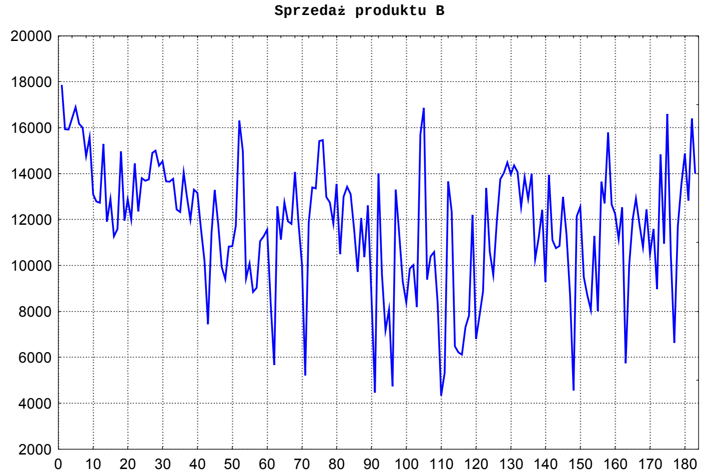
# PROGNOZOWANIE SPRZEDAŻY – STUDIUM PRZYPADKU

*prof. dr hab. Andrzej Sokołowski*[2]

W tym opracowaniu przedstawiony zostanie przebieg procesu poszukiwania modelu prognostycznego wykorzystującego jedynie przeszłe informacje na temat zjawiska prognozowanego. Celem postępowania jest zatem zbudowanie modelu szeregu czasowego dobrze oddającego jego przebieg w przeszłości, opisującego wewnętrzne prawidłowości analizowanego zjawiska oraz pozwalającego na sformułowanie sensownych prognoz. Informacje statystyczne dotyczą hurtowej sprzedaży pewnego wyrobu (oznaczamy go symbolem B). Dane są rzeczywiste i obejmują kolejne dni powszednie.

Każda analiza szeregu czasowego powinna rozpoczynać się od oglądu graficznej ilustracji przebiegu tego szeregu. Przy pomocy dostępnej z każdego modułu opcji *Wykresy* tworzymy liniowy wykres *Zmiennej B.*




Zazwyczaj w ekonomicznych szeregach czasowych doszukujemy się następujących składników: trend, wahania okresowe, jednorazowe interwencje oraz wahania losowe. W rozpatrywanym szeregu trudno jednoznacznie przewidzieć czy mamy tu do czynienia z trendem, a jeżeli tak – to którego stopnia. Zauważamy stosunkowo dużą wariancję procesu oraz liczne pojedyncze gwałtowne spadki i wzrosty sprzedaży. Przy poszukiwaniu modelu prognostycznego wykorzystamy 179 obserwacji szeregu, pozostawiając cztery ostatnie wartości do orientacyjnej oceny trafności prognoz. W tym celu wykorzystamy przycisk *Select cases (Wybierz obserwacje)*, definiując warunek jako *v0<180*.

[2] Akademia Ekonomiczna w Krakowie.

## Identyfikacja i estymacja trendu

W opcji *Wykresy* można dopasowywać funkcje trendu różnych postaci i mimo, że program podaje analityczną postać funkcji trendu, warto tę opcję wykorzystywać tylko do wstępnego zidentyfikowania typu tendencji rozwojowej. Jednoznaczna identyfikacja modelu trendu wymaga przeprowadzenia testowania istotności parametrów. Przy niezbyt skomplikowanych modelach trendu zaleca się skorzystanie z modułu *Regresja wielokrotna*. Wydaje się, że w naszym przypadku po ocenie przebiegu szeregu czasowego na wykresie, można rozsądnie założyć, że trend jest wielomianem stopnia nie wyższego niż trzeci. W celu oszacowania takiego trendu, do zbioru danych wprowadzamy zmienne *T2* oraz *T3* oznaczające odpowiednio drugą i trzecią potęgę zmiennej czasowej (numeru dnia w zbiorze danych). W module *Regresja wielokrotna* jako zmienną zależną wybieramy *B*, zaś jako zmienne niezależne *T, T2* oraz *T3*. Szacujemy model i przy pomocy przycisku *Podsumowanie regresji* uzyskujemy następujące okno wyników:

