

Grażyna Trzpiot

Uniwersytet Ekonomiczny w Katowicach

ODPORNĄ ANALIZĄ SZEREGÓW CZASOWYCH

Streszczenie: Obszar odpornych analiz szeregów czasowych pojawiał się w badaniach statystycznych w ostatnich latach, ostatnie dekady to największa aktywność na tym polu badawczym. Przedstawimy najbardziej powszechne modele szeregów czasowych, następnie typy obserwacji odstających w szeregach czasowych oraz odporną estymację szeregów czasowych. Statystyczny termin „odporny” wykorzystywany jest w znaczeniu odpornej estymacji współczynnika kierunkowego liniowej funkcji trendu.

Słowa kluczowe: szeregi czasowe, obserwacje odstające, odporna estymacja.

1. Model ARMA (Autoregressive Moving Average Model)

Szereg czasowy jest ciągiem n zmiennych losowych Y_1, Y_2, \dots, Y_n obserwowanych w jednakowych odstępach czasu. Często obserwacje możemy zapisać jako model:

$$Y_t = \mu + X_t \quad t = 1, 2, \dots, n, \quad (1)$$

gdzie μ jest parametrem położenia, a X_t wyznacza model (*Autoregressive Moving Average Model*) o średniej zero zapisywany jako ARMA(p, q). To oznacza, że prawdziwe jest równanie [1].

$$X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} = e_t - \beta_1 e_{t-1} - \dots - \beta_q e_{t-q}, \quad (2)$$

gdzie e_t ma rozkład $N(0, \sigma^2)$. Nieznanymi parametrami są: μ, σ oraz

$$\alpha = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_p \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_q \end{bmatrix}.$$

Lewa strona równania nazywana jest *częścią autoregresyjną*, ponieważ X_t zależy od wcześniejszych wartości X_{t-1}, \dots, X_{t-p} . Prawa strona równania nazywana jest czę-

ściąg średniej ruchomej, ponieważ aktualny błąd w czasie t jest liniową kombinacją wyjściowych $e_t, e_{t-1}, \dots, e_{t-q}$, które nie mogą być obserwowane bezpośrednio.

Model ARMA(p, q) jest trudny do estymacji, często wystarczające jest wykorzystanie modelu AR(p), który jest prosty, ponieważ zawiera jedynie część autoregresyjną. Taki model otrzymujemy, przyjmując $q = 0$

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + e_t, \quad (3)$$

Wykorzystując bieżące obserwacje, $Y_t = \mu + X_t$ można zapisać

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \dots + \alpha_p (Y_{t-p} - \mu) + e_t,$$

a po przekształceniu

$$Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \gamma + e_t, \quad (4)$$

gdzie $\gamma = \mu(1 - \alpha_1 - \dots - \alpha_p)$.

Mamy $n-p$ zestawów zbiorów zmiennych losowych $Y_t, Y_{t-1}, \dots, Y_{t-p}$ rangowanych względem t (od $p+1$ do n), które mogą być wykorzystane do estymacji $\alpha_1, \dots, \alpha_p, \gamma$, w znaczeniu estymacji współczynników regresji.

Prostszy, ale często wykorzystywany jest model AR(1)

$$Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \gamma + e_t. \quad (5)$$

Dla spełnienia warunku stacjonarności Y_t wartość bezwzględna α_1 musi być mniejsza niż 1.

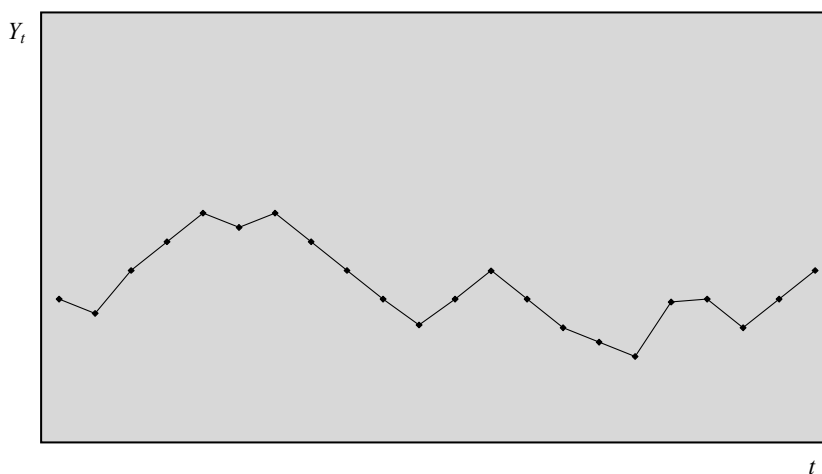
2. Obserwacje odstające w szeregach czasowych

