

Agnieszka Orwat
Uniwersytet Śląski w Katowicach
e-mail: agaorwat@poczta.onet.pl, aorwat@ux2.math.us.edu.pl

PRZYKŁAD ZASTOSOWANIA METODY ODPORNEJ W MODELOWANIU FINANSOWYCH SZEREGÓW CZASOWYCH

Streszczenie: Założenia, na których opierają się parametryczne metody estymacji dotyczące normalności rozkładu populacji oraz niezależności zmiennych, często nie są spełnione w przypadku danych finansowych. Dlatego istotną kwestią jest stosowanie metod estymacji odpornej (*robust estimation*) na te założenia, a tym samym na jakość obserwacji.

Celem pracy jest implementacja metody odpornej do modelowania szeregu czasowego wartości indeksu giełdowego WIG20 o piętnastominutowej częstotliwości notowań w dniu 13.02.2006. Do szacowania parametrów strukturalnych modelu zastosowano estymatory odporne Welscha.

Słowa kluczowe: obserwacje odstające, obserwacje nietypowe, obserwacje wpływowe, estymacja odporna, estymatory Welscha.

WSTĘP

Założenia, na których opierają się parametryczne metody estymacji, dotyczące normalności rozkładu populacji oraz niezależności zmiennych, często nie są spełnione w przypadku danych finansowych. Dlatego istotną kwestią jest stosowanie metod estymacji odpornej (*robust estimation*) na te założenia, a tym samym na jakość obserwacji.

Čzęsto w zbiorze danych można wyróżnić wartości wyraźnie różniące się pewnymi cechami od pozostałych, tzw. obserwacje odstające. Występowanie takich wartości może być zarówno wynikiem błędnych pomiarów jak również pochodzenia z innej populacji. Wpływ obserwacji odstających na wyniki szacowania parametrów klasyczną metodą najmniejszych kwadratów (KMNK) jest istotny.

Wykrywanie obserwacji odstających oraz zapobieganie ich wpływowi poprzez stosowanie metod estymacji odpornej stanowi ważny problem analizy statystyczno-ekonometrycznej. Znaczna grupa metod odpornych polega na odrzuceniu obserwacji uważanych za nietypowe, a następnie na estymowaniu parametrów na podstawie pozostałych obserwacji. Usunięcie obserwacji odstającej w przypadku trendu oznacza zaburzenie kolejności czasowej. Dlatego, w celu szacowania parametrów modelu szeregu czasowego zawierającego obserwacje odstające, zastosowana zostanie metoda odporna, oparta na przyporządkowywaniu tym obserwacjom wag zaproponowanych przez Welscha. Konsekwencją takiego

podejścia jest zastosowanie uogólnionej metody najmniejszych kwadratów (UMNK).

DIAGNOSTYKA TYPÓW OBSERWACJI

Rozważmy ogólny model postaci:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1.1)$$

gdzie: \mathbf{y} jest n wymiarowym wektorem losowym, którego wartości są niezależnymi obserwacjami zmiennej objaśnianej, \mathbf{X} jest $n \times (k+1)$ wymiarową macierzą obserwacji będących wartościami k niezależnych zmiennych objaśniających modelu, $\boldsymbol{\beta}$ jest $(k+1)$ wymiarowym wektorem nieznanych parametrów strukturalnych modelu.

Zakłada się, że składnik losowy $\boldsymbol{\varepsilon}$ ma n -wymiarowy rozkład normalny z wartością oczekiwaną $E(\boldsymbol{\varepsilon}) = 0$ oraz macierzą kowariancji Σ postaci: $\Sigma = Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$, gdzie \mathbf{I} jest macierzą jednostkową o wymiarach $n \times n$.