Podsumowanie regresji zmiennej zależnej: B

Dalej...
R= .46567545 R2= .21685363 Popraw. R^2= .20342826
F(3,175)=16,153 p<,00000 Błąd std. estymacji: 2482,2

| N=179 | BETA | Błąd st. BETA | B | Błąd st. B | t(175) | poziom p |
|---|---|---|---|---|---|---|
| W. wolny | | | 15874,52 | 757,9277 | 20,94463 | 0,000000 |
| T | -2,34009 | ,677530 | -125,60 | 36,3644 | -3,45385 | ,000693 |
| T2 | 3,19173 | 1,623006 | ,92 | ,4687 | 1,96655 | ,050816 |
| T3 | -1,15791 | 1,012373 | -,00 | ,0017 | -1,14376 | ,254206 |

Zauważamy, że nie wszystkie *wartości p* są mniejsze od poziomu istotności 0,05. Największa wartość prawdopodobieństwa testowego odpowiada zmiennej *T3*, co oznacza, że w analizowanym szeregu nie występuje trend trzeciego stopnia. Zgodnie z zasadą regresji krokowej eliminujemy zmienną, której odpowiada największa wartość *p* i ponownie szacujemy model. Otrzymujemy następujący wynik:

Podsumowanie regresji zmiennej zależnej: B

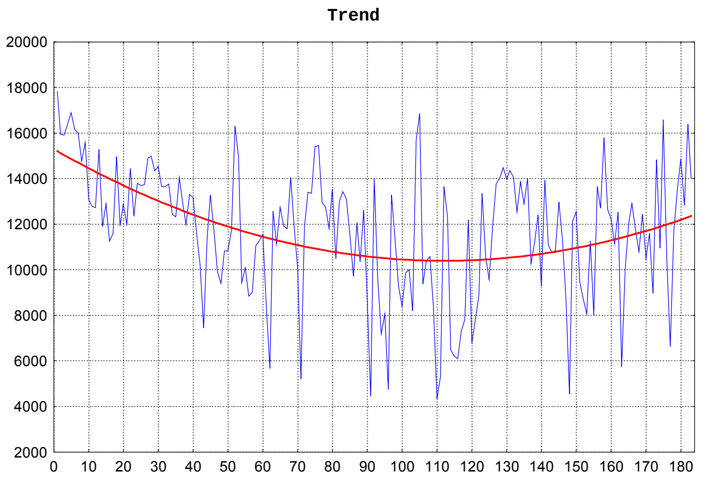
Dalej...
R= .45934666 R2= .21099936 Popraw. R^2= .20203344
F(2,176)=23,533 p<,00000 Błąd std. estymacji: 2484,3

| N=179 | BETA | Błąd st. BETA | B | Błąd st. B | t(176) | poziom p |
|---|---|---|---|---|---|---|
| W. wolny | | | 15293,96 | 563,3487 | 27,14829 | 0,000000 |
| T | -1,62886 | ,269238 | -87,42 | 14,4505 | -6,04987 | ,000000 |
| T2 | 1,36108 | ,269238 | ,39 | ,0778 | 5,05529 | ,000001 |

Tym razem wszystkie *wartości p* są bardzo małe i oszacowany model uznajemy za ostateczny. Stwierdzamy, że w badanym szeregu czasowym zaobserwowano istotny statystycznie trend wielomianowy stopnia drugiego opisany równaniem:

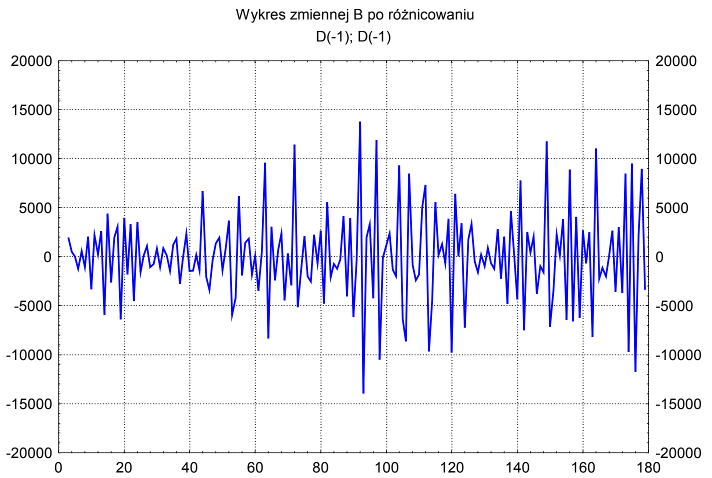
$$TREND = 15293{,}96 - 87{,}42 \cdot T + 0{,}39 \cdot T2$$

Zgodnie z tym równaniem zdefiniujemy w zbiorze danych zmienną *TREND*. Po jej wyliczeniu możemy wykreślić trend na rysunku. Ponownie uwidacznia się duża wariancja wokół linii trendu. Przypominamy, że trend był szacowany z opuszczeniem ostatnich czterech wartości. Są one jednak uwidocznione na rysunku i wszystkie znajdują się powyżej zwykłej ekstrapolacji trendu.

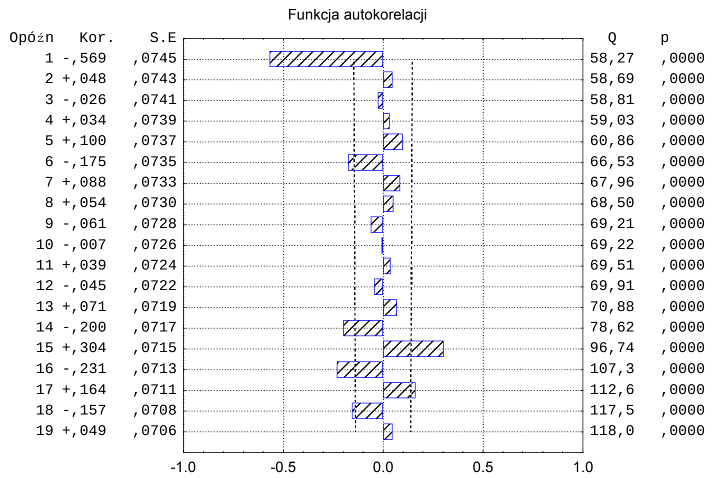



## Analiza struktury harmonicznej szeregu

Dla identyfikacji ewentualnych wahań okresowych obecnych w szeregu należy przede wszystkim wyeliminować z niego trend. Do eliminacji trendu wykorzystamy prostą metodę różnicowania. Ponieważ stwierdziliśmy występowanie trendu drugiego stopnia dlatego dla jego wyeliminowania należy dwukrotnie policzyć pierwsze różnice. Różnicowania dokonujemy w module *Szeregi czasowe/Prognozowanie*. Ponieważ w module tym program zapamiętuje automatycznie zadaną przez użytkownika liczbę ostatnio dokonanych przekształceń szeregu zatem nie ma potrzeby zapamiętywania wyników kolejnych różnicowań w zbiorze danych. Poniższy rysunek jest wykresem drugich różnic.




Teraz szereg jest pozbawiony trendu. Wydaje się też, że wariancja jest dość stabilna. Na podstawie szeregu pozbawionego trendu możemy wyznaczyć wartość funkcji autokorelacji.