Odporne obserwacje w szeregach czasowych mają różny charakter. Mogą to być jednorazowe obserwacje nietypowe stanowiące istotne odchylenie od przewidywanej wartości badanego zjawiska tylko w jednym okresie. Możemy też wskazać takie, które powodują zmianę procesu generującego dane zjawisko i zmianę trendu. Fox [2] oraz Martin [5] rozważali następujące rodzaje obserwacji odstających, które mogą występować w szeregach czasowych.

Pierwszą klasę tworzą **innowacyjne obserwacje odstające** (*innovation outliers*), które mogą być modelowane z wykorzystaniem wyjściowego modelu ARMA, z wyjątkiem jednego założenia: rozkład e_t nie jest normalny, charakteryzuje się grubymi ogonami.

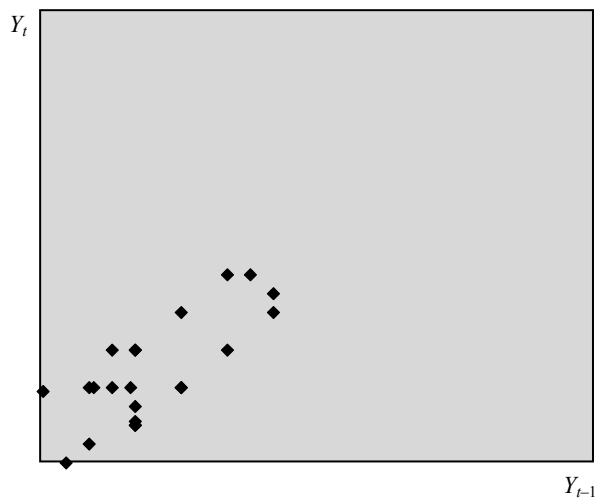
To implikuje, że szczególny obserwowalny rozkład e_{t_0} , który jest odmienny, odstający od innych obserwacji w szeregu (*outlying*) ma wpływ na X_{t_0} , X_{t_0+1} oraz X_{t_0+2} itd. Jednocześnie należy dodać, że jest to efekt wygasający.

Rysunki 3 i 4 przedstawiają wykresy analogiczne do rys. 1 i 2, ale pojawia się dodatkowo innowacyjna obserwacja odstająca w czasie t_0 . Spoglądając na wykres rozrzutu na rys. 4, zauważamy, że innowacyjna obserwacja odstająca pojawia się w punkcie $A = (Y_{t_0-1}, Y_{t_0})$, to jest obserwacja odstająca w kierunku pionowym. Obser-



Rys. 1. Wykres szeregu czasowego: brak obserwacji odstających

Źródło: [11].

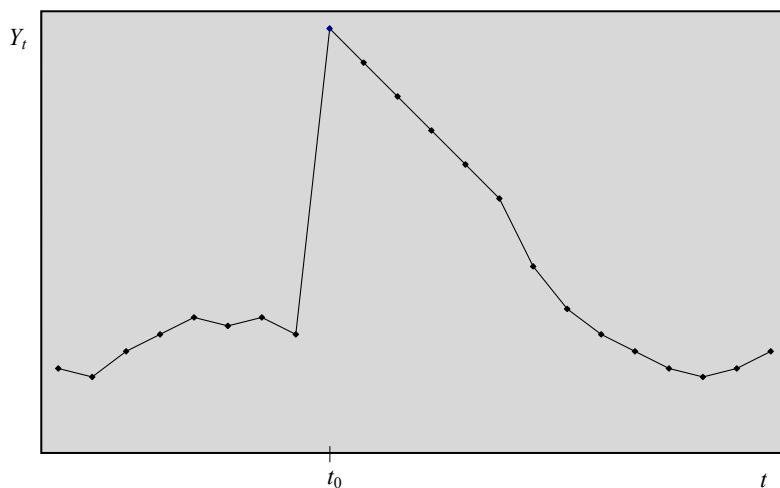


Rys. 2. Wykres rozrzutu Y_t względem Y_{t-1} : brak obserwacji odstających

Źródło: opracowanie własne na podstawie [11].

