserwacji, przy którym konstruowana jest karta autokorelacji. Jako model rozszerzony przyjęto model, w którym kolejne obserwacje opisuje kopuła FGM. Cechą charakterystyczną tego modelu, jest fakt, iż opisuje on słabe zależności pomiędzy obserwacjami (współczynnik korelacji przyjmuje wartości w przedziale  $[-\frac{1}{3}, \frac{1}{3}]$ ). Porównując wartości ARL otrzymane podczas badań symulacyjnych dla różnych warto-



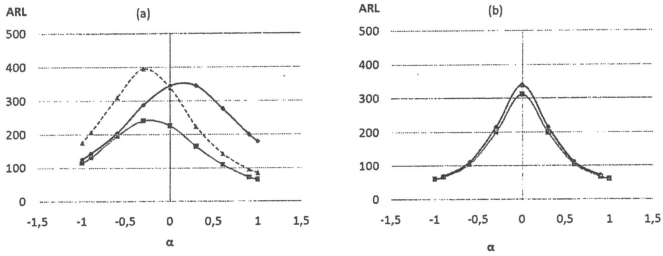
Rysunek 5: Porównanie ARL dla karty Kendalla (linia czarna), karty autokorelacji (linia szara) i zmodyfikowanej karty autokorelacji (przerywana linia szara) dla kopuły FGM o brzegowych rozkładach wykładniczych (a)  $n = 10$ , (b)  $n = 50$



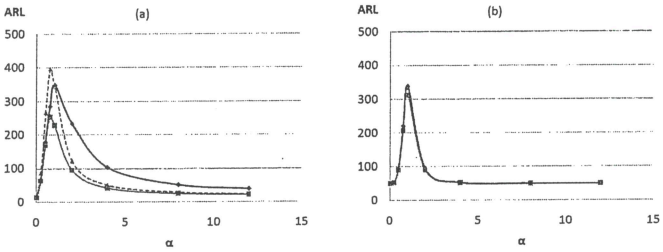
Rysunek 6: Porównanie ARL dla karty Kendalla (linia czarna), karty autokorelacji (linia szara) i zmodyfikowanej karty autokorelacji (przerywana linia szara) dla kopuły Placketta o brzegowych rozkładach wykładniczych (a)  $n = 10$ , (b)  $n = 50$



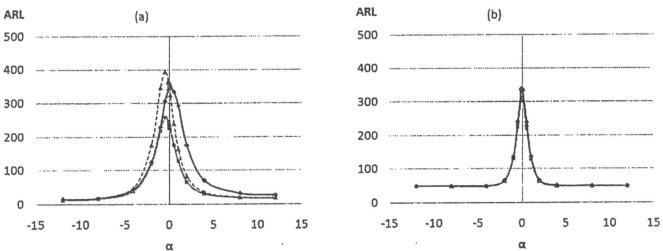
Rysunek 7: Porównanie ARL dla karty Kendalla (linia czarna), karty autokorelacji (linia szara) i zmodyfikowanej karty autokorelacji (przerywana linia szara) dla kopuły Franka o brzegowych rozkładach wykładniczych (a)  $n = 10$ , (b)  $n = 50$



Rysunek 8: Porównanie ARL dla karty Kendalla (linia czarna), karty autokorelacji (linia szara) i zmodyfikowanej karty autokorelacji (przerywana linia szara) dla kopuły FGM o brzegowych rozkładach jednostajnych (a)  $n = 10$ , (b)  $n = 50$



Rysunek 9: Porównanie ARL dla karty Kendalla (linia czarna), karty autokorelacji (linia szara) i zmodyfikowanej karty autokorelacji (przerywana linia szara) dla kopuły Placketta o brzegowych rozkładach jednostajnych (a)  $n = 10$ , (b)  $n = 50$



Rysunek 10: Porównanie ARL dla karty Kendalla (linia czarna), karty autokorelacji (linia szara) i zmodyfikowanej karty autokorelacji (przerywana linia szara) dla kopuły Franka o brzegowych rozkładach jednostajnych (a)  $n = 10$ , (b)  $n = 50$



ści parametru  $\alpha$  kopuły FGM dla poszczególnych rozmiarów próby, wyznaczono odporność kart wobec występowania słabych zależności dodatnich i ujemnych. Wyniki przedstawiają Rysunki 11 i 12.

