PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1126 — 2006
TAKSONOMIA 13
Klasyfikacja i analiza danych – teoria i zastosowania

Joanicjusz Nazarko, Mikołaj Rybaczuk, Arkadiusz Jurczuk
Politechnika Białostocka

WPŁYW POZIOMU ZAKŁÓCEŃ LOSOWYCH NA MOŻLIWOŚĆ IDENTYFIKACJI MODELI ARIMA DLA NIESTACJONARNYCH SZEREGÓW CZASOWYCH

1. Wstęp

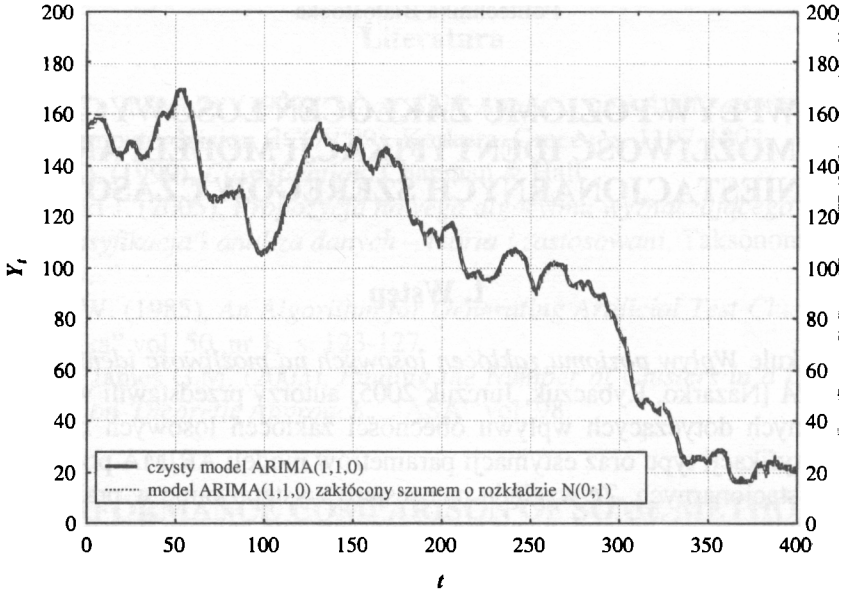
W artykule *Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA* [Nazarko, Rybaczuk, Jurczuk 2005] autorzy przedstawili wyniki badań symulacyjnych dotyczących wpływu obecności zakłóceń losowych na prawidłowość identyfikacji typu oraz estymacji parametrów modeli ARIMA podstawowych procesów stacjonarnych. Ze względu na częstsze występowanie w praktyce procesów niestacjonarnych zdecydowano się na rozszerzenie zakresu badań i określenie wpływu zakłóceń losowych na rozpoznawalność modeli klasy ARIMA również dla tego typu szeregów. W niniejszym artykule przedstawiono wyniki badań wpływu zakłóceń losowych o charakterze białego szumu o różnym poziomie wariancji oraz zakłóceń o rozkładzie niesymetrycznym na identyfikowalność postaci i jakość estymacji modeli niestacjonarnych: autoregresyjnych – ARIMA(1,1,0), średniej ruchomej – ARIMA(0,1,1) oraz mieszanych – ARIMA(1,1,1). Zgodnie z metodyką zaproponowaną przez Boxa i Jenkinsa do oceny typu modelu wykorzystano funkcje autokorelacji (ACF) i autokorelacji cząstkowej (PACF).

2. Autoregresyjne niestacjonarne szeregi czasowe – ARIMA(1, 1, 0)

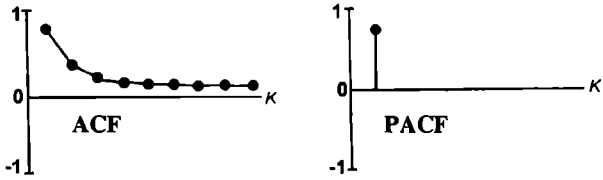
Poddawane zakłóceniom losowym modele ARIMA(1,1, 0) niestacjonarnych autoregresyjnych szeregów czasowych Y_t uzyskiwano poprzez konwersję białego szumu a_t o rozkładzie $N(0;\sigma_a)$ za pomocą filtra $y_t = a_t+(1+\phi)\, y_{t-1} - \phi\, y_{t-2}$ [Box, Jenkins

1983; DeLurgio 1998; Hanke, Reitsch, Wichern 2001]. Przedstawione w artykule wyniki dotyczą modelu generowanego dla wartości parametrów $\sigma_a = 1$ i $\phi = 0{,}8$.