Wyniki zastosowania określonych metod estymacji modelu (1.1) w istotny sposób zależą od jakości danych. Na i -tą daną z_i składają się w tym przypadku obserwacje będące wartościami k zmiennych objaśniających oraz zmiennej objaśnianej \mathbf{y} , tzn:

$$z_i = [x_{i1}, \dots, x_{ik}, y_i], \quad i = 1, \dots, n \quad (1.2)$$

Obserwacje z_i mogą różnić się zarówno wpływem na wartość ocen estymowanych parametrów, jak również położeniem w przestrzeni R^{k+1} . W konsekwencji, w zbiorze danych mogą występować tzw. obserwacje odstające, wyraźnie różniące się pewnymi cechami od pozostałych. Zbiór obserwacji, utworzony po usunięciu obserwacji odstającej nazywany jest rdzeniem.

Na obserwacje odstające składają się dwa rodzaje obserwacji, mianowicie obserwacje nietypowe (*outliers*) oraz obserwacje wpływowe (*influential observations*).

Obserwacje nietypowe charakteryzują się większymi wartościami reszt niż obserwacje należące do rdzenia. Natomiast obserwacje wpływowe silnie wpływają na wartości estymatorów parametrów oraz istotnie oddziałują na kształt modelu. Wartości reszt tych obserwacji nie są duże w stosunku do wartości reszt obserwacji należących do rdzenia.

Wykrywanie obserwacji odstających może być często problemem złożonym, zwłaszcza w przypadku wielowymiarowym. Zastosowanie odpowiednich narzędzi służących identyfikacji tego rodzaju obserwacji stanowi punkt wyjścia w stosowaniu metod estymacji odpornej.

WYKRYWANIE OBSERWACJI WPŁYWOWYCH

Podstawowym narzędziem służącym wykrywaniu obserwacji wpływowych jest macierz rzutowania \mathbf{H} (*hat matrix*) o wymiarach $n \times n$:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (1.1.1)$$

\mathbf{H} jest macierzą rzutu ortogonalnego przestrzeni R^n na podprzestrzeń $R(X)$ generowaną przez kolumny macierzy \mathbf{X} . Elementy diagonalne macierzy rzutowania, oznaczane w skrócie jako h_i , noszą nazwę wielkości wpływowych (*leverage*). Wielkości te określają wpływ poszczególnych obserwacji na ocenę parametrów modelu.

Z faktu, że macierz rzutowania jest idempotentna ($\mathbf{H}^2 = \mathbf{H}$), wynika, że

$$0 \leq h_i \leq 1 \quad (1.1.2)$$

Wpływ i -tej obserwacji na zmianę teoretycznej wartości zmiennej objaśnianej zależy wyłącznie od wielkości reszty oraz i -tej wielkości wpływowej. Wartości h_i bliskie jedności istotnie wpływają na wartości teoretyczne i -tej zmiennej, nawet jeśli wartość i -tej reszty jest mała (Ostasiewicz, 1998). R.E Welsch (Hoaglin i in., 1978) zaproponował próg równy $h_H = 2(k+1)/n$, pozwalający stwierdzić, czy dana obserwacja jest obserwacją wpływową. Przyjmuje się, że jeśli wartość h_i odpowiadająca i -tej obserwacji spełnia warunek:

$$h_i > h_H \quad (1.1.3)$$

wówczas i -tą obserwację można uznać za wpływową.

WYKRYWANIE OBSERWACJI NIETYPOWYCH

Jednym z narzędzi służących identyfikacji obserwacji nietypowych są standaryzowane reszty $e_{(i)}^*$ postaci:

$$e_{(i)}^* = \frac{e_i}{s_{(i)} \sqrt{1 - h_i}}, \quad (1.2.1)$$

gdzie $s_{(i)}^2 = \frac{n-k-1}{n-k-2} s^2 - \frac{e_i^2}{(n-k-2)(1-h_i)}$ jest oceną wariancji σ^2 składnika

losowego po usunięciu i -tej obserwacji.