Na Rysunkach 11, 12 obserwujemy większą odporność, a zatem mniejszą wrażliwość na występowanie słabych zależności wśród kolejnych obserwacji dla karty Kendalla, szczególnie wyraźną różnicę można zauważyć w przypadku dodatnich słabych zależności. Odporność obu kart jest porównywana dla prób o liczności co najmniej 9, gdyż z uwagi na dyskretny charakter statystyki  $\tau$  Kendalla porównanie kart nie jest możliwe dla mniejszych prób. Otrzymane wyniki w sposób bardziej jednoznaczny niż dotychczas pozwalają ocenić działanie obu kart.

Przewaga karty Kendalla była najwyraźniej widoczna w sytuacji, gdy rozkłady brzegowe były różne od normalnego. Potwierdzają to również wyniki odporności wobec odstępstwa od założeń o normalności. Przyjmując jako model rozszerzony rodzinę rozkładów Gamma( $a, 1$ ), gdzie  $a = 1, 2, 3, 4$  obliczono wartości ARL dla obu kart. Wyniki przedstawia Rysunek 13.

## 6 Rozmyta karta Kendalla

W klasycznej teorii statystycznego sterowania procesem wszystkie parametry kart kontrolnych oraz wyniki kontroli są opisane liczbami rzeczywistymi, a zatem są wyznaczone precyzyjnie. Sytuacja ulega skomplikowaniu, jeśli obserwacje procesu są nieprecyzyjne i określone liczbami rozmytymi. Proponowana w drugim punkcie karta Kendalla dla danych precyzyjnych posłużyła jako swego rodzaju podstawa do stworzenia rozmytej karty Kendalla, innymi słowy karty Kendalla dla danych rozmytych.

Rozmytą postać statystyki  $\tau$  Kendalla wprowadzono w pracy Héberta, Massona, Denœux (2003). W nowszym artykule tych samych autorów (Denœux, Massona Héberta, 2005) zaproponowano również ciekawą metodę wyznaczania jej funkcji przynależności. Hryniewicz (2007) używał rozmytej statystyki  $\tau$  Kendalla w analizie szeregów czasowych składających się z dwumianowych danych rozmytych. Prace te nie wykluczają jednakże rozpatrzenia bardziej ogólnego przypadku, kiedy szereg kolejnych obserwacji jest danym wektorem liczb rozmytych  $(\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n)$ .

W celu wyznaczenia rozmytej postaci statystyki  $\tau$  Kendalla dla rozważanego szeregu załóżono, że każda rozmyta obserwacja  $\tilde{Z}_i$  opisana jest za pomocą funkcji przynależności  $\mu_i(z)$ ,  $i = 1, \dots, n$ . Do wyznaczenia rozmytego odpowiednika statystyki Kendalla w rozważanym szeregu wykorzystano zasadę rozszerzania (Zadeh, 1965).

Łatwo zauważyć, że wielkość  $V_i$  daną wzorem (2) można zapisać równoważnie jako

$$V_i = \frac{\text{card}_{j \neq i} \{(Z_j, Z_{j+1}) : Z_j < Z_i, Z_{j+1} < Z_{i+1}\}}{n-2}, i = 1, \dots, n-1. \quad (15)$$

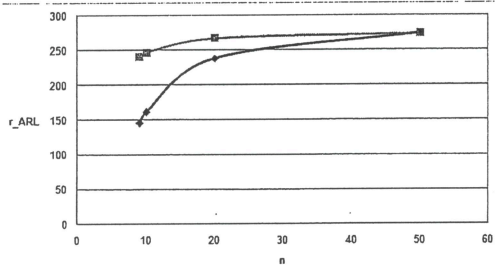
Przyjmując, że każda obserwacja  $\tilde{Z}_i$  zdefiniowana jest przez  $\alpha$ -cięcie  $[Z_{i,L}^\alpha, Z_{i,U}^\alpha]$ ,  $0 < \alpha \leq 1$ , podano rozmyty odpowiednik wielkości  $V_i$ . Oznaczono go jako  $\tilde{V}_i$  i określono za pomocą  $\alpha$ -cięcia  $[V_{i,L}^\alpha, V_{i,U}^\alpha]$ ,  $0 < \alpha \leq 1$ , gdzie