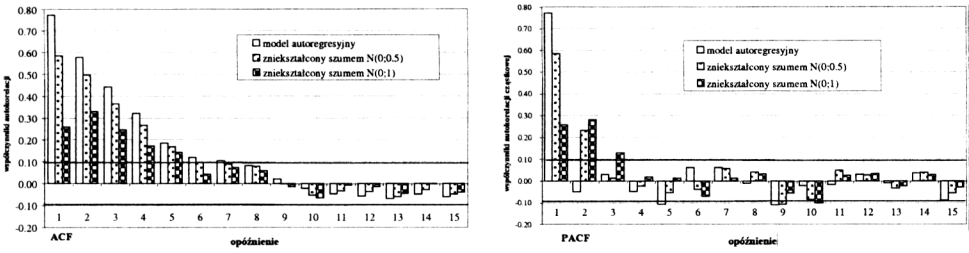
Na rysunku 1 pokazano przebieg realizacji analizowanego niestacjonarnego procesu autoregresyjnego. W wyniku operacji jednokrotnego różnicowania szereg ten został przekształcony do postaci stacjonarnej [Box, Jenkins 1983]. Przebiegi funkcji autokorelacji ACF i autokorelacji cząstkowej PACF zróżnicowanego szeregu ΔY_t oceniono, porównując z wzorcowymi funkcjami charakterystycznymi dla modeli autoregresyjnych rzędu pierwszego przy $\phi > 0$, przedstawionymi na rys. 2. Funkcje ACF i PACF dla różnic ΔY_t wygenerowanego „wzorcowego" szeregu Y_t oraz tego szeregu z nałożonym zakłóceniem o rozkładzie $N(0; 1)$ przedstawiono na rys. 3.



Rys. 1. Wygenerowana realizacja niestacjonarnego procesu autoregresyjnego ($\sigma_a = 1$, $\phi = 0{,}8$) oraz procesu zakłóconego szumem $N(0; 1)$
Źródło: opracowanie własne.



Rys. 2. Wzorcowa postać funkcji ACF i PACF dla procesu ARIMA(1,1,0)
Źródło: na podstawie [Hanke, Reitsch, Wichern 2001, s. 455].



Rys. 3. Funkcje autokorelacji ACF i autokorelacji cząstkowej PACF różnic ΔY_t niestacjonarnego modelu autoregresyjnego bez zakłóceń ($\sigma_a = 1$, $\phi = 0{,}8$) oraz modelu zakłóconego szumami o rozkładach $N(0; 0{,}5)$ i $N(0; 1)$
Źródło: opracowanie własne.

W przypadku niestacjonarnych modeli autoregresyjnych zniekształcenia funkcji ACF i PACF nasilają się znacznie szybciej niż w przypadku modeli stacjonarnych [Nazarko, Rybaczuk, Jurczuk 2005]. Już przy nieznacznych odchyleniach kształtów wykresów szeregu czasowego bez zakłóceń i szeregu z losowym zakłóceniem $N(0; 1)$ (rys. 1) zmiany w kształtach funkcji autokorelacji i autokorelacji cząstkowej są znaczne w stosunku zarówno do funkcji ACF i PACF modelu „czystego" (rys. 3), jak i do wykresu wzorcowego (rys. 2) i prowadzić mogą do rozpoznania niewłaściwego typu modelu. Mogą one wręcz sugerować, iż modelem adekwatnym w badanym przypadku jest model autoregresji rzędu trzeciego ARIMA(3,1,0).

W tabeli 1 przedstawiono wyniki estymacji parametru ϕ dla wygenerowanego niestacjonarnego procesu autoregresyjnego z $\sigma_a = 1$ i $\phi = 0{,}8$ (n = 400) oraz procesów zakłóconych.

