

Joanicjusz Nazarko, Mikołaj Rybaczuk, Arkadiusz Jurczuk
Politechnika Białostocka

WPLYW POZIOMU ZAKŁÓCEŃ LOSOWYCH NA MOŻLIWOŚĆ IDENTYFIKACJI MODELI ARIMA

1. Wstęp

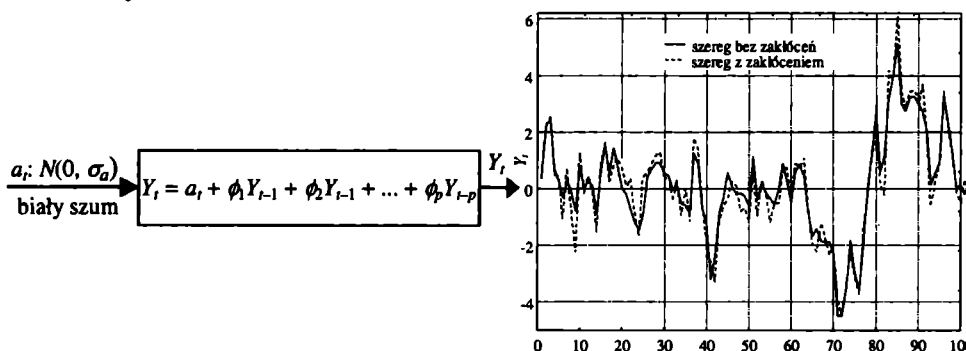
Ważną klasę modeli szeregów czasowych są modele: autoregresyjne, średniej ruchomej oraz mieszane modele autoregresyjne i średniej ruchomej. Są one stosowane do modelowania i prognozowania na podstawie stacjonarnych szeregów czasowych. Metody szacowania parametrów takich modeli szeregów czasowych są zawarte w dostępnych programach statystycznych i ekonometrycznych. W modelach ARIMA bardzo ważnym etapem analizy szeregów, poprzedzającym estymację parametrów, jest identyfikacja rodzaju i rzędu modelu przeprowadzana na podstawie wzorców generycznych oraz wiedzy eksperckiej analityka. Pomiary sygnałów stochastycznych są obarczone zwykle losowymi błędami. Poziom błędów losowych pomiaru, akceptowalny ze względu na rozpoznawalność mierzonego sygnału, zależy od amplitudy wahań sygnału i jego charakteru. Biorąc pod uwagę specyfikę modelowania szeregów metodą Boxa-Jenkinsa, postanowiono ocenić stopień wpływu zakłóceń losowych (błędów losowych pomiaru) o charakterze białego szumu o różnym poziomie wariancji na identyfikowalność postaci modeli oraz jakość estymacji parametrów szeregów: autoregresyjnych – ARIMA(1,0,0), średniej ruchomej – ARIMA(0,0,1) oraz mieszanych – ARIMA(1,0,1).

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

Model autoregresyjnego szeregu czasowego Y_t można uzyskać poprzez konwersję białego szumu a_t o rozkładzie $N(0, \sigma_a)$ za pomocą modelu podanego na rys. 1 [Box, Jenkins 1983; Hanke, Reitsch, Wichern 2001]. Na rysunku tym przedstawiono także

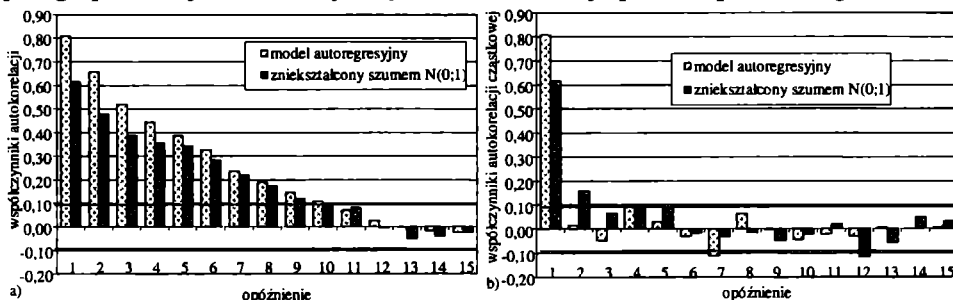
wygenerowaną dla pierwszych 100 okresów realizację stacjonarnego szeregu autoregresyjnego ARIMA(1,0,0) dla $Y_0 = 0$, $\sigma_a = 1$, $p = 1$ oraz $\phi_1 = 0,8$ ($n = 350$). Wariancja wygenerowanego pierwotnego szeregu czasowego (bez zakłóceń) wynosi 2,83.

