

Analiza szeregów czasowych:

6. Liniowe modele niestacjonarne

P. F. Góra

<http://th-www.if.uj.edu.pl/zfs/gora/>

semestr letni 2007/08

Warunki stacjonarności modelu AR(p)

$$y_n = \beta_1 y_{n-1} + \beta_2 y_{n-2} + \dots + \beta_p y_{n-p} + \alpha_0 \eta_n. \quad (1)$$

Proces (1) jest stacjonarny wtedy i tylko wtedy, gdy pierwiastki równania

$$\lambda^p - \beta_1 \lambda^{p-1} - \beta_2 \lambda^{p-2} - \dots - \beta_p = 0 \quad (2)$$

leżą wewnątrz okręgu jednostkowego. Jeżeli któryś pierwiastek leży poza okręgiem jednostkowym, model wybucha.

Co się dzieje, jeśli jakiś pierwiastek leży *na* okręgu jednostkowym? Model jest niestacjonarny, ale czy możemy jakoś sobie z tym poradzić?

Trendy liniowe

Rozważmy proces

$$y_{n+1} = y_n + \alpha\eta_n \quad (3)$$

Proces ten jest niestacjonarny, równanie (2) ma dla niego postać $\lambda - 1 = 0$, ale **szereg pierwszych różnic** jest stacjonarny:

$$z_n^{(1)} \equiv y_{n+1} - y_n = \alpha\eta_n \quad (4)$$

jest stacjonarny!

Ogólnie, jeśli zbudowany ze współczynników procesu wielomian występujący w równaniu (2) ma postać $(\lambda - 1)B_p(\lambda)$, gdzie B_p jest wielomianem stopnia p , mającym wszystkie pierwiastki wewnątrz okręgu jednostkowego, proces taki nazywamy procesem ARIMA(1, p ,0). Proces taki ma (lokalny) trend liniowy.

Trendy kwadratowe

Podobnie, jeśli zbudowany ze współczynników procesu wielomian występujący w równaniu (2) ma postać $(\lambda - 1)^2 B_p(\lambda)$, gdzie B_p jest wielomianem stopnia p , mającym wszystkie pierwiastki wewnątrz okręgu jednostkowego, proces taki nazywamy procesem ARIMA(2, p ,0). Proces taki ma (lokalny) trend kwadratowy. Szereg jego **drugich różnic** jest stacjonarny.

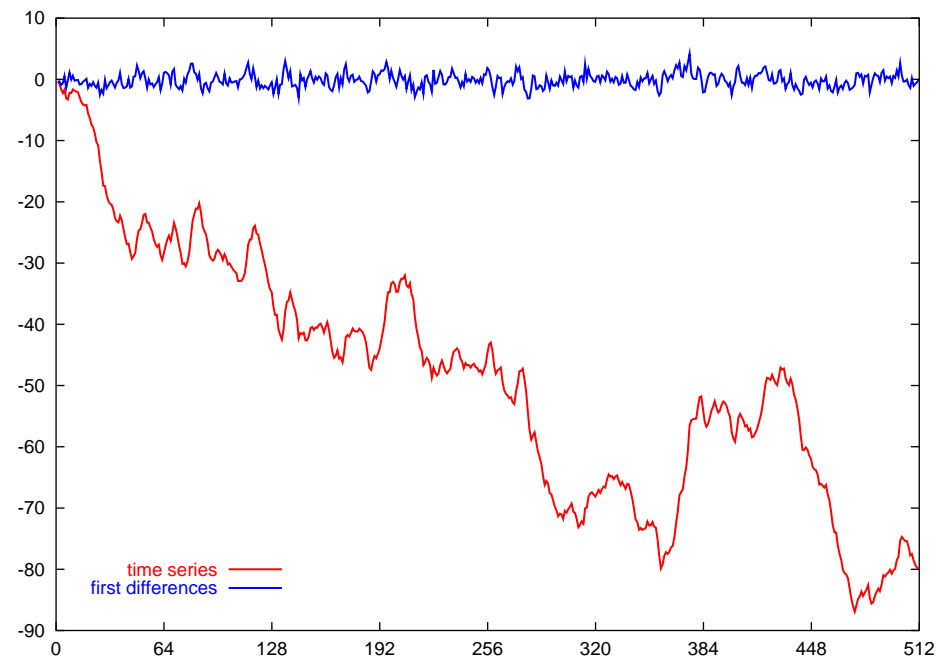
Przykład:

$$y_{n+1} = 2y_n - y_{n-1} + \alpha_0 \eta_n \quad (5a)$$

$$z_n^{(1)} \equiv y_{n+1} - y_n = z_{n-1}^{(1)} + \alpha_0 \eta_n \quad (5b)$$

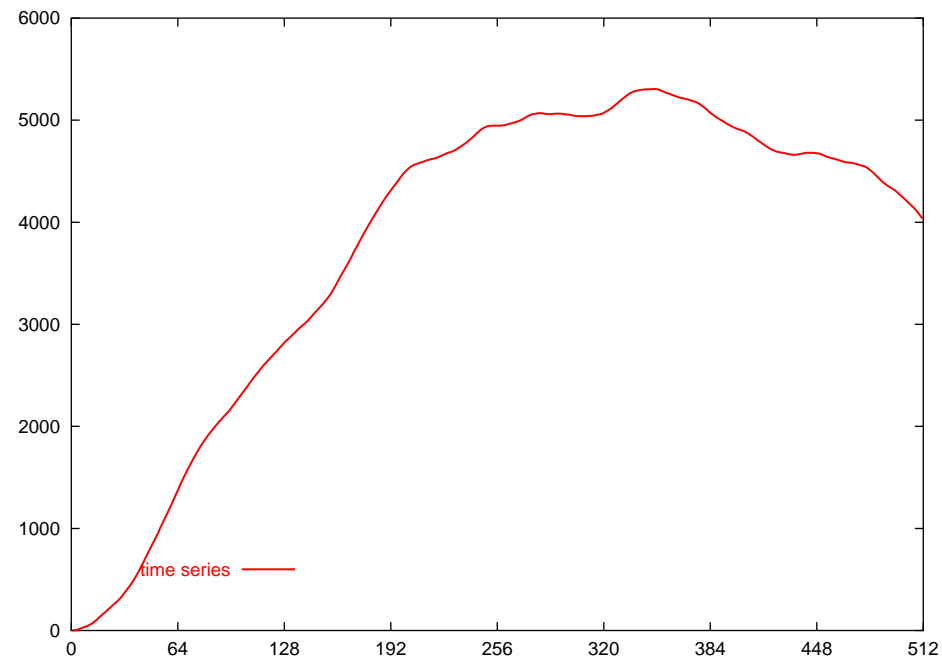
$$z_n^{(2)} \equiv z_n^{(1)} - z_{n-1}^{(1)} = \alpha_0 \eta_n \quad (5c)$$