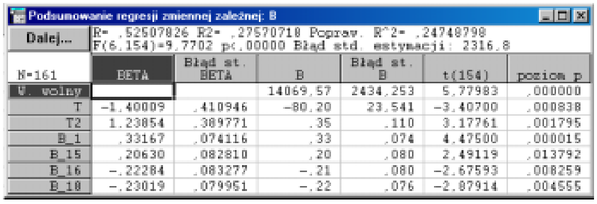



| Opóźn | Kor. | S.E | Q | p |
|---|---|---|---|---|
| 1 | -,569 | ,0745 | 58,27 | ,0000 |
| 2 | +,048 | ,0743 | 58,69 | ,0000 |
| 3 | -,026 | ,0741 | 58,81 | ,0000 |
| 4 | +,034 | ,0739 | 59,03 | ,0000 |
| 5 | +,100 | ,0737 | 60,86 | ,0000 |
| 6 | -,175 | ,0735 | 66,53 | ,0000 |
| 7 | +,088 | ,0733 | 67,96 | ,0000 |
| 8 | +,054 | ,0730 | 68,50 | ,0000 |
| 9 | -,061 | ,0728 | 69,21 | ,0000 |
| 10 | -,007 | ,0726 | 69,22 | ,0000 |
| 11 | +,039 | ,0724 | 69,51 | ,0000 |
| 12 | -,045 | ,0722 | 69,91 | ,0000 |
| 13 | +,071 | ,0719 | 70,88 | ,0000 |
| 14 | -,200 | ,0717 | 78,62 | ,0000 |
| 15 | +,304 | ,0715 | 96,74 | ,0000 |
| 16 | -,231 | ,0713 | 107,3 | ,0000 |
| 17 | +,164 | ,0711 | 112,6 | ,0000 |
| 18 | -,157 | ,0708 | 117,5 | ,0000 |
| 19 | +,049 | ,0706 | 118,0 | ,0000 |

Wartości funkcji autokorelacji wskazują, że można podejrzewać iż w szeregu występują składowe okresowe o okresach: 6, 14, 15, 16, 17, 18 dni. Wyraźnie zaznacza się składowa autoregresyjna rzędu pierwszego. Ta identyfikacja ma charakter wstępny.

## Budowa modelu prognostycznego

Przechodzimy do modułu *Regresja wielokrotna* i szacujemy model zawierający trend oraz wstępnie zidentyfikowane składowe okresowe w postaci odpowiednio opóźnionych zmiennych autoregresyjnych. Zmienne takie tworzymy w zbiorze danych kopiując oryginalną zmienną *B* i odpowiednio ją opóźniając oraz nadając właściwe nazwy. Po estymacji modelu wstępnego i wyeliminowaniu zmiennej *B* opóźnionej o sześć, czternaście i siedemnaście jednostek czasu otrzymujemy model ostateczny.



Podsumowanie regresji zmiennej zależnej: B

Dalej...

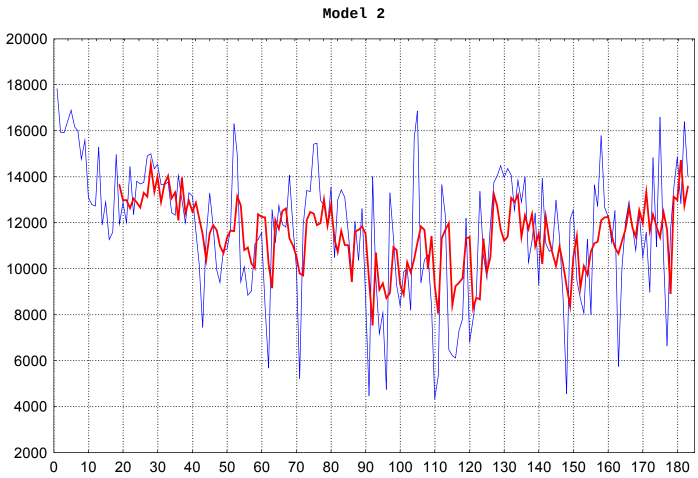
R= ,52507826 R2= ,27570718 Poprav. R^2= ,24748798
F(6,154)=9,7702 p<,00000 Błąd std. estymacji: 2316,8