W celu zbadania wpływu zakłóceń losowych na identyfikowalność parametru ϕ_1 przy różnym stopniu zakłócenia realizację procesu pierwotnego zakłócano szumem o różnym poziomie wariancji. Odchylenie standardowe zakłócenia zwiększano od 0 co 0,05, aż do maksymalnego zastosowanego zakłócenia $\sigma = 1,0$. Przebiegi szeregu czasowego pierwotnego oraz szeregu zakłóconego białym szumem $N(0;1,0)$ przedstawiono na rys. 1.



Rys.1. Algorytm oraz fragment wygenerowanej realizacji pierwotnego procesu autoregresyjnego ($Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$) oraz procesu zakłóconego szumem $N(0; 1,0)$
Źródło: opracowanie własne.

Na rysunku 2 przedstawiono przebiegi funkcji AC i PAC realizacji pierwotnego procesu ARIMA(1,0,0) oraz tej samej realizacji zakłóconej szumem $N(0; 1)$. Wariancja tego procesu jest 1,37 razy większa niż wariancja procesu pierwotnego.



Rys. 2. Funkcje autokorelacji ACF (a) i autokorelacji cząstkowej PACF (b) modelu autoregresyjnego bez zakłóceń ($Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$) oraz modelu zakłóconego szumem $N(0;1)$
Źródło: opracowanie własne.

Model pierwotny charakteryzuje się wygasającymi współczynnikami autokorelacji oraz wyraźnie dominującym współczynnikiem autokorelacji cząstkowej pierwszego rzędu; pozostałe mieszczą się w zasadzie w granicach błędu (rys. 2).

Przebieg funkcji autokorelacji sygnału zakłóconego nie zmienił się wyraźnie, wykres funkcji PACF sugeruje zaś, że badany szereg czasowy jest procesem autoregresyjnym rzędu drugiego – ARIMA(2,0,0) [Hanke, Reitsch, Wichem 2001]. Wpływ narastania wariancji zakłócenia na identyfikowalność rzędu i parametru modelu przedstawiono w tab. 1 oraz na rys. 3.

Zakłócenie losowe z odchyleniem standardowym większym od ok. 0,75 powoduje, że przedziały ufności szacowanego parametru dotyczące sygnału bez zakłócenia i z zakłóceniem stają się rozłączne. Oceniając punktowo – już zakłócenie o odchyleniu standardowym ok. 0,5 daje wyraźną różnicę w oszacowanych wartościach parametru ϕ_1 . Zakłócenia tego rzędu powodują, iż coraz bardziej istotny staje się parametr ϕ_2 , co sugeruje identyfikację modelu innego rzędu – ARIMA(2,0,0). Również na wykresie PACF przy zakłóceniach o odchyleniu standardowym powyżej 0,5 współczynnik autokorelacji cząstkowej rzędu drugiego przekracza wartości krytyczne.

Tabela 1. Wyniki estymacji parametru ϕ_1 dla wygenerowanego sygnału autoregresyjnego z $Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$ ($n=350$) oraz sygnałów zakłóconych

Model	Zakłócenie	Wartość estymatora	Asympt. błąd standardowy	Istotność p	95% przedział ufności	
					od	do
ARIMA (1,0,0)	brak	0,8050	0,0324	0,0000	0,7413	0,8686
	N(0;0,3)	0,7817	0,0340	0,0000	0,7149	0,8485
	N(0;0,5)	0,7439	0,0363	0,0000	0,6725	0,8152
	N(0;0,75)	0,6807	0,0396	0,0000	0,6027	0,7587
	N(0;1,0)	0,6098	0,0428	0,0000	0,5256	0,6940

Źródło: opracowanie własne za pomocą programu STATISTICA.