Reszty dane wzorem (1.2.1) nazywane resztami studentyzowanymi, mają rozkład t -Studenta o $(n-k-2)$ stopniach swobody [Staudte i in. 1990, str. 210-211]. Fakt ten umożliwia przeprowadzenie testu statystycznego weryfikującego hipotezę zerową postaci: H_0 : i -ta obserwacja jest elementem rdzenia, wobec hipotezy alternatywnej H_1 : i -ta obserwacja nie jest elementem rdzenia, czyli jest nietypowa.

Jeżeli zachodzi warunek $|e_{(i)}^*| > t_{n-k-2}(\alpha)$, hipotezę zerową H_0 odrzucamy na korzyść hipotezy alternatywnej H_1 na poziomie istotności α .

ŁĄCZNE WYKRYWANIE OBSERWACJI WPŁYWOWYCH ORAZ NIETYPOWYCH

Wskaźnikiem identyfikującym w sposób łączny nietypowość i wpływowość i -tej obserwacji jest standaryzowana różnica między wartościami teoretycznymi zmiennej objaśnianej wyznaczonymi dla modelu oszacowanego na podstawie wszystkich obserwacji oraz modelu oszacowanego na podstawie zbioru nie zawierającego i -tej obserwacji:

$$DFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{s_{(i)}\sqrt{h_i}} = e_{(i)}^* \sqrt{\frac{h_i}{1-h_i}} \quad (1.3.1)$$

Przyjmuje się, że i -ta obserwacja jest odstająca, jeśli zachodzi następujący warunek:

$$|DFITS_i| > 2\sqrt{(k+1)/n} \quad (1.3.2)$$

METODOLOGIA ESTYMACJI ODPORNEJ OPARTEJ NA WAŻONYCH ESTYMATORACH WELSCHA

Wśród szerokiej grupy metod estymacji odpornej znajdują się m.in. metody polegające na odrzuceniu obserwacji uważanych za nietypowe, a następnie estymowaniu parametrów modelu na podstawie pozostałych obserwacji. Otrzymane tym sposobem estymatory nazywane są estymatorami uciętymi. Usunięcie obserwacji odstającej w przypadku trendu oznacza zaburzenie kolejności czasowej. Dlatego, w celu szacowania parametrów modelu szeregu czasowego na podstawie wszystkich obserwacji, poszczególnym obserwacjom przyporządkowywane są wagi w_i o wartościach z przedziału $(0,1]$. Obserwacjom odstającym tego szeregu przypisywane są wówczas wagi o mniejszych wartościach, natomiast elementom rdzenia wagi równe 1. Zauważmy, że uogólniając tą metodę na przypadek, w którym obserwacjom odstającym przyporządkowywane są wagi równe 0, a elementom rdzenia wagi równe 1, otrzymujemy metodę opartą na odrzucaniu obserwacji odstających.

Konsekwencją podejścia opartego na przypisywaniu poszczególnym obserwacjom wag w_i , spełniających warunek $0 < w_i \leq 1$ jest zastosowanie uogólnionej metody najmniejszych kwadratów do estymacji parametrów strukturalnych szeregu czasowego.

W metodzie tej zakłada się, że składnik losowy ϵ ma rozkład normalny z wartością oczekiwaną $E(\epsilon) = 0$ oraz macierzą kowariancji Σ postaci $\Sigma = Cov(\epsilon) = \sigma^2 \mathbf{W}^{-1}$, gdzie \mathbf{W} jest znaną symetryczną dodatnio określoną macierzą o wymiarach $n \times n$.

Istnieje wówczas symetryczna i dodatnio określona macierz $\mathbf{W}^{1/2}$ taka, że $\mathbf{W} = \mathbf{W}^{1/2} \mathbf{W}^{1/2}$.