Tabela 1. Wyniki estymacji parametru ϕ dla wygenerowanego niestacjonarnego szeregu autoregresyjnego z $\sigma_a = 1$, $\phi = 0{,}8$ ($n = 400$) oraz sygnałów zakłóconych

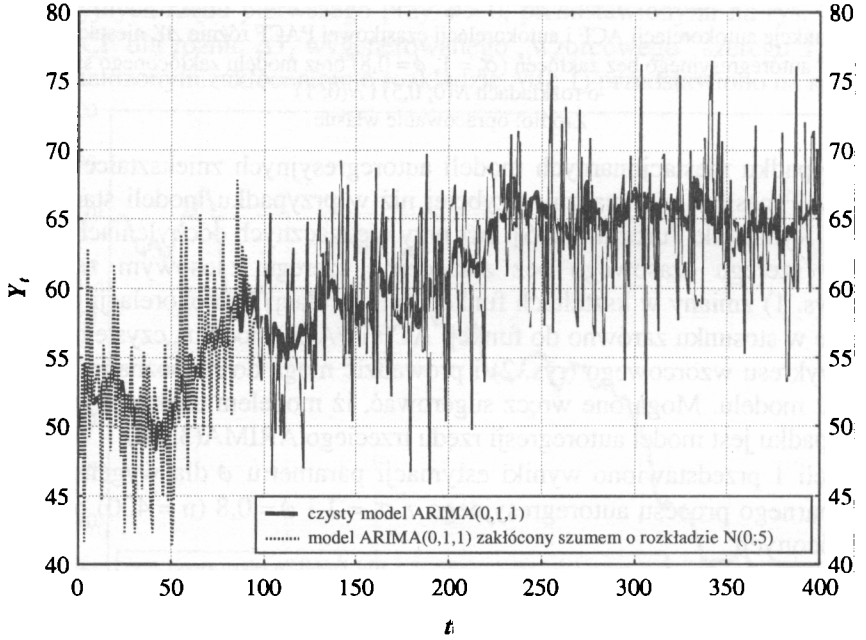
Model	Zakłócenie	Wartość estymatora	Asymptotyczny błąd standardowy	Istotność p	95% przedział ufności	
					od	do
ARIMA (1,1,0)	brak	0,7820	0,0311	0,0000	0,7209	0,8431
	$N(0;0{,}3)$	0,7737	0,0316	0,0000	0,7115	0,8358
	$N(0;0{,}5)$	0,5999	0,0401	0,0000	0,5211	0,6787
	$N(0;0{,}8)$	0,6856	0,0364	0,0000	0,6141	0,7572
	$N(0;1)$	0,2773	0,0482	0,0000	0,1825	0,3720

Źródło: opracowanie własne na podstawie obliczeń przeprowadzonych z wykorzystaniem programu Statistica.

Uzyskane wyniki estymacji wskazują, jak szybko w miarę wzrostu zaburzeń pogarsza się jakość oszacowania parametru ϕ procesu autoregresji. Już przy zniekształceniu sygnału szumem o rozkładzie $N(0; 0{,}5)$ wartość estymatora znajduje się poza 95% przedziałem ufności wartości parametru ϕ oszacowanej dla szeregu bez zakłóceń. Można zatem stwierdzić, że identyfikowalność rzędu modelu oraz proces estymacji parametru modeli ARIMA(1, d, 0) są bardzo wrażliwe na występowanie w szeregach czasowych zakłóceń losowych.

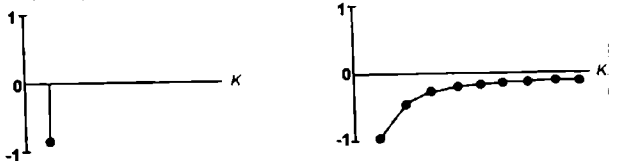
3. Modele niestacjonarnych szeregów czasowych średniej ruchomej – ARIMA(0,1,1)

Model niestacjonarnego szeregu czasowego średniej ruchomej Y_t można uzyskać poprzez przekształcenie białego szumu a_t o rozkładzie $N(0,\sigma_a)$ za pomocą filtru $y_t = y_{t-1} + a_t - \theta a_{t-1}$ [Box, Jenkins 1983; DeLurgio 1998; Hanke, Reitsch, Wichern 2001].