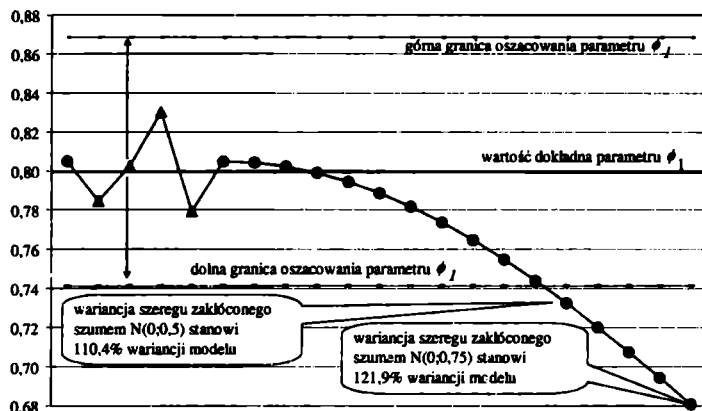
Kolejnym etapem przeprowadzonej analizy była próba oceny wpływu zaburzeń losowych na jakość prognoz otrzymanych z modeli poddanych zakłóceniom. W tabeli 2 przedstawiono oszacowane wartości prognozy na okres 351 przy różnym poziomie zakłócenia.

Tabela 2. Prognoza na okres $t = 351$ dla wygenerowanego sygnału autoregresyjnego z $Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$

ARIMA(1,0,0)	Okres t	Y_t	Prognoza Y_t	95% przedz. ufności		Błąd	
				od	do	standardowy	bezwzględny
Sygnał pierwotny	351	4,0813	2,7899	0,8231	4,7568	1,0000	1,2914
Sygnał + szum N(0;0,3)	351	4,0813	2,7267	0,6182	4,8354	1,0720	1,3546
Sygnał + szum N(0;0,5)	351	4,0813	2,6058	0,2839	4,9277	1,1806	1,4755
Sygnał + szum N(0;0,75)	351	4,0813	2,3971	-0,2737	5,0679	1,3579	1,6842
Sygnał + szum N(0;1,0)	351	4,0813	2,1587	-0,9097	5,2270	1,5601	1,9226

Źródło: opracowanie własne.

Dla szumu o odchyleniu standardowym powyżej 0,3 zwiększa się błąd standardowy i rozszerza się przedział ufności prognozy przy jednoczesnym zwiększaniu się błędu bezwzględnego prognozy.

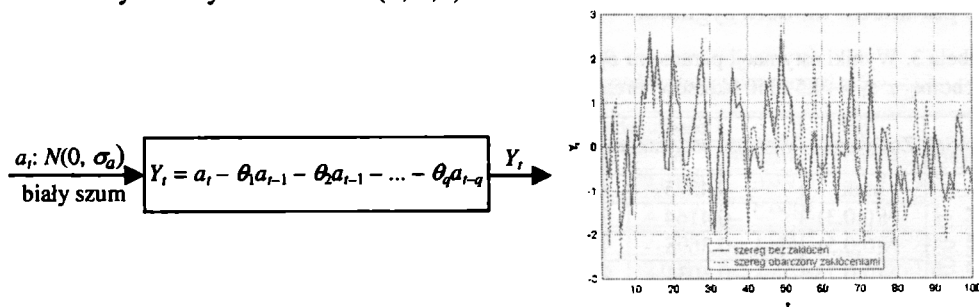


Rys. 3. Oszacowania parametru ϕ_1 metodą Mc Leoda i Salesa (STATISTICA) dotyczące różnych realizacji procesu (ozn. ▲) oraz pierwszej realizacji przy coraz silniejszych zakłóceńach (ozn. ●)
Źródło: opracowanie własne.

Przedstawione na rys. 3 wyniki symulacji ilustrują zmiany odchylania się oszacowań wartości parametru ϕ_1 od wartości dokładnej równej 0,8 w miarę jak wzrasta odchylenie standardowe szumu zakłócającego (z krokiem 0,05). Przy wzroście odchylenia standardowego zakłócenia powyżej 0,5 oszacowana wartość parametru ϕ_1 wykracza poza przedział ufności wyznaczony względem dokładnej wartości parametru modelu.