Tak określony problem można sprowadzić do klasycznego przypadku MNK, wprowadzając następujące oznaczenia:

$$\mathbf{y}_w = \mathbf{W}^{1/2} \mathbf{Y} \quad (2.1)$$

$$\mathbf{X}_w = \mathbf{W}^{1/2} \mathbf{X} \quad (2.2)$$

$$\boldsymbol{\varepsilon}_w = \mathbf{W}^{1/2} \boldsymbol{\varepsilon} \quad (2.3)$$

Ogólne równanie modelu sprowadza się wówczas do postaci:

$$\mathbf{y}_w = \mathbf{X}_w \boldsymbol{\beta} + \boldsymbol{\varepsilon}_w, \quad \boldsymbol{\varepsilon}_w \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (2.6)$$

Estymator \mathbf{b}_w wektora parametrów $\boldsymbol{\beta}$ będący rozwiązaniem następującego równania normalnego:

$$\mathbf{X}_w^T \mathbf{y}_w = (\mathbf{X}_w^T \mathbf{X}_w) \mathbf{b}_w \quad (2.7)$$

jest równy:

$$\mathbf{b}_w = (\mathbf{X}_w^T \mathbf{X}_w)^{-1} \mathbf{X}_w^T \mathbf{y}_w \quad (2.8)$$

Uwzględniając w powyższej równości zależności (2.1) oraz (2.2), łatwo zauważyć, że estymator \mathbf{b}_w wektora parametrów $\boldsymbol{\beta}$ postaci (2.8) wyraża się wzorem:

$$\mathbf{b}_w = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (2.9)$$

Jest on więc uogólnionym estymatorem MNK.

Wartość oczekiwana i macierz kowariancji estymatora \mathbf{b}_w są postaci:

$$E(\mathbf{b}_w) = \boldsymbol{\beta}, \quad \Sigma = \text{Cov}(\mathbf{b}_w) = \sigma^2 (\mathbf{X}_w^T \mathbf{X}_w)^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \quad (2.10)$$

Estymator wariancji σ^2 wynosi:

$$s_w^2 = \mathbf{e}_w^T \mathbf{e}_w / (n - k - 1) \quad (2.11)$$

gdzie wektor reszt jest postaci (2.12)

$$\mathbf{e}_w = \mathbf{y}_w - \mathbf{X}_w \mathbf{b}_w \quad (2.12)$$

Rozważać będziemy macierz \mathbf{W} diagonalną o elementach będących wagami odpowiednio przyporządkowanymi poszczególnym obserwacjom. Wówczas estymator \mathbf{b}_w nazywany jest ważonym estymatorem MNK.

Jeden ze sposobów wyznaczania wartości wag w_i i -tej obserwacji został opracowany przez R.E Welscha. Zaproponowane wagi w_i są następującej postaci:

$$w_i = w \left(\mathbf{x}_i, \frac{e_i}{s_{(i)}} \right) = \min \left\{ \frac{c}{|DFITS|}, 1 \right\} \quad (2.13)$$

Stała c określona jest zależnością:

$$c = p \sqrt{(p+1)/n}, \quad \text{gdzie } p=1,2, \quad (2.14)$$

W opisanym powyżej procesie estymacji, obserwacjom odstającym szeregu czasowego przypisywane są wagi o wartościach równych $\frac{c}{|DFITS|}$.

Pozostałym obserwacjom przypisywane są wagi równe 1.

Estymator ważony \mathbf{b}_w parametrów $\boldsymbol{\beta}$, uzyskany przy wagach postaci (2.13) nazywany jest ważonym estymatorem Welscha. Estymator ten jest odporny zarówno na obserwacje wpływowe, jak i na obserwacje nietypowe.