$$V_{i,L}^\alpha = \min_{\substack{z_i \in [Z_{i,L}^\alpha, Z_{i,U}^\alpha] \\ i=1, \dots, n}} \frac{\text{card}_{j \neq i} \{(z_j, z_{j+1}) : z_j < z_i, z_{j+1} < z_{i+1}\}}{n-2} \quad (16)$$

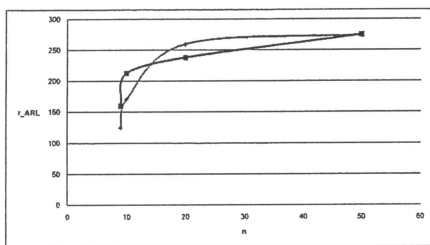
oraz

$$V_{i,U}^\alpha = \max_{\substack{z_i \in [Z_{i,L}^\alpha, Z_{i,U}^\alpha] \\ i=1, \dots, n}} \frac{\text{card}_{j \neq i} \{(z_j, z_{j+1}) : z_j < z_i, z_{j+1} < z_{i+1}\}}{n-2} \quad (17)$$

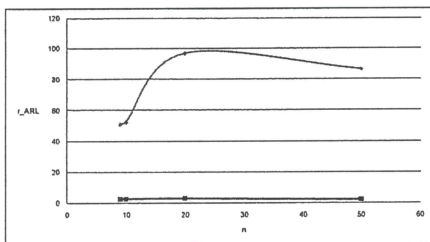
dla  $i = 1, \dots, n-1$ . Znając  $\alpha$ -cięcia  $[V_{i,L}^\alpha, V_{i,U}^\alpha]$ ,  $0 < \alpha \leq 1$  dla każdego  $i = 1, \dots, n-1$  można bezpośrednio wyznaczyć  $\alpha$ -cięcia dla rozmytej statystyki  $\tau$  Kendalla  $[\tau_L^\alpha, \tau_U^\alpha]$ ,  $0 < \alpha \leq 1$  i w ten sposób otrzymać jej funkcję przynależności.



Rysunek 11: Odporność karty Kendalla (linia czarna) oraz karty autokorelacji (linia szara) wobec słabych zależności dodatnich



Rysunek 12: Odporność karty Kendalla (linia czarna) oraz karty autokorelacji (linia szara) wobec słabych zależności ujemnych



Rysunek 13: Odporność karty Kendalla (linia czarna) oraz karty autokorelacji (linia szara) wobec odstępstwa od normalności

Pomimo zwięzłych postaci wzorów (16) i (17) wyznaczenie funkcji przynależności rozmytej statystyki  $\tau$  Kendalla okazuje się, w ogólnym przypadku, zadaniem trudnym i skomplikowanym obliczeniowo. W sytuacji, gdy liczba obserwacji rozmytych w całym zbiorze jest niewielka można skorzystać z ogólnej metody obliczania dokładnej postaci funkcji przynależności zaproponowanej przez Héberta, Massona, Denœux (2003). Innym rozwiązaniem jest zastosowanie algorytmów optymalizacji stochastycznej opartych na metodach Monte Carlo (Denœux, Massona Héberta, 2005). Niemniej, w szczególnym przypadku rozmytej karty Kendalla, kiedy liczba analizowanych obserwacji rozmytych może być stosunkowo duża, wymienione metody wcale nie muszą być skuteczne. Istnieje zatem potrzeba zaprojektowania efektywnego algorytmu obliczającego  $\tau_L^\alpha$  i  $\tau_U^\alpha$ .