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

Model szeregu czasowego średniej ruchomej Y_t można uzyskać poprzez przekształcenie białego szumu a_t o rozkładzie $N(0, \sigma_a^2)$ za pomocą modelu podanego na rys. 4 [Box, Jenkins 1983; Hanke, Reitsch, Wichern 2001]. Na rysunku tym przedstawiono też realizację stacjonarnego szeregu średniej ruchomej dla pierwszych 100 okresów ($n = 350$) dla $\sigma_a = 1$, $q = 1$ oraz $\theta_1 = -0,5$ oraz sygnał pierwotny z nałożonym białym szumem $N(0; 1,0)$.

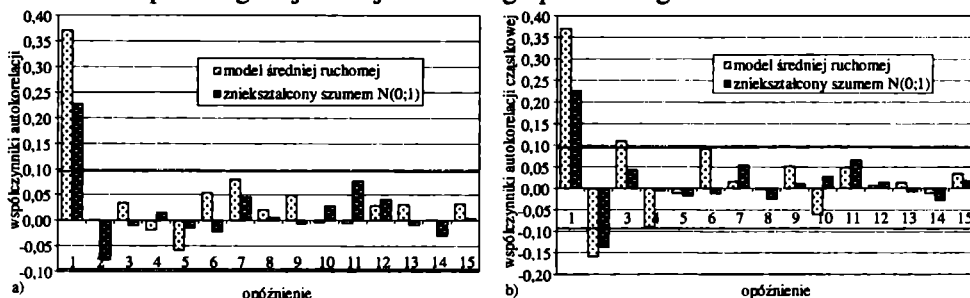


Rys. 4. Algorytm i fragment wygenerowanej realizacji pierwotnego procesu średniej ruchomej bez zakłóceń ($\sigma_a = 1$, $\theta_1 = -0,5$) oraz szeregu zakłóconego szumem $N(0; 1,0)$

Źródło: opracowanie własne.

Wygenerowany za pomocą wymienionego modelu sygnał pierwotny charakteryzuje się niewielką wariancją (1,11). Po zakłóceniu sygnału pierwotnego białym szumem z maksymalnym odchyleniem standardowym ($\sigma_a = 1$) wariancja jest 1,88 raza większa niż wariancja szeregu pierwotnego. Wyraźnie jest widoczna większa wrażliwość modelu ARIMA(0,0,1) na zakłócenia losowe niż wrażliwość modelu autoregresyjnego. Identyczny poziom szumu działającego na proces autoregresyjny zwiększał wariancję tylko 1,37 razy.

Model pierwotny charakteryzuje się wysoką wartością funkcji autokorelacji dla $k = 1$ i nieistotnymi współczynnikami wyższych rzędów. Wartości PACF zanikają naprzemiennie do zera (rys. 5). W modelu zakłóconym szumem $N(0; 1,0)$ zarysowuje się wyraźniej wartość funkcji autokorelacji dla opóźnienia $k = 2$, jednak nie wykracza ona poza wartości krytyczne; poza tym przebieg ACF nie odbiega znacznie od przebiegu tej funkcji dla szeregu pierwotnego.



Rys. 5. Funkcje autokorelacji ACF (a) i autokorelacji cząstkowej PACF (b) modelu średniej ruchomej bez zakłóceń ($\sigma_a = 1$, $\theta_1 = -0,5$) oraz szeregu zakłóconego szumem $N(0;1)$

Źródło: opracowanie własne.

Wprowadzenie zakłócenia losowego utrudnia jednoznaczne rozpoznanie przebiegu PACF jako tłumionej sinusoidy, a co za tym idzie – identyfikację rodzaju modelu. W celu sprawdzenia zasadności zwiększenia rzędu modelu oceniono istotność parametru θ_2 . Uzyskane wyniki obliczeń nie wykazały istotności dodatkowego parametru modelu sygnału.