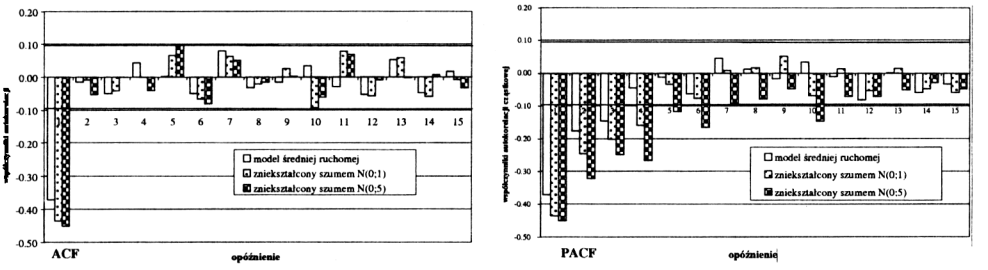


Rys. 4. Wygenerowana realizacja niestacjonarnego procesu średniej ruchomej ARIMA(0,1,1), przy $\sigma_a = 1$ i $\theta = 0{,}5$ oraz procesu zakłóconego szumem $N(0; 5)$
Źródło: opracowanie własne.

Zakłóceniom poddano realizację niestacjonarnego szeregu średniej ruchomej ($n = 400$) dla $\sigma_a = 1$ oraz $\theta = 0{,}5$. W celu podkreślenia różnicy pod względem wrażliwości na zakłócenia procesów autoregresji i średniej ruchomej na rys. 4 przedstawiono realizację modelu ARIMA(0,1,1) czystego oraz zakłóconego znacznym szumem o rozkładzie $N(0; 5)$.



Rys. 5. Wzorcowa postać funkcji ACF i PACF dla procesu ARIMA (0,1,1)
Źródło: na podstawie [Hanke, Reitsch, Wichern 2001, s. 456].



Rys. 6. Funkcje autokorelacji ACF i autokorelacji cząstkowej PACF różnic ΔY_t niestacjonarnego modelu średniej ruchomej bez zakłóceń ($\sigma_a = 1$, $\theta = 0{,}5$) oraz modelu zakłóconego szumami o rozkładach $N(0;1)$ i $N(0;5)$
Źródło: opracowanie własne.

Analizując wykres przedstawiony na rys. 4, można zauważyć, że przy $\sigma = 5$ amplituda zakłóceń jest zbliżona do zakresu wartości od maksimum do minimum wyjściowego szeregu czasowego, przy czym funkcje ACF i PACF (rys. 6) nie zmieniają istotnie kształtu. Porównując wykresy funkcji autokorelacji i autokorelacji cząstkowej różnic ΔY_t z wzorcowymi przebiegami ACF i PACF procesu średniej ruchomej rzędu pierwszego (rys. 5), można uznać badany proces za zgodny z modelem ARIMA(0,1,1). Nieznaczne zmiany kształtu funkcji ACF i PACF w miarę narastania zakłóceń mogą sugerować, że proces identyfikacji w tym przypadku nie jest zbyt wrażliwy na wprowadzane zakłócenia losowe. Jednakże ocena ich wpływu na dokładność oszacowania parametru θ wskazuje na istotny wpływ poziomu zakłóceń na poprawność estymacji modelu procesu pierwotnego. Uzyskane wyniki estymacji przedstawiono w tab. 2.

Tabela 2. Wyniki estymacji parametru θ dla wygenerowanego niestacjonarnego szeregu średniej ruchomej z $\sigma_a = 1$ i $\theta = 0{,}5$ ($n = 400$) oraz sygnałów zakłóconych

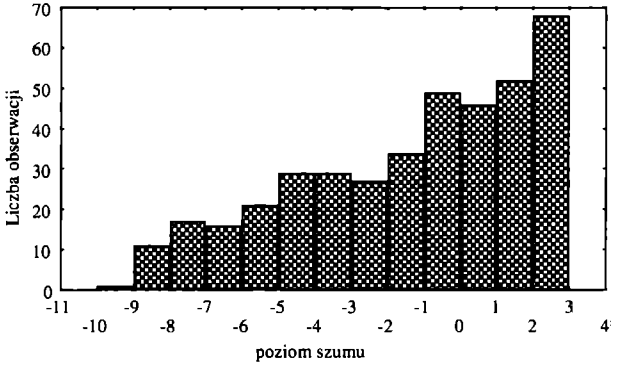
Model	Zakłócenie	Wartość estymatora	Asymptotyczny błąd standardowy	Istotność p	95% przedział ufności	
					od	do
ARIMA (0,1,1)	brak	0,4832	0,0454	0,0000	0,3939	0,5725
	$N(0;0{,}5)$	0,5469	0,0424	0,0000	0,4636	0,6303
	$N(0;1)$	0,6455	0,0682	0,0000	0,5703	0,7207
	$N(0;3)$	0,8371	0,0282	0,0000	0,7815	0,8926
	$N(0;5)$	0,9021	0,0232	0,0000	0,8565	0,9477

Źródło: opracowanie własne na podstawie obliczeń przeprowadzonych z wykorzystaniem programu Statistica.