| N=161 | BETA | Błąd st. BETA | B | Błąd st. B | t(154) | poziom p |
|---|---|---|---|---|---|---|
| W. wolny | | | 14069,57 | 2434,253 | 5,77983 | ,000000 |
| T | -1,40009 | ,410946 | -80,20 | 23,541 | -3,40700 | ,000838 |
| T2 | 1,23854 | ,389771 | ,35 | ,110 | 3,17761 | ,001795 |
| B_1 | ,33167 | ,074116 | ,33 | ,074 | 4,47500 | ,000015 |
| B_15 | ,20630 | ,082810 | ,20 | ,080 | 2,49119 | ,013792 |
| B_16 | -,22284 | ,083277 | -,21 | ,080 | -2,67593 | ,008259 |
| B_18 | -,23019 | ,079951 | -,22 | ,076 | -2,87914 | ,004555 |

Model (2) można zapisać jako:

$$\hat{B}_t = 14069{,}57 - 80{,}20 \cdot t + 0{,}35 \cdot t^2 + 0{,}33 \cdot B_{t-1} + 0{,}20 \cdot B_{t-15} - 0{,}21 \cdot B_{t-16} - 0{,}22 \cdot B_{t-18}$$

Parametry tego modelu są istotne statystycznie. Po wprowadzeniu nowej zmiennej o nazwie *PROGN_2* (prognoza otrzymana z modelu 2) sporządzamy rysunek szeregu empirycznego oraz wartości teoretycznych Modelu 2.




## Próba opisu reszt odstających

Model 2 na ogół dobrze podąża z szeregiem, ale nie jest w stanie zadowalająco opisać dużych odchyleń dodatnich i ujemnych. Ponieważ szereg czasowy dotyczy sprzedaży hurtowej, więc można przypuszczać, że wystąpienie dużej reszty może mieć jakiś wpływ na kształtowanie się szeregu w najbliższej przyszłości. Jeżeli odbiorcy kupili bardzo dużo danego towaru w pewnym dniu, to przy założeniu pewnej równomierności sprzedaży detalicznej można przypuszczać, że sprzedaż hurtowa w następnym dniu będzie mniejsza. Ten wpływ obserwacji odstających może być oczywiście dłuższy niż jeden dzień. W naszym modelu przetestujemy ten wpływ w okresie do trzech dni. W związku z tym przeglądamy reszty Modelu 2 (używając kolejno przycisków *Analiza reszt* oraz *Wartości przewidywane i reszty*) i znajdujemy obserwacje dla których reszty standaryzowane przekroczyły ±1,5. Liczba ta została wybrana drogą empiryczną. Do zbioru danych wprowadzamy zmienne zero-jedynkowe identyfikujące te „duże” reszty – osobno dodatnie, osobno ujemne. Zmienne te oznaczamy *RPLUS* oraz *RMIN*. Zmiennych tych nie ma sensu bezpośrednio wykorzystywać do modelowania, gdyż przy procesie prognozowania znamy tylko reszty przeszłe, a nie przyszłe. Dlatego w zbiorze danych tworzymy opóźnione zmienne resztowe (opóźnione o 1, 2 oraz 3 dni) oznaczając je przez *RPLUS_1, RPLUS_2, RPLUS_3* dla reszt dodatnich i analogicznie dla reszt ujemnych. Po zastosowaniu procedury regresji krokowej otrzymujemy następujący rezultat:

Podsumowanie regresji zmiennej zależnej: B

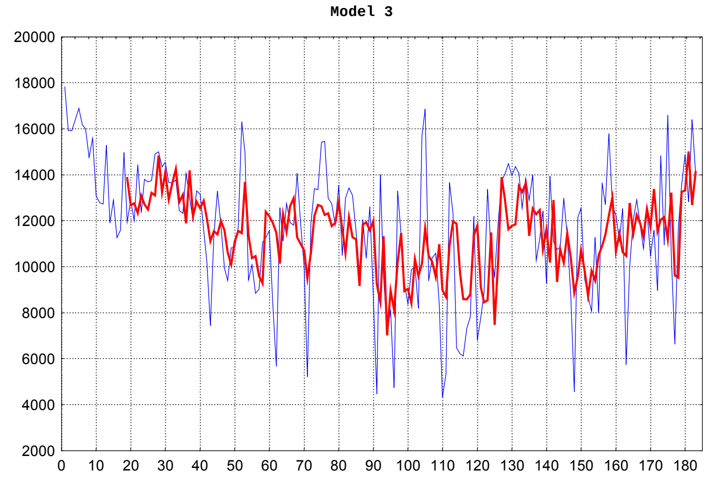
Dalej...

R= .57467540 R2= .33025182 Poprav. R^2= .29500191
F(8,152)=9.3689 p<.00000 Błąd std. estymacji: 2242.4

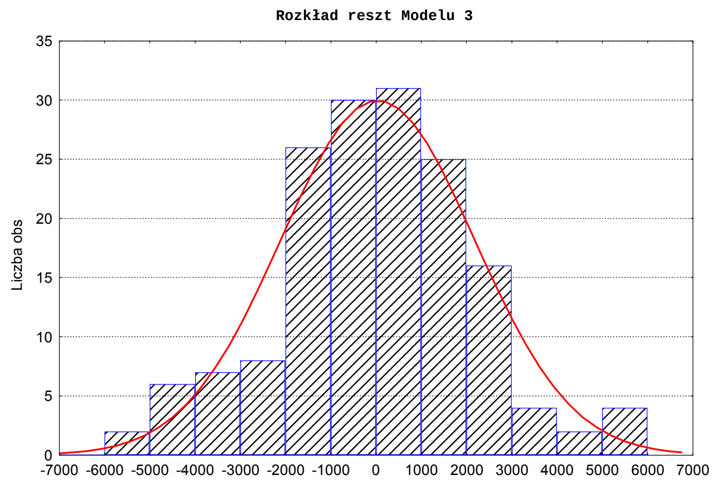
| N=161 | BETA | Błąd st. BETA | B | Błąd st. B | t(152) | poziom p |
|---|---|---|---|---|---|---|
| W. wolny | | | 10670,58 | 2604,947 | 4,09627 | ,000068 |
| T | -1,06707 | ,408987 | -61,13 | 23,429 | -2,60905 | ,009986 |
| T2 | ,95869 | ,385686 | ,27 | ,109 | 2,48568 | ,014013 |
| B_1 | ,49523 | ,092400 | ,49 | ,092 | 5,35960 | ,000000 |
| B_15 | ,28508 | ,083980 | ,28 | ,081 | 3,39459 | ,000877 |
| B_16 | -,27978 | ,082248 | -,27 | ,079 | -3,40163 | ,000856 |
| B_18 | -,19929 | ,078013 | -,19 | ,074 | -2,55453 | ,011616 |
| RMIN_1 | ,20074 | ,085859 | 2034,94 | 870,362 | 2,33804 | ,020689 |
| RPLUS_2 | -,18327 | ,068649 | -1933,98 | 724,412 | -2,66972 | ,008417 |

Widzimy, że duża reszta dodatnia ma jak się można było spodziewać ujemny wpływ na popyt w następnym dniu, natomiast duża reszta ujemna jest rekompensowana zwiększonymi zakupami za dwa dni.

Model, którego opis zawiera okno wyników powyżej nazwiemy Modelem 3. Po zdefiniowaniu zmiennej prognozowanej w zbiorze danych sporządzamy rysunek przedstawiający szereg oryginalny i wartości teoretyczne wynikające z Modelu 3.