Pierwsza propozycja takiego algorytmu została przedstawiona w pracy Hryniewicza, Szewi (2008). Wyniki otrzymane w trakcie wielu doświadczeń symulacyjnych skłaniały jednak do modyfikacji owego algorytmu w celu uzyskania poprawy jego efektywności. Wprowadzono kilka zmian w jego budowie. W autoreferacie w dalszej części tego punktu ograniczono się jedynie do podania komentarza na temat działania proponowanego algorytmu oraz przedstawienia rozmytej karty Kendalla. Dokładny opis algorytmu można znaleźć w (Hryniewicz, Szewi, 2008).

Uzyskane wyniki symulacji zamieszczone w wymienionej poprzednio pracy prowadzą do wniosku, iż z uwagi na specyfikę działania algorytmów nie zawsze wyznaczona para  $[\tau_L, \tau_U]$  jest właściwym oszacowaniem  $\alpha$ -cięcia dla rozmytej statystyki  $\tau$  Kendalla. Niemniej, w przypadku dużych prób otrzymywane wyniki okazują się bardziej wiarygodne, zwłaszcza te dotyczące wielkości  $\tau_U$ .

Na podstawie rozmytej statystyki  $\tau$  Kendalla, skonstruowana została rozmyta karta Kendalla. W przeciwieństwie do kart kontrolnych konstruowanych dla danych precyzyjnych, na rozmytych kartach nie są zaznaczane kolejne pomiary procesu. Nie możliwe jest zatem jednoznaczne określenie, czy obserwowany układ pomiarów wskazuje na istnienie zależności pomiędzy kolejnymi pomiarami procesu. Problem ten można ominąć np. znajdując precyzyjną reprezentację rozmytych wartości  $\tilde{\tau}_n$ , czy też wprowadzając dodatkowy warunek przy podejmowaniu decyzji. W pracy zastosowano w tym celu posybilistyczne podejście zaproponowane przez Hryniewicza (2006). Zgodnie z nim odpowiednie decyzje podejmowane są po wyznaczeniu tzw. wskaźników możliwości i konieczności, iż obserwowana wielkość  $\tilde{\tau}_n$  znajdzie się poza granicami kontrolnymi (5) i (6).

W celu podjęcia decyzji, czy kolejne obserwacje rozmyte na karcie Kendalla są zależne zaproponowano użycie wskaźnika możliwości PD oraz wskaźnika konieczności NSD (Dubois, Prade, 1983). Bez odwoływania się w tym miejscu do definicji owych wskaźników, można przedstawić przyjętą zasadę podejmowania odpowiednich decyzji.

Jako granice kontrolne, odpowiednio, górną i dolną potraktowano następujące wartości wskaźnika NSD

$$NSD_U = Ness(\tilde{\tau}_n > K_U) = 1 - \alpha_{N,U}, \quad (18)$$

gdzie

$$\alpha_{N,U} = \inf_{\alpha} (\alpha : \tau_{n,L}^\alpha \geq K_U) \quad (19)$$

oraz

$$NSD_L = Ness(K_L > \tilde{\tau}_n) = 1 - \alpha_{N,L}, \quad (20)$$

gdzie

$$\alpha_{N,L} = \sup_{\alpha} (\alpha : \tau_{n,U}^\alpha \leq K_L). \quad (21)$$

Do podjęcia jednoznacznej decyzji konieczne jest przyjęcie odpowiedniej wartości minimalnej dla  $NSD_U$  i  $NSD_L$ . Jeśli obliczona wartość wskaźnika NSD jest większa bądź równa od tej wartości minimalnej,

wówczas możemy stwierdzić, że obserwowane dane rozmyte wykazują charakterystyczny układ dla danych zależnych.