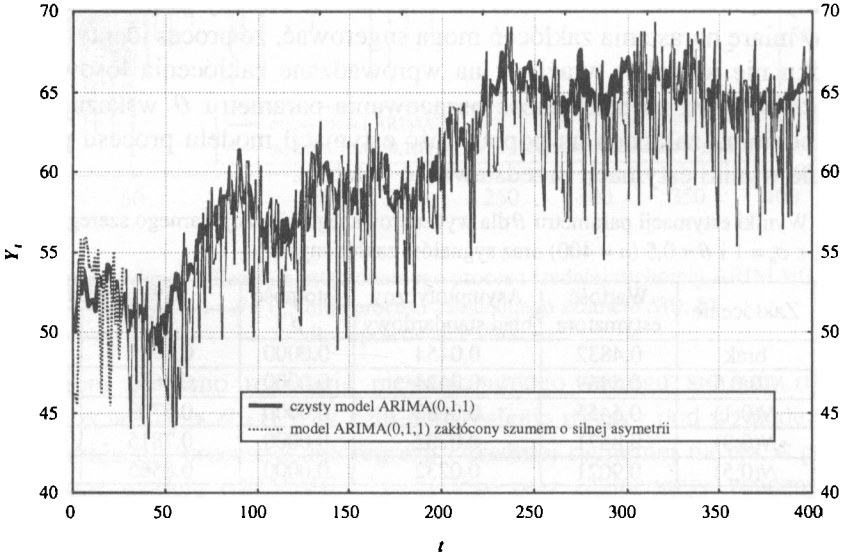
Przeprowadzona analiza wykazała, że szacowane wartości parametru θ już przy zakłóceniu szumem $N(0;1)$ znacznie odbiegają od założonej pierwotnie wartości i wykraczają poza zakres 95% przedziału ufności oszacowanej wartości parametru ϕ szeregu bez zakłóceń (tab. 2).

Uzyskane wyniki oceny, wskazujące na nieznaczny wpływ silnych zakłóceń losowych na poprawność identyfikacji typu modelu ARIMA (0,1,1), skłoniły auto-

rów do sprawdzenia wpływu asymetrii zakłóceń na proces identyfikacji. W przeanalizowanym dodatkowo wypadku, na modelowy niestacjonarny szereg średniej ruchomej nałożono szum o rozkładzie silnie niesymetrycznym (rys. 7). Oba szeregi, czysty i zakłócony, przedstawiono na rys. 8.



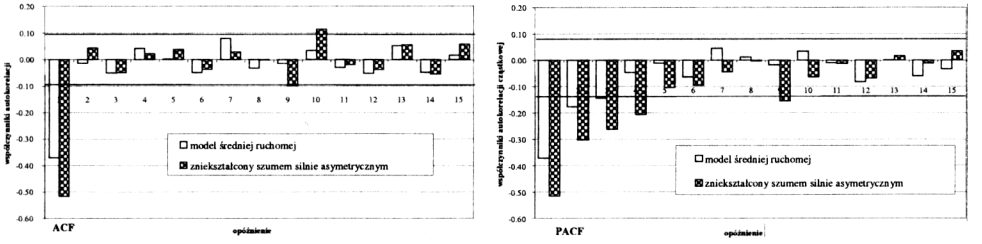
Rys. 7. Empiryczny rozkład szumu silnie asymetrycznego o zakresie wartości [-10; 3] – rozstęp prawie osiąga wartości wygenerowanego niestacjonarnego szeregu czasowego klasy ARIMA(0,1,1)
Źródło: opracowanie własne.



Rys. 8. Wygenerowana realizacja niestacjonarnego procesu średniej ruchomej ARIMA(0,1,1) przy $\sigma_a = 1$ i $\theta = 0{,}5$ oraz procesu zakłóconego szumem o rozkładzie wg rys. 7
Źródło: opracowanie własne.

Wprowadzenie znacznego zakłócenia z silną asymetrią nadal nie zmienia wyraźnie kształtu funkcji ACF i PACF (rys. 9). Ich przebieg sugeruje, że proces może być nadal identyfikowany jako ARIMA(0,1,1). Jednakże, podobnie jak poprzednio,

oszacowane wartości parametrów znacznie odbiegają od wartości przyjętej w czasie generowania (tab. 3).