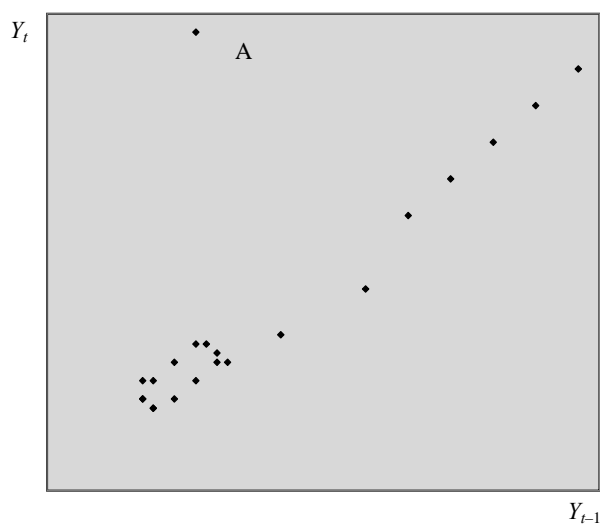
wujemy kolejne duże wartości (Y_{t-1} , Y_t), które leżą blisko wyjściowego kierunku opisu zmian liniowych oraz wyjściowego współczynnika kierunkowego α_1 . Na wykresie rozrzutu, innowacyjna obserwacja odstająca ma zatem wpływ na jedną wartość zmiennej objaśniającej oraz na pewien ciąg „dobrych” wartości wpływowych, które mają szansę poprawić dokładność estymacji współczynnika regresji α_1 . Kiedy

się stosuje metodę najmniejszych kwadratów, te „dobre” obserwacje mogą skompensować efekt obserwacji odstającej. W rzeczywistości, jeżeli rozkład dystrybucyjny e_t o grubych ogonach ma własność symetrii, to możemy uzyskać estymatory α_1 o większej precyzji niż w przypadku normalnego rozkładu e_t .



Rys. 3. Wykres szeregu czasowego: innowacyjna obserwacja odstająca w czasie t_0

Źródło: opracowanie własne na podstawie [11].



Rys. 4. Wykres rozrzutu Y_t względem Y_{t-1} : innowacyjna obserwacja odstająca w czasie t_0

Źródło: opracowanie własne na podstawie [11].

Drugi rodzaj obserwacji odstających nazywamy addytywnymi obserwacjami odstającymi (*additive outliers*). Takie obserwacje odstające pojawiają się, jeżeli obserwujemy pewne wartości, traktowane jako zaburzenie, dodane do każdej zmiennej losowej Y_t . W rezultacie obserwacje nie mogą być zapisane z wykorzystaniem modelu ARMA.

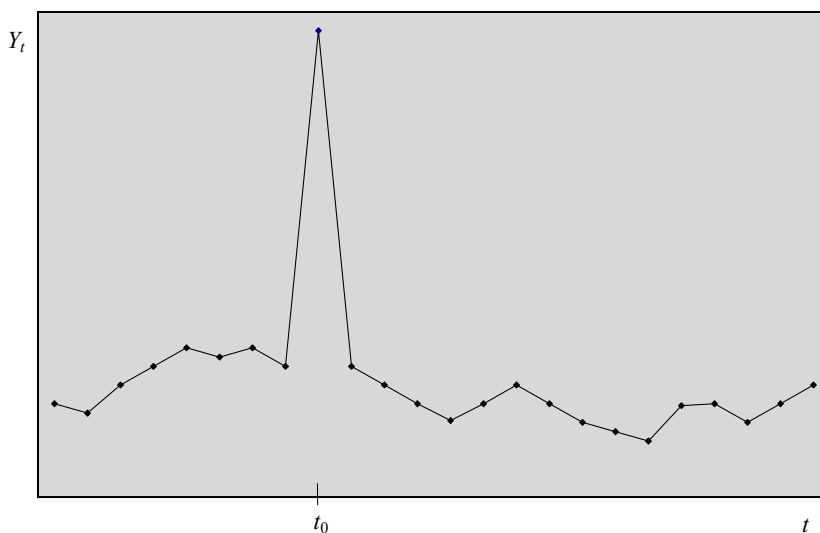
Formalnie

$$Y_t = \mu + X_t + V_t. \quad (6)$$

Dla większości obserwacji zachodzi $V_t = 0$, ale zakładamy, że $P(V_t \neq 0) = \varepsilon$, gdzie ta frakcja addytywnych obserwacji odstających jest dodatnia i nieduża. To jest sytuacja analogiczna do zaburzenia zastępowalnego, które jest podstawowym opisem dla prób skończonych z wykorzystaniem podejścia punktu załamania. Martin [5] proponuje, aby szeregi czasowe zachowywały (w praktyce) obserwacje odstające, dla których wartość ε mieści się pomiędzy 1 a 25%.