Metoda estymacji odpornej oparta na ważonych estymatorach Welscha jest metodą iteracyjną. W każdym kroku iteracyjnym wyznaczane są wartości wag, estymatory parametrów strukturalnych oraz wariancji σ^2 . Macierz rzutowania \mathbf{H} w kolejnych iteracjach wynosi:

$$\mathbf{H} = \mathbf{X}_w (\mathbf{X}_w^T \mathbf{X}_w)^{-1} \mathbf{X}_w^T \quad (2.15)$$

Uwzględniając w powyższej równości postać macierzy \mathbf{X}_w danej zależnością (2.2), otrzymujemy:

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \quad (2.16)$$

IMPLEMENTACJA METODY

Przedmiotem analizy jest szereg czasowy wartości indeksu giełdowego WIG20 o piętnastominutowej częstotliwości notowań w dniu 13.02.2006. Okres badawczy obejmuje obserwacje zidentyfikowane w kolejnych kwadransach między momentem określenia wartości kursu na otwarciu (godzina 9.30) a momentem określenia wartości kursu na zamknięciu (godzina 16.30).

Dla szeregu czasowego tych obserwacji stosowanie KMNK celem szacowania parametrów strukturalnych modelu liniowego nie znajduje uzasadnienia, m.in. ze względu na fakt, że nie są spełnione założenia tej metody. Wyznaczone w pracy wartości ocen estymatorów parametrów modelu liniowego za pomocą KMNK oraz wartości statystyk określających jakość uzyskanej funkcji trendu, służą przede wszystkim uzasadnieniu potrzeby zastosowania metod estymacji odpornych na odstępstwa od metody klasycznej. Wyniki estymacji uzyskane przy użyciu KMNK zawiera kolumna 1 tabeli 1, natomiast wykres liniowej funkcji trendu otrzymanej tą metodą oznaczono linią ciągłą na rysunku 1.

W przypadku analizowanego szeregu, na wyniki szacowania parametrów KMNK mają wpływ obserwacje odstające. Z porównania wartości wpływowych h_i i -tej obserwacji, $i=1, \dots, 27$, wyliczonymi na podstawie (1.1.1) z wartością progową równą $h_H = 0,14$, wynika, że obserwacjami mogącymi oddziaływać na położenie linii trendu wyznaczonego KMNK oraz wartości ocen szacowanych parametrów są obserwacje pierwsza i ostatnia. Ponieważ dla trzech ostatnich obserwacji szeregu czasowego zachodzi warunek $|e_{(i)}^*| > t_{24}(0,1) = 1,71$, można

wnioskować na poziomie istotności 0,1, że obserwacje te stanowią grupę obserwacji nietypowych badanego szeregu wobec trendu liniowego uzyskanego w wyniku zastosowania KMNK. Wartości reszt studentyzowanych oraz wskaźnika $DFITS$ ostatnich czterech obserwacji zawiera kolumna druga i trzecia tabeli 2. Ponieważ warunek $|DFITS_i| > 0,54$ zachodzi tylko dla ostatnich czterech obserwacji, można przyjąć, że rdzeń w przypadku trendu liniowego wyznaczonego KMNK, składa się z wszystkich pozostałych obserwacji.

Wykorzystując wagi Welscha, wyliczone w oparciu o wartości wskaźników $DFITS$, wyznaczono wartości estymatorów odpornych na zidentyfikowane obserwacje odstające. Wyniki zawarte, w kolumnie drugiej tabeli 1, uzyskane w pierwszej iteracji metodą odporną, wskazują na lepszą jakość otrzymanego modelu. Stopień dopasowania wartości teoretycznych do danych empirycznych tworzących rdzeń, nieznacznie się polepszył.

W przeciwieństwie do wyników, jakie uzyskano klasyczną metodą estymacji, wartości ocen parametrów uzyskane metodą odporną, są istotnie różne od zera na poziomie istotności 0,05. Wartościami odstającymi od trendu liniowego wyznaczonego po pierwszej iteracji przez oceny estymatorów odpornych są dane obserwowane w czasie $t = 24, 25, 26$.