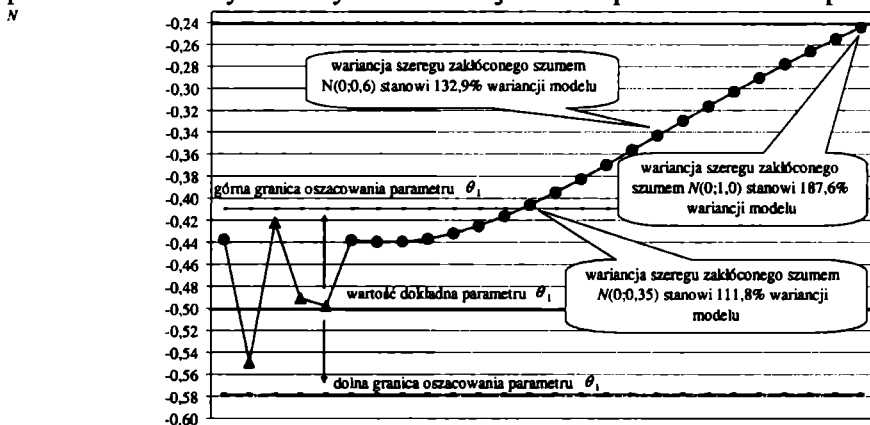
Tabela 3. Wyniki estymacji parametru θ_1 dla wygenerowanego z szumu $N(0;1)$ modelu średniej ruchomej z $\theta_1 = -0,5$ (350 okresów) oraz modeli zakłóconych

Model	Zakłócenie	Wartość estymatora	Asympt. błąd standardowy	Istotność p	95% prz.ufn.	
					od	do
ARIMA (0,0,1)	brak	-0,4375	0,0493	0,0000	-0,5344	-0,3405
	$N(0;0,3)$	-0,4164	0,0503	0,0000	-0,5153	-0,3174
	$N(0;0,5)$	-0,3698	0,0522	0,0000	-0,4724	-0,2670
	$N(0;0,75)$	-0,3031	0,0544	0,0000	-0,4100	-0,1962
	$N(0;1,0)$	-0,2441	0,0558	0,0000	-0,3538	-0,1343

Źródło: opracowanie własne.

W tabeli 3 oraz na rys. 6 przedstawiono także wpływ wzrostu wariancji zakłócenia na jakość estymacji parametru θ_1 i jednoznaczność identyfikacji rodzaju modelu. Oszacowanie wartości parametru dla szeregów zakłóconych szumem o odchyleniu standardowym, powyżej 0,5, wykraczają poza przedział ufności parametru modelu pierwotnego. Zakłócenie sygnału pierwotnego szumem o odchyleniu standardowym, nie przekraczającym 0,3 (wariancja zakłóconego szeregu jest 1,09 razy większa niż niezakłóconego modelu), nie powoduje wyraźnych zmian w poziomie szacowanych wartości parametru; większe zakłócenia prowadzą do liniowego wzrostu szacowanych parametrów.

Wykres przedstawiony na rys. 6 wskazuje, że przy odchyleniu standardowym szumu przekraczającym poziom $\sigma_a = 0,35$ wartość parametru θ_1 wykracza poza przedział ufności wyznaczony dla dokładnej wartości parametru modelu pierwotnego.



Rys. 6. Oszacowania parametru θ_1 metodą Mc Leoda i Salesa (STATISTICA) dla różnych realizacji procesu średniej ruchomej (ozn. ▲) oraz przy coraz silniejszych zakłóceniach (ozn. ●)
Źródło: opracowanie własne.

Ocenę jakości prognoz sporządzonych na okres 351 przy różnym poziomie sygnału zakłócającego przedstawiono w tab. 4.

Tabela 4. Prognoza na okres $t = 351$ dla wygenerowanego sygnału modelu średniej ruchomej z $\theta_1 = -0,5$

ARIMA(0,0,1)	Okres t	Y_t	Prognoza Y_t	95% przedz. ufności		Błąd	
				od	do	standardowy	bezwzględny
Sygnał pierwotny	351	0,7315	0,1813	-1,6871	2,0497	0,9500	0,5502
Sygnał + szum $N(0;0,3)$	351	0,7315	0,2044	-1,7588	2,1676	0,9982	0,5271
Sygnał + szum $N(0;0,5)$	351	0,7315	0,2252	-1,8993	2,3498	1,0802	0,5063
Sygnał + szum $N(0;0,75)$	351	0,7315	0,2299	-2,1713	2,6311	1,2208	0,5016
Sygnał + szum $N(0;1,0)$	351	0,7315	0,2132	-2,5196	2,9460	1,3895	0,5183