Zakładamy, że V_t mają takie same rozkłady i są niezależnymi zmiennymi losowymi, które generują model „izolowanych” lub „rozproszonych” obserwacji odstających, pojawiających się niezależnie. Z drugiej strony szczególna zależność od struktury V_t może być wykorzystana do opisu obserwacji odstających.

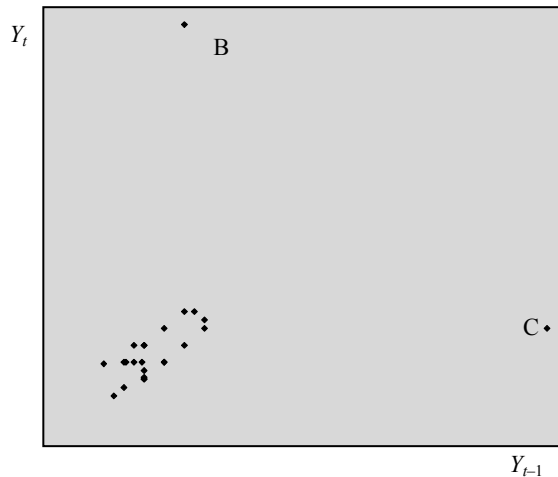
Rysunki 5 i 6 przedstawiają ilustrację izolowanych addytywnych obserwacji odstających, które mogą być spowodowane przez wyjątkowe wydarzenie, ale też mogą być rezultatem błędu zapisu lub pewnym wyjątkowym zachowaniem instrumentu pomiarowego. Pojedyncza obserwacja odstająca Y_{t_0} pojawia się w postaci *dwóch* nie-



Rys. 5. Wykres szeregu czasowego: izolowane addytywne obserwacje odstające

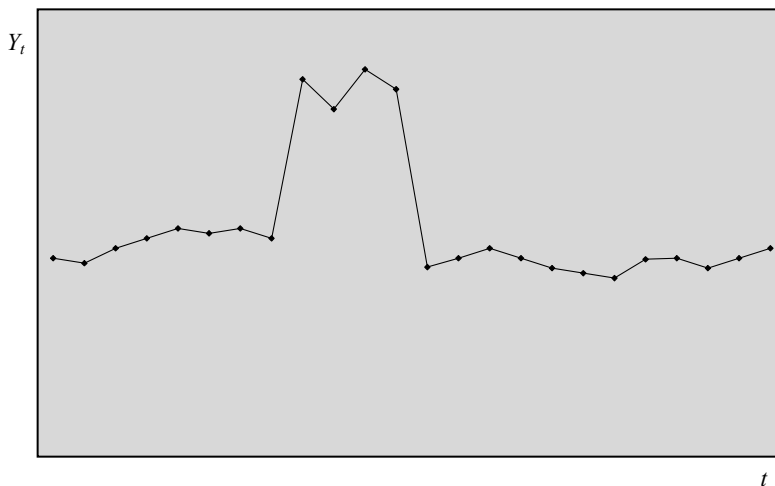
Źródło: opracowanie własne na podstawie [11].

zależnych punktów na wykresie rozrzutu: punkt $B = (Y_{t-1}, Y_t)$, który jest wartością odstającą zmiennej objaśniającej oraz punkt $C = (Y_t, Y_{t+1})$, który jest „złym” punktem wpływowym. W modelu $AR(p)$ otrzymujemy jedną obserwację odstającą w kierunku pionowym oraz p punktów wpływowych. Addytywne obserwacje odstające mają większe znaczenie w badanym szeregu czasowym niż innowacyjne obserwacje odstające, ponieważ wykorzystywana do estymacji metoda najmniejszych kwadra-



Rys. 6. Wykres rozrzutu Y_t względem Y_{t-1} : izolowane addytywne obserwacje odstające

Źródło: opracowanie własne na podstawie [11].