Wyniki dotyczące estymacji odpornej zawarte w tabeli 1 wskazują, że wraz z każdą z pięciu dokonanych iteracji jakość uzyskiwanych modeli odpornych polepsza się. Zmiany położenia linii trendu w przestrzeni R^2 zilustrowane na rysunku 1, odzwierciedlają wzrost współczynników kierunkowych prostych wyznaczanych w kolejnych iteracjach. Tym samym wzrasta wartość wpływu h_1 pierwszej obserwacji szeregu, podczas gdy wartości wpływowe obserwacji odstających maleją z każdą iteracją. Przykładowo, wyniki uzyskane w pierwszych dwóch iteracjach dotyczące wartości współczynników $DFITS$ oraz reszt studentyzowanych ostatnich czterech obserwacji zamieszczono w tabeli 2.

Istotnym problemem identyfikacji obserwacji odstających oraz stosowania metod odpornych jest odpowiednia specyfikacja modeli trendu, uwzględniających charakter kształtowania się wartości analizowanych zmiennych. Najlepszą funkcją trendu opisującą zmiany wartości indeksu giełdowego WIG20 jest wielomian szóstego stopnia, gdyż współczynnik determinacji dla tego trendu przyjmuje największą wartość w klasie trendów będących funkcjami elementarnymi.

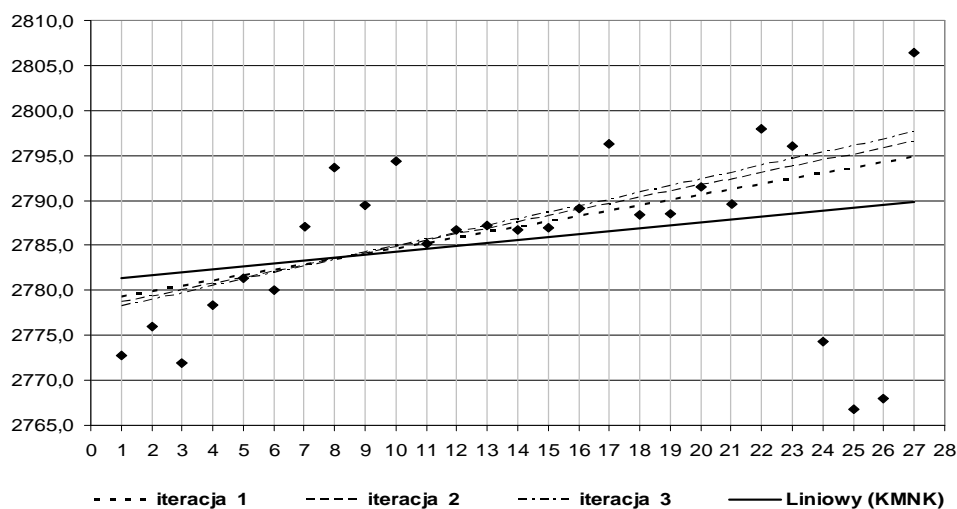
Oceny szacowanych za pomocą KMNK nieznanymi parametrów strukturalnych modelu oraz wartości statystyk określających jakość uzyskanej funkcji trendu zawiera kolumna 1 tabeli 3. Wykres wielomianowej funkcji trendu otrzymanej tą metodą oznaczono linią ciągłą na rysunku 2. Podobnie, jak w przypadku trendu liniowego wyznaczonego za pomocą KMNK, najwyższe wartości wpływowe h_i równe 0,84 posiadają obserwacje pierwsza oraz ostatnia. Ponieważ wartości reszt studentyzowanych odpowiadające obserwacjom w momencie $t = 22, 23, 26, 27$, spełniają warunek $|e^*_{(t)}| > t_{20}(0,1) = 1,72$, można wnioskować na poziomie istotności 0,1, że obserwacje te są nietypowe wobec

trendu wielomianowego, którego parametry strukturalne oszacowano KMNK. Wyznaczone wartości wskaźnika DFITS dla poszczególnych obserwacji wskazują, że obserwacja pierwsza oraz dwie ostatnie nie są elementami rdzenia. Wyniki jakie uzyskano na wskutek estymacji parametrów strukturalnych modelu wielomianowego metodą klasyczną oraz w kolejnych dwóch iteracjach odpornej metody Welscha, zawarte są w tabeli 3. Wykresy otrzymanych funkcji trendów wielomianowych przedstawia rysunek 2.