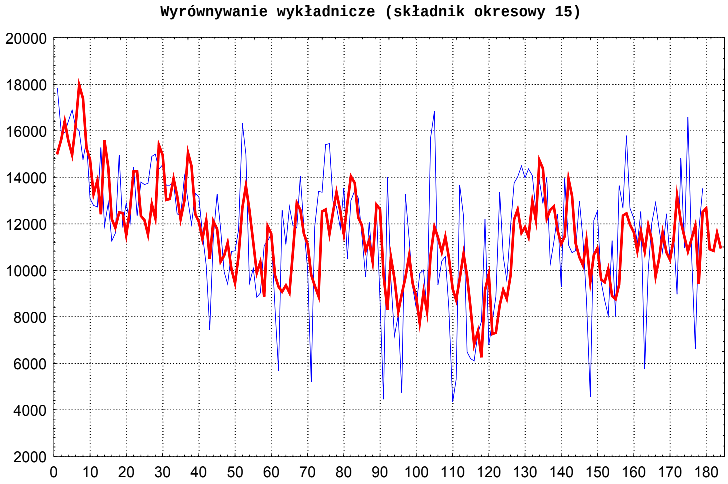

Z trzech porównywanych modeli, ten ostatni charakteryzuje się najmniejszą wartością średniego błędu dopasowania (około 2242). Statystyka Durbina-Watsona jest niemal dokładnie równa swej wartości przeciętnej w warunkach braku autokorelacji i wynosi 1,948. Współczynnik autokorelacji reszt rzędu pierwszego wynosi tylko 0,024. Kolejny rysunek przedstawia rozkład reszt.




Rozkład reszt nieźle pasuje do rozkładu normalnego.

## Wyrównywanie wykładnicze

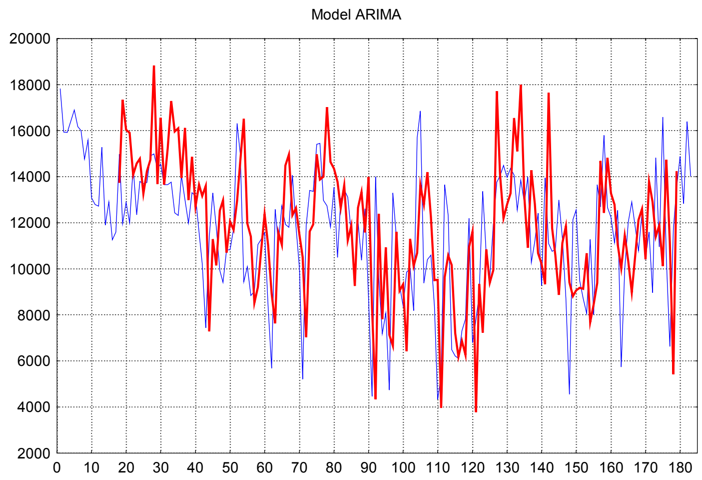
Obok przedstawionego powyżej modelu zaprezentujemy też dwie konkurencyjne procedury prognostyczne. Stosunkowo atrakcyjną procedurą, nawet dla mało wprawnych ekonometryków jest wyrównywanie wykładnicze. Wyjątkowo przyjazny moduł tej metody oraz pełna automatyzacja procesu poszukiwania najlepszych ocen parametrów modelu pozwala na błyskawiczne uzyskiwanie samego modelu oraz prognoz. Użytkownik programu musi na wstępie zdefiniować charakter trendu, wiodące wahanie okresowe oraz sposób powiązania składników szeregu czasowego (addytywny lub multiplikatywny). Z zaproponowanych postaci trendu wybieramy trend wykładniczy, jako najbardziej zbliżony do tendencji zaobserwowanej w naszym szeregu, przyjmujemy za wiodącą okresowość 15 dni (największy moduł współczynnika autokorelacji poza opóźnieniem 1) oraz addytywną konstrukcję modelu (wahania okresowe nie są proporcjonalne do trendu). Po tych decyzjach uzyskujemy model, którego dopasowanie przedstawiamy na rysunku.