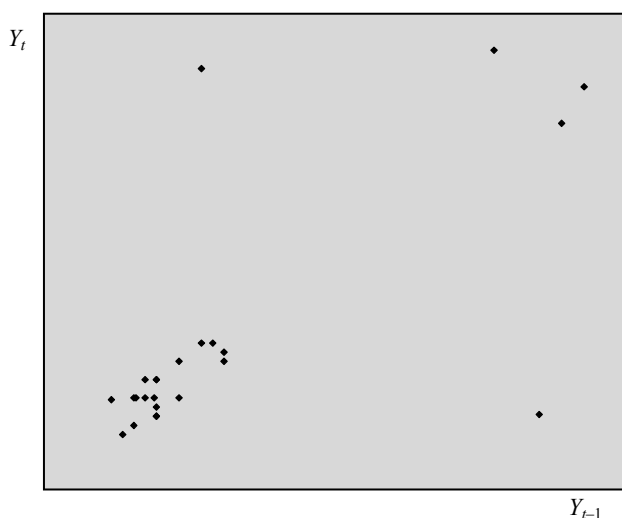


Rys. 7. Wykres szeregu czasowego: grupa addytywnych obserwacji odstających

Źródło: opracowanie własne na podstawie [11].

tów nie jest odporna na punkty wpływowe, a wyznaczone estymatory są obciążone. W przypadku szeregu zaprezentowanego na rys. 5 i 6 estymacja metodą najmniejszych kwadratów wyznacza estymatory z obciążeniem zmiernym do zera [5].

Na rysunkach 7 i 8 przedstawiamy grupę addytywnych obserwacji odstających, co nie jest rzadkością w badaniach empirycznych. Przykładowo, tygodniowa sprzedaż może dodatkowo być zmniejszona (zaburzona) przez miesięczny rabat. Inny przykład to produkcja przemysłowa, która czasowo jest zmniejszona (zaburzona) przez strajk. Rezultat, który widzimy na wykresie rozrzutu, jest niekorzystny, ponieważ estymacja punktów wpływowych metodą najmniejszych kwadratów wyznacza estymatory z obciążeniem zmiernym do jedności.



Rys. 8. Wykres rozrzutu Y_t względem Y_{t-1} : grupa addytywnych obserwacji odstających

Źródło: opracowanie własne na podstawie [11].

Addytywne obserwacje odstające pojawiają się w badaniach częściej niż innowacyjne obserwacje odstające, powodują poważniejsze problemy przy estymacji modelu, ponieważ metoda najmniejszych kwadratów, wykorzystywana do estymacji zadań regresji, nie jest odporna na punkty wpływowe. Takie zadania wymagają odpornych analiz szeregów czasowych.

3. Wykorzystanie M - oraz GM -estymatorów w szeregach czasowych

Jeżeli model z procesem ARMA zawiera jedynie innowacyjne obserwacje odstające, to wykorzystujemy M -estymatory stosowane w odpornej analizie regresji [3; 7; 11]. Ograniczmy zatem zadanie do modelu $AR(p)$ (zapisanego jako równanie (4)). W tym przypadku klasyczna metoda najmniejszych kwadratów odpowiada zadaniu:

$$\min_{\hat{\gamma}, \hat{\alpha}} \sum_t (Y_t - \hat{\gamma} - \hat{\alpha}_1 Y_{t-1} - \dots - \hat{\alpha}_p Y_{t-p})^2. \quad (7)$$

Zadanie powyższe możemy uogólnić i zapisać problem wyznaczenia M -estymatora jako:

$$\min_{\hat{\gamma}, \hat{\alpha}} \sum_t \rho \left(\frac{Y_t - \hat{\gamma} - \hat{\alpha}_1 Y_{t-1} - \dots - \hat{\alpha}_p Y_{t-p}}{\hat{\sigma}} \right)^2, \quad (8)$$

gdzie $\hat{\sigma}$ jest odpornym estymatorem parametru skali.

Estymator lokalny parametru położenia μ jest wyznaczany jako:

$$\hat{\mu} = \frac{\hat{\gamma}}{1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p}. \quad (9)$$

Jeżeli spełnione są klasyczne założenia modelu, jest to estymator zgodny o rozkładzie asymptotycznie normalnym. Wyznaczone ograniczenie Rao-Cramera oraz odporną efektywność znajdziemy w pracy Martina [6].

W ogólnym modelu ARIMA(p, q) klasyczna metoda najmniejszych kwadratów odpowiada zadaniu

$$\min_{\hat{\gamma}, \hat{\alpha}, \hat{\beta}} \sum_t r_t (\hat{\gamma}, \hat{\alpha}, \hat{\beta})^2, \quad (10)$$

gdzie: r_t – reszty modelu,

$\hat{\gamma}, \hat{\alpha}, \hat{\beta}$ – estymatory parametrów modelu ARIMA(p, q).