Rys. 9. Funkcje autokorelacji ACF i autokorelacji cząstkowej PACF różnic ΔY_t niestacjonarnego modelu średniej ruchomej bez zakłóceń ($\sigma_a = 1$, $\theta = 0{,}5$) oraz modelu zakłóconego szumem o rozkładzie silnie asymetrycznym
Źródło: opracowanie własne.

Tabela 3. Wyniki estymacji parametru θ dla wygenerowanego niestacjonarnego szeregu średniej ruchomej z $\sigma_a = 1$ i $\theta = 0{,}5$ ($n = 400$) oraz sygnału zniekształconego silnie asymetrycznym szumem

Zakłócenie	Wartość estymatora	Asymptotyczny błąd standardowy	Istotność p	95% przedział ufności	
				od	do
Brak	0,4832	0,0454	0,0000	0,3939	0,5725
Z silną asymetrią	0,8289	0,0282	0,0000	0,7735	0,8843

Źródło: opracowanie własne na podstawie obliczeń przeprowadzonych z wykorzystaniem programu Statistica.

W wypadku niestacjonarnego procesu szeregu średniej ruchomej zakłócenie go szumem z silną asymetrią powoduje powstanie znaczących rozbieżności estymowanych wartości parametru θ modelu.

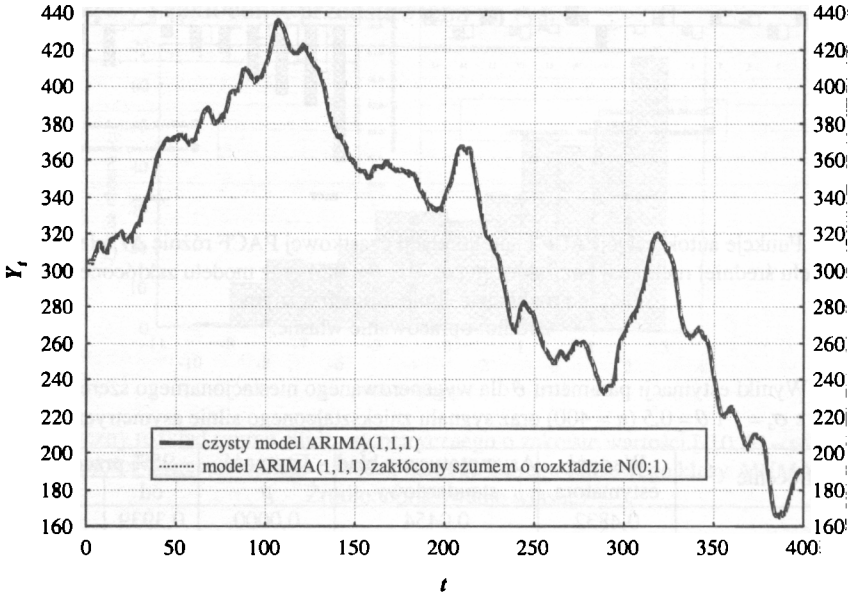
4. Mieszany niestacjonarny model autoregresyjny i średniej ruchomej – ARIMA(1, 1, 1)

Kolejnym etapem badań była ocena wpływu zakłóceń losowych na identyfikowalność niestacjonarnego procesu mieszanego rzędu pierwszego. Na rysunku 10 przedstawiono szereg klasy ARIMA(1,1,1) wygenerowany za pomocą przekształcenia $y_t = (1+\phi)\, y_{t-1} - \phi y_{t-2} + a_t - \theta a_{t-1}$ [Box, Jenkins 1983; DeLurgio 1998; Hanke, Reitsch, Wichern 2001]. Realizacje procesu uzyskiwano, przyjmując $\sigma_a = 1$, $\phi = 0{,}8$, $\theta = -0{,}5$ ($n = 400$).

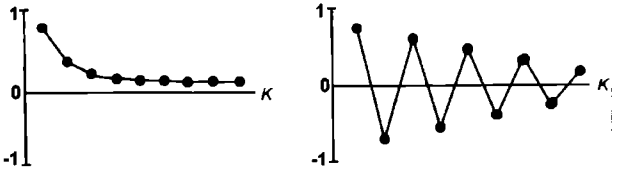
Na rysunku 10 przedstawiono realizację procesu pierwotnego oraz wykres tego szeregu z nałożonym szumem o rozkładzie $N(0; 1)$. Jak widać, wprowadzone zakłócenia powodują jedynie „pogrubienie" linii szeregu czystego. Amplituda zakłóceń jest bardzo mała w stosunku do amplitudy szeregu.