Tabela 1. Wyniki estymacji KMNK oraz OMW dla trendu liniowego

	KMNK		Odporna Metoda Welscha (OMW)									
	1		Iteracja 1		Iteracja 2		Iteracja 3		Iteracja 4		Iteracja 5	
	$b^{(0)}$	Stat. t	$b^{(1)}$	Stat. t	$b^{(2)}$	Stat. t	$b^{(3)}$	Stat. t	$b^{(4)}$	Stat. t	$b^{(5)}$	Stat. t
b_0	2781,03 (3,724)	746,86	2778,75 (2,740)	1013,99	2777,98 (2,477)	1121,71	2777,49 (2,273)	1221,92	2777,05 (2,072)	1308,31	2776,76 (2,004)	1426,04
b_1	0,3253 (0,232)	1,4	0,5956 (0,188)	3,17	0,6862 (0,172)	3,98	0,7427 (0,159)	4,66	0,8723 (0,125)	4,98	0,9307 (0,103)	5,03
R^2	0,0727		0,2694		0,3689		0,4389		0,5299		0,6745	
S_e^2	88,4853		64,5536		55,7583		49,5725		41,5339		28,7583	
S_{ew}^2			43,6984		35,1170		29,3015		22,574		17,834	

Źródło: obliczenia własne

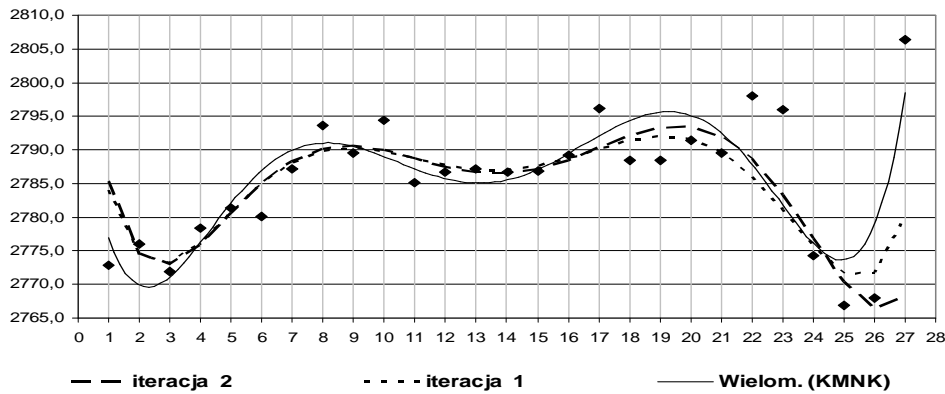


Rysunek 1. Wykresy liniowych funkcji trendu wartości indeksu WIG20 uzyskanych metodami: klasyczną i odporną

Tabela 2. Charakterystyki obserwacji odstających względem trendu liniowego

Nr obserwacji	KMNK		Odporna Metoda Welscha			
	DFITS	$ e^*_{(i)} $	Iteracja 1		Iteracja 2	
			DFITS	$ e^*_{(i)} $	DFITS	$ e^*_{(i)} $
24	0,56	1,69	0,98	3,57	0,98	4,65
25	1,01	2,85	1,56	7,27	3,38	19,16
26	1,03	2,74	1,54	6,84	2,93	15,48
27	0,82	2,02	0,52	1,91	0,49	1,75

Źródło: Obliczenia własne



Rysunek 2. Wykresy funkcji trendów wielomianowych szóstego stopnia wartości indeksu WIG20 uzyskanych metodami: klasyczną i odporną