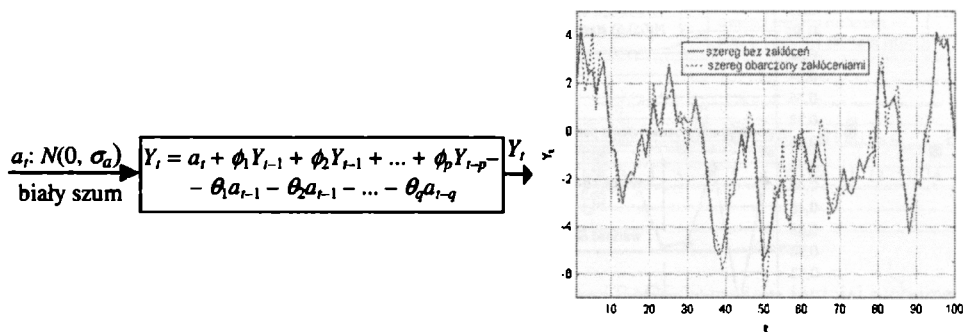
Źródło: opracowanie własne.

Analiza wyników wskazuje, że wraz ze wzrostem wielkości zakłócenia zwiększa się błąd standardowy prognozy a wraz z nim rozszerza się przedział ufności progno-

zowanej wartości. Błąd bezwzględny prognozy rośnie dopiero na poziomie maksymalnego stosowanego zniekształcenia.

4. Mieszany model autoregresyjny i średniej ruchomej – ARIMA(1,0,1)

Na rysunku 7 przedstawiono przekształcenie [Box, Jenkins 1983; Hanke, Reitsch, Wichern 2001], za pomocą którego z białego szumu a_t o rozkładzie $N(0, \sigma_a)$ zbudowano proces autoregresji i średniej ruchomej ARIMA(1,0,1). Realizacje procesu uzyskiwano, przyjmując $Y_0 = 0$, $\sigma_a = 1$, $p = 1$, $q = 1$, $\phi_1 = 0,8$, $\theta_1 = -0,5$ ($n = 350$). Na rysunku 7 przedstawiono także graficzną ilustrację przebiegu tego procesu dla 100 początkowych okresów. Model mieszany pierwotny charakteryzuje się stosunkowo dużą wariancją (5,12) w porównaniu z modelami autoregresyjnym i średniej ruchomej.



Rys. 7. Algorytm i fragment wygenerowanej realizacji pierwotnego procesu mieszanego bez zakłóceń ($Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$, $\theta_1 = -0,5$) oraz szeregu zakłóconego szumem $N(0; 1,0)$

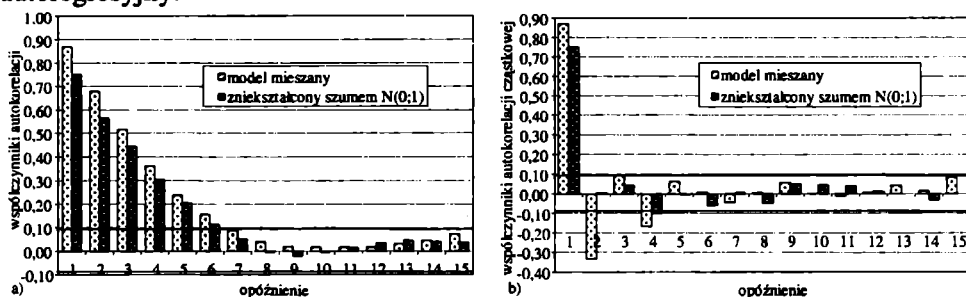
Źródło: opracowanie własne.

Podobnie jak w dwóch poprzednich modelach, proces ARIMA(1,0,1) zakłócono białym szumem. Maksymalny szum $N(0; 1,0)$ stosowany w badaniach powoduje, że wariancja sygnału zniekształconego jest jedynie 1,16 raza większa niż wariancja sygnału pierwotnego. Można zatem stwierdzić, że przy takim poziomie wariancji jest on mało wrażliwy na zakłócenia losowe w porównaniu z modelami AR(1) i MA(1).