Oceniając funkcje autokorelacji ACF, stwierdzono ich nieznaczną zmienność względem wzorcowej krzywej (rys. 11 i 12) dla zakłóceń na poziomie od $N(0;0{,}25)$

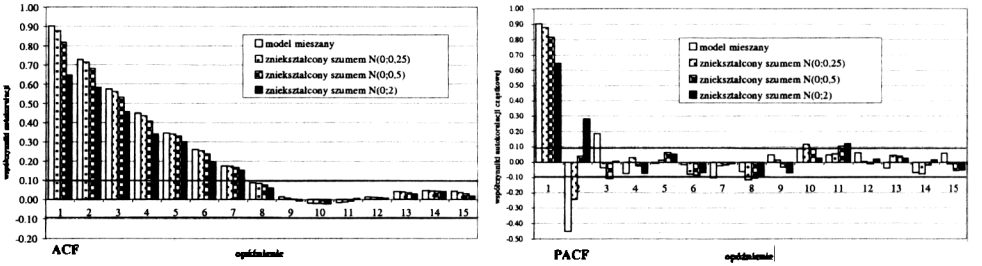
do $N(0;2)$. Funkcja PACF już przy szumie $N(0;0,5)$ sugeruje, iż jest to model klasy ARIMA(1,1,0). Składowa średniej ruchomej całkowicie zanika. Przy dalszym zwiększaniu amplitudy szumu pojawia się, ale z przeciwnym znakiem.



Rys. 10. Wygenerowana realizacja niestacjonarnego procesu mieszanego ARIMA(1,1,1) przy $\sigma_a = 1$, $\phi = 0{,}8$ i $\theta = -0{,}5$ oraz procesu zakłóconego szumem o rozkładzie $N(0; 1)$
Źródło: opracowanie własne.



Rys. 11. Wzorcowa postać funkcji ACF i PACF dla procesu ARIMA (1,1,1)
Źródło: na podstawie [Hanke, Reitsch, Wichern 2001, s. 457].



Rys. 12. Funkcje autokorelacji ACF i autokorelacji cząstkowej PACF różnic ΔY_t niestacjonarnego modelu mieszanego bez zakłóceń ($\sigma_a = 1$, $\phi = 0{,}8$ i $\theta = -0{,}5$) oraz modeli zakłóconych szumem o rozkładzie z różnymi parametrami σ
Źródło: opracowanie własne.

W tabeli 4 przedstawiono wyniki analizy wpływu zakłóceń losowych na proces estymacji parametrów ϕ i θ modelu mieszanego ARIMA(1,1,1).

Tabela 4. Wyniki estymacji parametrów ϕ i θ dla wygenerowanego niestacjonarnego szeregu mieszanego ARIMA(1,1,1) z $\sigma_a = 1$, $\phi = 0{,}8$ i $\theta = -0{,}5$ ($n = 400$) oraz sygnałów zakłóconych

Zakłócenie	Estymowany parametr	Wartość	Asymptotyczny błąd standardowy	Ocena istotności p	95% przedział ufności	
					od	do
Brak	ϕ	0,8316	0,0292	0,0000	0,7742	0,8890
	θ	-0,5321	0,0401	0,0000	-0,6109	-0,4533
$N(0;0{,}25)$	ϕ	0,8297	0,0309	0,0000	0,7688	0,8952
	θ	-0,2418	0,0504	0,0000	-0,3407	-0,1428
$N(0;0{,}5)$	ϕ	0,8325	0,0328	0,0000	0,7681	0,8970
	θ	0,0382	0,0546	0,4844	-0,0691	0,1454
$N(0;1)$	ϕ	0,8434	0,0361	0,0000	0,7723	0,9145
	θ	0,3662	0,0575	0,0000	-0,2531	0,4793
$N(0;2)$	ϕ	0,8422	0,0359	0,0000	0,7717	0,9126
	θ	0,3420	0,0574	0,0000	0,2290	0,4550

Źródło: opracowanie własne na podstawie obliczeń przeprowadzonych z wykorzystaniem programu Statistica.