Stosując klasyczne narzędzia SPC dodatkowo korzysta się z tak zwanych sygnałów ostrzegawczych. Dla omawianej rozmytej karty Kendalla jako narzędzia generującego takie sygnały można użyć wskaźnika możliwości PD. Podobnie do rozważanego wcześniej wskaźnika konieczności, wartości wskaźnika PD odpowiadające granicom kontrolnym wyznaczamy jako

$$PD_U = Poss(\tilde{\tau}_n > K_U) = 1 - \alpha_{P,U}, \quad (22)$$

gdzie

$$\alpha_{P,U} = \sup_{\alpha} (\alpha : \tau_{n,U}^{\alpha} \geq K_U) \quad (23)$$

oraz

$$PD_L = Poss(K_L > \tilde{\tau}_n) = 1 - \alpha_{P,L}, \quad (24)$$

gdzie

$$\alpha_{P,L} = \inf_{\alpha} (\alpha : \tau_{n,L}^{\alpha} \leq K_L). \quad (25)$$

W tym przypadku również przyjmowana jest wartość minimalna dla  $PD_U$  i  $PD_L$ . Jeżeli obliczona wartość wskaźnika PD jest większa bądź równa od ustalonej wartości minimalnej, to możemy przyjąć ostrzeżenie, że układ obserwowanych danych rozmytych jest charakterystyczny dla danych zależnych.

Własności zaproponowanej rozmytej wersji karty Kendalla nie zostały jeszcze dokładnie zbadane. Kierunki dalszych badań związane z analizą zachowania się karty dla nieprecyzyjnych danych zostały określone w ostatnim punkcie autoreferatu.

## 7 Kierunki przyszłych badań

W pracy zaproponowano oryginalne zastosowanie nieparametrycznej procedury statystycznej - karty Kendalla - do wykrywania rozregulowań procesu ujawniających się w postaci zależności występujących pomiędzy kolejnymi jego obserwacjami. Wykonane dotychczas badania symulacyjne niosą ciekawe i obiecujące wyniki.

Statystyczne własności karty Kendalla nie zależą od postaci rozkładu obserwowanych danych. Z tego powodu może być ona bezpiecznie stosowana podczas wstępnych badań procesu, kiedy dostępne informacje nie pozwalają na dokładny opis procesu, a w związku z tym na wybór odpowiednio efektywniejszego narzędzia statystycznego. Aspekt ten pozostaje uzasadniony także w przypadku, gdy dane nie są określone precyzyjnie. Jednakże z uwagi na nieprecyzyjność danych konieczne obliczenia stają się znacznie bardziej zawiłe. Proponowany algorytm do wyznaczenia  $\alpha$ -cięcia dla rozmytej statystyki  $\tau$  Kendalla ma raczej prostą postać, niemniej jego efektywność nie jest w dalszym ciągu zadowalająca.

Dalsze badania dotyczące postaci algorytmu obliczającego  $\alpha$ -cięcia dla rozmytej statystyki  $\tau$  Kendalla powinny prowadzić do zaproponowania algorytmu znacznie bardziej efektywnego, wyznaczającego w dokładniejszy sposób  $\alpha$ -cięcia.

Z kolei innym ciekawym i ważnym problemem jest analiza wartości ARL dla rozmytej karty Kendalla, czyli sprawdzenie w jaki sposób nieprecyzyjność danych wpływa na istotną charakterystykę kart kontrolnych, jaką jest średnia liczba pomiarów do pojawienia się pierwszego sygnału alarmowego - ARL.