Jak widać model „ignoruje" gwałtowne pojedyncze wzrosty lub spadki sprzedaży. Ponadto podobnie jak następny przedstawiany model umożliwia uwzględnienie tylko jednej składowej harmonicznej.

## Model ARIMA

Przy budowie modelu ARIMA wykorzystujemy te informacje, które uzyskaliśmy w toku poprzednich rozważań. Ponieważ zidentyfikowano trend drugiego stopnia zatem przed estymacją współczynników modelu przeprowadzamy dwukrotnie różnicowanie rzędu pierwszego, co doprowadza szereg do stacjonarności względem wartości przeciętnej. Przyjmujemy, że podstawową harmoniką występującą w zjawisku jest 15 dni. Podejmujemy próbę eliminacji tej składowej przez różnicowanie sezonowe z opóźnieniem 15. Poprzez kolejne powiększanie wartości parametrów *p, q, P, Q*, kontrolę istotności parametrów oraz funkcji autokorelacji reszt znajdujemy dopuszczalny model ARIMA(2,2,1)(0,1,1). Opis danych empirycznych przez ten model ilustruje rysunek.



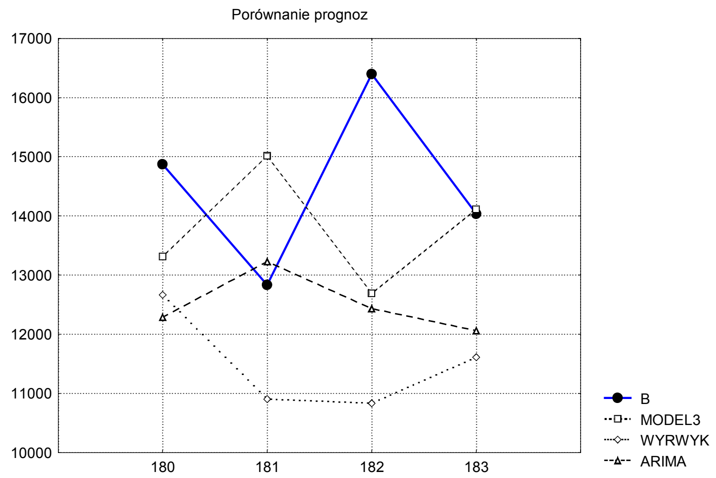

Wartości teoretyczne modelu ARIMA mają o wiele większą wariancję niż w przypadku pozostałych modeli. Obserwujemy pojedyncze „skoki" modelu w górę lub w dół. Po bliższej analizie można zauważyć, że ujemne odchylenia od ogólnej tendencji następują w modelu przeważnie w dzień po rzeczywistym skoku. To zjawisko jest niepokojące dla procesu prognozowania.

## Porównanie prognoz

Na wstępie analizy opuściliśmy w procesie estymacji cztery ostatnie obserwacje. Teraz wartości te wykorzystamy do porównania trafności prognoz. Porównanie to ma charakter wstępny. Model ostateczny został przekazany użytkownikowi i on zadecyduje, czy trafność prognoz jest wystarczająca z punktu widzenia skuteczności decyzji ekonomicznych podejmowanych na ich podstawie.

Porównanie trafności prognoz.

| Dzień | Wartość rzeczywista | Model 3 | Wyrównywanie wykładnicze | ARIMA |
|---|---|---|---|---|
| 180 | 14871,0 | 13312,6 | 12664,5 | 12288,5 |
| 181 | 12830,9 | 15015,2 | 10903,2 | 13228,0 |
| 182 | 16396,2 | 12690,5 | 10832,8 | 12432,8 |
| 183 | 14034,4 | 14113,4 | 11611,7 | 12061,3 |
| Średni błąd prognozy | - | 2288 / 15,7% | 3369 / 23,2% | 2570 / 17,7% |




Zarówno dane zawarte w Tabeli porównujące prognozy jak i analiza rysunku przemawiają za przyjęciem *Modelu 3* jako ostatecznego modelu prognostycznego.