Reszty r_t są nieliniową funkcją parametrów, więc mamy problem nieliniowej estymacji w klasycznym zadaniu estymacji. W przypadku innowacyjnych obserwacji odstających Martin [5] proponuje zastąpić powyższe zadanie (10) przez zadanie z wykorzystaniem M -estymatorów

$$\min_{\hat{\gamma}, \hat{\alpha}, \hat{\beta}} \sum_t \rho \left(\frac{r_t (\hat{\gamma}, \hat{\alpha}, \hat{\beta})}{\hat{\sigma}} \right)^2. \quad (11)$$

dla pewnej funkcji ρ , gdzie $\hat{\sigma}$ jest estymatorem parametru skali, a następnie estymować μ z wykorzystaniem estymatora (9). Otrzymujemy asymptotyczne własności analogiczne jak w przypadku zadań autoregresji [4].

Zapisane M -estymatory nie są odporne w przypadku addytywnych obserwacji odstających. Wówczas należy wykorzystać GM -estymatory, ponieważ zapewniają dodatkowo ograniczenie skali wpływu odstających obserwacji wpływowych, poprzez wprowadzenie pewnej funkcji wagowej w . M -estymatory zapisane powyżej zastępujemy GM -estymatorami [10].

4. Podsumowanie

W analizie szeregów czasowych modele stochastyczne stosowane są bardzo często. Modele regresji są również wykorzystywane w analizie ekonomicznych szeregów czasowych, a zwłaszcza danych finansowych. Model czynnikowy ma zastosowanie w analizie portfelowej oraz w opisie ryzyka tego portfela. Podejścia te często są wykorzystywane wspólnie do pełniejszego opisu zależności (przykładowo ryzyko specyficzne i ryzyko systematyczne), a zwłaszcza do badania szeregu skorelowanych reszt w modelu wielowymiarowej regresji. W rzeczywistych szeregach czasowych nie unikniemy obserwacji odstających. Metoda największej wiarygodności, tak jak metoda najmniejszych kwadratów, również nie wyznacza estymatorów odpornych parametrów estymowanych modeli. Wykorzystanie podejścia odpornego w estymowaniu [7; 9; 10] oraz testowaniu [8] parametrów modeli szeregów czasowych jest narzędziem badań w ekonomii.

Literatura

- [1] Box G., Jenkins G., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco 1976.
- [2] Fox A.J., *Outliers in times series*, „Journal of Royal Statistic Society” 1972, vol. 34, s. 350-363.
- [3] Hampel F., Ronchetti E., Rousseeuw P., Stahel W., *Robust Statistics: The Approach Based On Influence Functions*, Wiley, New York 1986.
- [4] Lee C.H., Martin R.D., *M-Estimates for ARMA Process*, Technical Report, University of Washington, 1982.
- [5] Martin R.D., *Robust methods for times series analysis*, Applied Time Series Analysis, Academic Press, New York 1981, s. 683-759.
- [6] Martin R.D., *The Cramer Rao bound and robust M-Estimates for autoregressions*, „Biometrika” 1982, vol. 69, s. 437-442.
- [7] Trzpiot G., *Extreme value distributions and robust estimation*, Acta Universitatis Lodziensis, Folia Economica 228, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2009, s. 85-92.
- [8] Trzpiot G., *Some properties of the robust trend tests*, Acta Universitatis Lodziensis, Folia Economica, Wydawnictwo Uniwersytetu Łódzkiego Łódź 2010 [w druku].
- [9] Trzpiot G. (red.), *Wielowymiarowe metody statystyczne w analizie ryzyka inwestycyjnego*, PWE, Warszawa 2010.
- [10] Trzpiot G., Majewska J., *Estimation of value at risk: Extreme value and robust approaches*, „Operation Research and Decisions”, vol. 20, no. 1, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2010, s. 131-143.
- [11] Yohai V.J., *High breakdown point and high efficiency robust estimates for regression*, „The Annals of Statistics” 1987, vol. 15, s. 642-656.

ROBUST TIME SERIES ANALYSIS

Summary: The field of robust time series analysis has come into existence in the last decades. We describe the most common used time series models together with main type of outliers in time series. The last part of the paper presents application of well known *M*- and *GM*-estimators for time series analysis.