Przy dodatnim parametrze ϕ_1 oraz ujemnym θ_1 proces mieszany charakteryzuje się dość szybkim spadkiem funkcji autokorelacji ACF i naprzemiennym szybkim spadkiem funkcji autokorelacji cząstkowej PACF (rys. 8) [Hanke, Reitsch, Wichern 2001].

Szereg zakłócony szumem $N(0;1,0)$ ma funkcje ACF i PACF charakterystyczne dla modelu autoregresyjnego (rys. 8) [Hanke, Reitsch, Wichern 2001]. Obserwacja funkcji autokorelacji i autokorelacji cząstkowej wykazała, że w miarę nasilania się zakłócenia zanikają cechy modelu mieszanego. Powyżej poziomu zakłócenia $N(0; 0,6)$ funkcja AC charakteryzuje się wykładniczym zanikaniem wartości do zera, a

funkcja PAC ma jeden dodatni impuls przy $k = 1$. Sugeruje to, że modelem zidentyfikowanym jako adekwatny w sytuacji tak zakłóconego sygnału może być model autoregresyjny.



Rys. 8. Funkcje autokorelacji ACF (a) i autokorelacji cząstkowej PACF, (b) modelu mieszanego bez zakłóceń ($Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$, $\theta_1 = -0,5$) oraz szeregu zakłóconego szumem $N(0; 1)$

Źródło: opracowanie własne.

Tabela 5. Wyniki estymacji parametrów ϕ_1 oraz θ_1 dla wygenerowanego modelu mieszanego z $Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$, $\theta_1 = -0,5$ (350 okresów) oraz modeli zakłóconych

Model	Zakłócenie	Wartość estymatora	Asympt. błąd standardowy	Istotność p	95% przedział ufności	
ARIMA (1,0,1)	brak	$\phi_1: 0,7876$	0,0370	0,0000	0,7148	0,8605
		$\theta_1: -0,4419$	0,0589	0,0000	-0,5578	-0,3260
	N(0;0,3)	$\phi_1: 0,7771$	0,0393	0,0000	0,6998	0,8545
		$\theta_1: -0,3838$	0,0659	0,0000	-0,5133	-0,2542
	N(0;0,5)	$\phi_1: 0,7802$	0,0408	0,0000	0,6999	0,8606
		$\theta_1: -0,2278$	0,0700	0,0013	-0,3655	-0,0901
	N(0;0,6)	$\phi_1: 0,7817$	0,04158	0,0000	0,6999	0,8635
		$\theta_1: -0,1531$	0,0710	0,0318	-0,2928	-0,0134
	N(0;0,65)	$\phi_1: 0,7822$	0,0420	0,0000	0,6997	0,8648
		$\theta_1: -0,1185$	0,0714	0,0982	-0,2590	0,0221
	N(0;1,0)	$\phi_1: 0,7816$	0,0455	0,0000	0,6921	0,8711
		$\theta_1: 0,0753$	0,0746	0,3137	-0,0715	0,2220

Źródło: opracowanie własne.

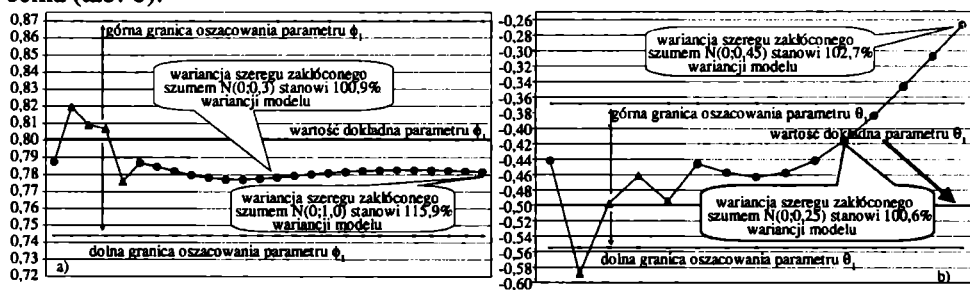
Zarówno w tab. 5, jak i na rys. 9 widać, iż składowa autoregresyjna modelu mieszanego nie ulega zniekształceniu nawet przy maksymalnym stosowanym szumie zakłócającym $N(0; 1,0)$. Zarówno estymowane wartości parametru ϕ_1 , błędy standardowe oszacowań i zakresy przedziałów ufności dla parametru ϕ_1 kształtują się na porównywalnym poziomie, niezależnie od stopnia zniekształcenia modelu. Parametr θ_1 przy zniekształceniu większym niż szumem o odchyleniu standardowym 0,6 nie różni się istotnie od zera. Model zachowuje własności modelu mieszanego, oszacowania parametrów są zaś zbliżone do poziomu parametrów używanych przez model czysty przy zakłóceniach do $N(0; 0,2)$. Powyżej tego poziomu zakłócenia model szybko przekształca się w model autoregresyjny.