Oszacowania parametru ϕ przy wszystkich rozpatrywanych poziomach zakłóceń mieszczą się w 95% przedziale ufności wokół przyjętego do generacji szeregu poziomu 0,8. Parametr θ od poziomu ok. –0,5 zmierza w kierunku dodatnim w miarę wzrostu wariancji zakłóceń (tab. 4).

5. Wnioski

Porównując uzyskane przez autorów wyniki badań, można stwierdzić, że reakcja niestacjonarnych szeregów czasowych na zakłócenia losowe znacznie różni się od szeregów stacjonarnych [por. Nazarko, Rybaczuk, Jurczuk 2005]. W przypadku niestacjonarnych szeregów typu ARIMA(1,1,0) kształty funkcji ACF i PACF zmieniają się już przy nieznacznych zakłóceniach na tyle, że nie pozwalają trafnie zidentyfikować typu modelu zakłócanego szeregu.

Funkcje autokorelacji ACF i PACF szeregów niestacjonarnych typu ARIMA(0,1,1) nie zmieniają się istotnie przy wszelkich badanych poziomach zakłóceń, wskazując zawsze na właściwy typ modelu. Znacznie zmieniające się jednocześnie wartości oszacowań parametru θ wskazują, iż szereg jest rozpoznawany jako tego samego typu, ale zupełnie inny ze względu na wartość parametru.

Funkcja ACF zakłócanego niestacjonarnego szeregu mieszanego typu ARIMA(1,1,1) zmienia swój kształt tylko nieznacznie przy badanych poziomach zakłóceń, funkcja PACF zaś znacznie – od zaniku składowej średniej ruchomej aż

do pojawienia się estymowanych wartości parametru θ o znaku przeciwnym niż przyjęty do generowania modelu.

Otrzymane wyniki mają swoje uzasadnienie w mechanizmach generowania procesów autoregresji i średniej ruchomej jako liniowej kombinacji impulsów losowych, a także we własności modeli ARIMA. Identyfikacja modeli ARIMA nie jest ściśle skodyfikowana i w dużej mierze opiera się na wiedzy empirycznej i intuicji badacza. Często dany szereg czasowy można opisać różnymi modelami ARIMA. Człony autoregresyjne i średniej ruchomej mogą znosić wzajemnie swoje działanie [Akaike 1974]. Istnieje także zależność między krotnością różnicowania a rzędem autoregresji i średniej ruchomej. Przeróżnicowanie szeregu może zostać skompensowane przez uwzględnienie dodatkowego członu autoregresji w modelu, a niedoróżnicowanie – przez dodatkowy człon średniej ruchomej.

W wyniku przeprowadzonych badań stwierdzono, że poprawność identyfikacji typu modelu ARIMA zależy w dużym stopniu (różnym w zależności od typu procesu) od poziomu zakłóceń losowych obecnych w szeregu. Uzyskane wyniki mogą sugerować użyteczność rozszerzenia klasycznego podejścia Boxa-Jenkinsa o etap wstępnej filtracji szeregu.

Literatura

Akaike H. (1974), *A New Look at the Statistical Model Identification*, IEEE Transaction on Automatic Control, vol. 19, nr 6, s. 716-723.

Box G.E.P., Jenkins G.M. (1983), *Analiza szeregów czasowych. Prognozowanie i sterowanie,* PWN, Warszawa.

DeLurgio S.A. (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, New York.

Hanke J.E., Reitsch A.G., Wichern D.W. (2001), *Business Forecasting,* Prentice Hall, Inc., New Jersey.

Nazarko J., Rybaczuk M., Jurczuk A. (2005), *Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA, Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 12, AE, Wrocław, s. 238-247.

INFLUENCE OF RANDOM NOISE ON AN IDENTIFICATION OF ARIMA MODELS FOR NON-STATIONARY PROCESSES

Summary

According to Box-Jenkins' methodology a key issue in time-series modelling with ARIMA models is to identify proper class and their order. The basis for the identifying is an analysis of plot of autocorrelation (ACF) and partial correlation functions (PACF). The paper presents results of the second part of a simulation

research on influence of white noise presence and its variance level on identification of ARIMA models for three types of non-stationary processes: ARIMA(1,1,0), ARIMA(0,110) and ARIMA(1,1,1). The effect of noise variance level on a quality of estimation of ARIMA models has been also presented in the paper. It has been considered that possibility of correct identification and estimation strongly depends on presence and variance of a random noise.