Tabela 3. Wyniki estymacji KMNK oraz OMW dla trendu wielomianowego

	KMNK		Odporna Metoda Welscha (OMW)					
	1		Iteracja 1		Iteracja 2		3	
	$\mathbf{b}^{(0)}$	Stat. t	$\mathbf{b}^{(1)}$	Stat. t	$\mathbf{b}^{(2)}$	Stat. t	$\mathbf{b}^{(2)}$	Stat. t
h_0	2799,3949 (14,444)	193,81	2807,5223 (7,989)	351,42	2810,5821 (19,771)	142,15		
b_1	-32,7649 (12,912)	-2,54	-32,6950 (7,289)	-4,49	-35,1736 (15,054)	-2,34		
b_2	12,0174 (3,750)	3,20	10,5827 (2,156)	4,91	11,2191 (4,027)	2,79		
b_3	-1,7339 (0,485)	-3,57	-1,4159 (0,284)	-4,97	-1,4839 (0,503)	-2,95		
b_4	0,1186 (0,031)	3,83	0,0919 (0,018)	4,95	0,0949 (0,032)	2,99		
b_5	-0,0038 (0,001)	-4,02	-0,0029 (0,0005)	-4,90	-0,0029 (0,001)	-2,94		
b_6	0,00004 (0,00001)	4,18	0,00001 (0,00001)	4,83	3,40486E-05 (0,00001)	2,85		
R^2	0,6541		0,4938		0,7174			
S_e^2	41,2491		44,7157		24,9683			
S_{ew}^2			17,5595		12,0123			

Źródło: Obliczenia własne

PODSUMOWANIE

Występowanie obserwacji odstających ma szczególnie znaczenie w modelowaniu wartości zmiennych mających postać finansowych szeregów czasowych. Ponieważ usunięcie obserwacji odstających w przypadku trendu oznacza zaburzenie kolejności czasowej, kluczowym problemem staje się dobór odpowiednich metod estymacji parametrów modelu szeregu czasowego.

W niniejszej pracy omówiono i zastosowano jedną z metod estymacji, które uwzględniają obserwacje odstające i jednocześnie są na nie odporne. W tym celu do szacowania parametrów strukturalnych szeregu czasowego wartości indeksu giełdowego WIG20 wykorzystano ważone estymatory Welscha. Istotną zaletą tych estymatorów jest odporność na obserwacje odstające, zarówno ze względu na wpływowość jak i nietypowość. Własność ta ma szczególne znaczenie w procesie modelowania finansowych szeregów czasowych, w którym ważną rolę odgrywa odpowiednia specyfikacja możliwych do zastosowania funkcji trendu. Wybór postaci funkcji trendu determinuje bowiem identyfikację obserwacji odstających, które z kolei mają istotny wpływ na oceny szacowanych parametrów modelu.

LITERATURA

- Hoaglin D.C., Welsch R.E. (1978), "The Hat Matrix in Regression and ANOVA", *The American Statistician*, Vol.32, Issue 1, str. 17
- Ostasiewicz W. (1998), „Statystyczne metody analizy danych”, Wyd. AE Wrocław str: 235-274.
- Rousseeuw P. J., Leroy A. M. (2003), "Robust Regression and Outlier Detection", John Wiley&Sons, New Jersey.
- Staudte R.G, Sheather S.J. (1990), "Robust Estimation and Testing", John Wiley&Sons, New York.

Sample applications of the robust estimation to financial time series

Summary: The assumptions which constitute the basis for parametric estimation techniques, relating to the normality of distribution of the population as well as independent variables, are often not fulfilled in the case of financial data. Therefore it is important to use robust estimation methods, both in terms of these assumptions and in terms of the quality of observations. The paper aims to implement the robust estimation to modelling the time series of the WIG20 index of the 15-minute frequency on 13 February 2006. The structural parameters of the model were estimated with the use of the Welsch estimators.

Key words: outliers, influential observations, robust estimation, Welsch estimators