Tabela 6. Prognoza na okres $t = 351$ dotycząca wygenerowanego modelu mieszanego z $Y_0 = 0$, $\sigma_a = 1$, $\phi_1 = 0,8$, $\theta_1 = -0,5$

ARIMA(1,0,1)	Okres t	Y_t	Prognoza Y_t	95% przedz. ufności		Błąd	
				od	do	standardowy	bezwzględny
Sygnał pierwotny	351	-0,0262	-1,4074	-3,0666	0,2518	1,0060	1,3812
Sygnał + szum $N(0;0,3)$	351	-0,0262	-1,2135	-2,9862	0,5592	1,0749	1,1873
Sygnał + szum $N(0;0,5)$	351	-0,0262	-0,8580	-2,8429	1,1268	1,2035	0,8318
Sygnał + szum $N(0;0,75)$	351	-0,0262	-0,5609	-2,8628	1,7409	1,3957	0,5347
Sygnał + szum $N(0;1,0)$	351	-0,0262	-0,3833	-3,0264	2,2597	1,6026	0,3571

Źródło: opracowanie własne.

Wraz ze wzrostem wielkości zakłócenia zwiększa się błąd standardowy prognozy. Wraz z nim rozszerza się przedział ufności prognozowanej wartości, a błąd bezwzględny prognozy maleje nawet do poziomu maksymalnego stosowanego zniekształcenia (tab. 6).



Rys. 9. Oszacowania parametru ϕ_1 (a) oraz θ_1 (b) metodą McLeoda i Salesa (STATISTICA) dla wygenerowanego modelu mieszanego przy różnych białych szumach (ozn. ▲) oraz z coraz silniejszymi zakłóceniami (ozn. ●)

Źródło: opracowanie własne.

Na rysunku 9 pokazano odchylenie się oszacowań wartości parametrów ϕ_1 od dokładnej wartości równej 0,8 oraz θ_1 od wartości dokładnej równej -0,5 w miarę wzrostu odchylenia standardowego szumu zakłócającego co 0,05.

5. Wnioski

W wyniku obliczeń symulacyjnych stwierdzono, że możliwości poprawnej identyfikacji modeli ARIMA, opisujących szeregi czasowe, zależą od poziomu zakłóceń losowych szeregu. Jest pewien graniczny poziom udziału wariancji zakłóceń w wariancji szeregu, przy którym wartości estymowanych parametrów modelu wykraczają poza przedziały ufności względem dokładnych wartości parametrów. W miarę wzrostu wariancji zakłóceń rośnie ryzyko błędnej identyfikacji rzędu i rodzaju modelu. Obecność zakłóceń wpływa też na błąd standardowy i szero-

kość przedziału ufności prognozy. Wraz ze wzrostem poziomu zakłóceń wzrastają też błędy bezwzględne prognozy.

Literatura

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

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

INFLUENCE OF RANDOM NOISE ON AN IDENTIFICATION OF ARIMA MODELS

Summary

The main problem of time series modelling with ARIMA models is to identify class (autoregression, moving average, autoregression and moving average) and their order. The basis for the identifying is an analysis of plot of autocorrelation (AC) and partial correlation functions (PAC). The paper presents results of simulation research on influence of white noise presence and its variance level on identification of basic ARIMA models. 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.