## 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.
- Borror C. M., Montgomery, D. C., Runger G. C., (1999), Robustness of the EWMA Control Chart to Non-normality. *Journal of quality Technology* 31, 309–316.
- Box, G. E. P., Anderson, S.L., (1955), Permutation theory in the derivation of robust criteria and the study departures from assumptions. *Journal of the Royal Statistical Society Series B* 17, 1–34.
- Chan, L. K., Hapuarachchi, K. P., Macpherson B. D., (1988), Robustness of  $\bar{X}$  and  $R$  Charts. *IEEE Transactions on Reliability* 37, 117–123.
- Denceux, T., Masson, M.-H., Hébert, P.A., (2005), Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets and Systems* 153, 1-28.
- Dubois, D., Prade, H., (1983), Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences* 30, 184–244.
- Embrechts, P., Lindskog, F., McNeil, A., (2003), Modeling dependence with copulas and applications to risk management. In: *Rachev S. (eds.) Handbook of Heavy Tailed Distributions in Finance* Elsevier, New York, 329–384.
- Ferguson, T.S. and Genest, C. and Hallin, M., (2000), Kendall's tau for Serial Dependence. *The Canadian Journal of Statistics* 28, 587-604.
- Hallin, M., Mélard, G., (1988), Rank-Based Tests for Randomness Against First-Order Serial Dependence. *Journal of the American Statistical Association* 83, 1117-1128.
- Hampel, F. R., (1974), The influence curve and its role in robust estimation. *Journal of the American Statistical Association* 62, 1179–1186.
- Hébert, P.-A., Masson, M.H., Denceux, T., (2003), Fuzzy rank correlation between fuzzy numbers. *Proc. of IFSA World Congress Istanbul, Turkey*, 224–227.
- Hryniewicz, O., (2006), Possibilistic decisions and fuzzy statistical tests. *Fuzzy Sets and Systems* 157, 2665–2673.
- Hryniewicz, O., (2007), Looking for dependencies in short time series using imprecise statistical data. In: *Castillo O., Melin P., Ross O.M., Cruz R.S., Pedrycz W., Kacprzyk J. (Eds.): Theoretical Advances and Applications of Fuzzy Logic and Soft Computing* Springer-Verlag, Berlin Heidelberg, 201–208.
- Hryniewicz, O., Szediw, A. (2008), Kendall Control Chart for Fuzzy Data. In: *Atanassov K., Chountas P., Kacprzyk J., Krawczak M., Melo-Pinto P., Szmidt E., Zadrozny S. (Eds.): Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Applications.* EXIT, Warszawa, 2008 (w druku).
- Huber, P. J., (1964), Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73–101.

- Jiang, W., Tsui, K.L., Woodall, W.H., (2000), The New SPC Monitoring Method: The ARMA Chart. *Technometrics* 42, 399–410.
- Johnson, R.A., Bagshaw, M., (1974), The Effect of Serial Correlation on the Performance of CUSUM Tests. *Technometrics* 16, 103–112.
- Knoth, S., Schmid, W., Schone, A., (2001), Simultaneous Shewhart-Type Charts for the Mean and the Variance of a Time Series. In: *Frontiers in Statistical Quality Control VI*, Lenz, H.J. and Wilrich, P.T. (eds.), Physica Verlag, Heidelberg, 61–79.
- Lu, C.W., Reynolds Jr., M.R., (1999a), EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology* 31, 166–188.
- Lu, C.W., Reynolds Jr., M.R., (1999b), Control Charts for Monitoring the Mean and Variance of Auto-correlated Processes. *Journal of Quality Technology* 31, 259–274.
- Lu, C.W., Reynolds Jr., M.R., (2001), CUSUM Charts for Monitoring An Autocorrelated Process. *Journal of Quality Technology* 33, 316–334.
- Maragah, H.D. and Woodall, W.H., (1992), The Effect of Autocorrelation on the Retrospective X-Chart. *Journal of Statistical Computation and Simulation* 40, 29–42.
- Mardia, K. W., (1970), *Families of Bivariate Distributions* Griffin, London.
- Montgomery, D.C., Mastrangelo, C.M., (1991), Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology* 23, 179–193.
- Nelsen, R.B., (2006), *An Introduction to Copulas*, 2nd ed. Springer, New York.
- Schmid, W., (1995), On the run length of a Shewhart control chart for correlated data. *Statistical Papers* 36, 111–130.
- Schmid, W., (1997), On EWMA charts for time series. In: *Frontiers in Statistical Quality Control V*, Lenz, H.J. and Wilrich, P.T. (eds.), Physica Verlag, Heidelberg, 115–137.
- Sklar, A., (1959), Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ Inst Statist Univ Paris* 8, 229–231.
- Timmer, D.H., Pignatello Jr, J., Longnecker, M., (1998), The development and evaluation of CUSUM-based control charts for an AR(1) process. *IIE Transactions* 30, 525–534.
- 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.
- 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.
- 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.

Zadeh, L.A., (1965), Fuzzy Sets. *Information and Control* 8, 338–353.

Zhang, N.F., (1998), Statistical Control Chart for Stationary Process Data. *Technometrics* 40, 24–38.

Zieliński, R., (1978), O mierzeniu odporności statystyk. *Zastosowania Matematyki* 12, 71–76.