Przykład szeregu z trendem liniowym

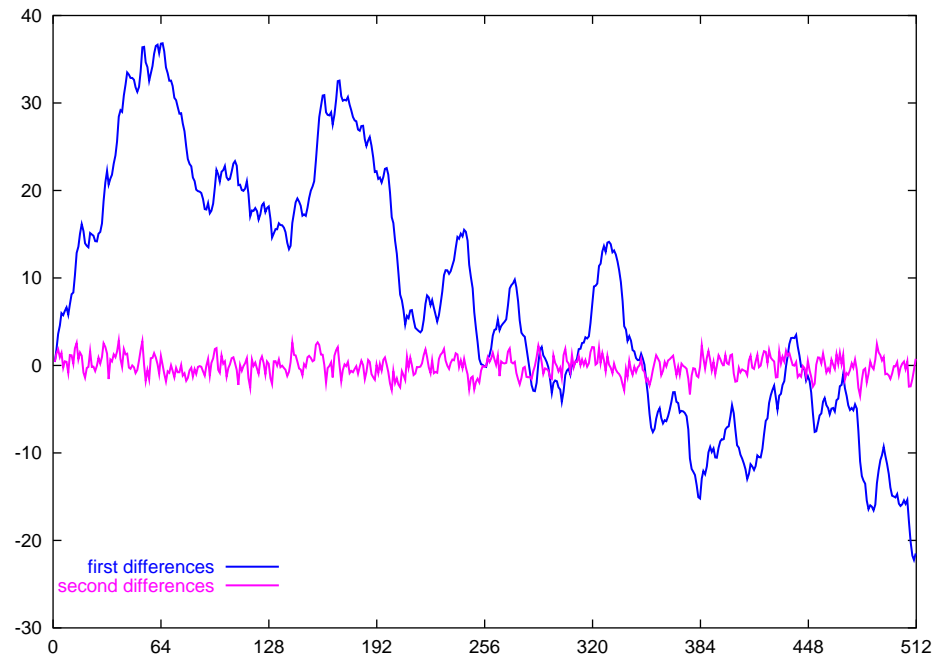


Niestacjonarny **szereg czasowy** i stacjonarny **szereg pierwszych różnic**

Przykład szeregu z trendem kwadratowym



Niestacjonarny **szereg czasowy**



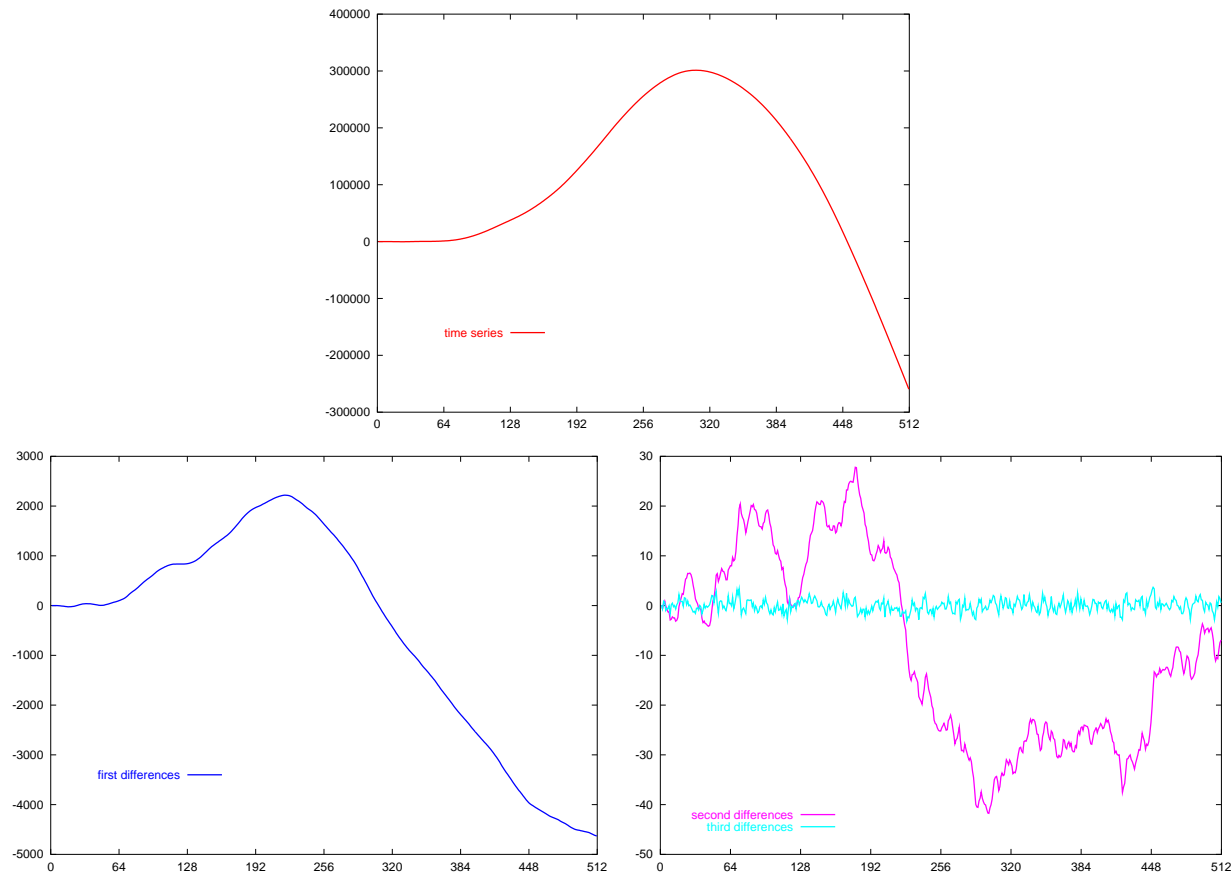
Niestacjonarny **szereg pierwszych różnic** szeregu z poprzedniego rysunku
i stacjonarny **szereg drugich różnic**

“Empiryczny” sposób postępowania

Jeśli domyślamy się, że badany szereg czasowy *może* mieć trend liniowy, badamy szereg jego pierwszych różnic. Jeśli szereg pierwszych różnic jest niestacjonarny, badamy szereg drugich różnic.

W praktycznych zastosowaniach szeregi z trendami liniowymi i kwadratowymi na ogół wystarczają, możemy się więc ograniczyć do badania pierwszych i drugich różnic.

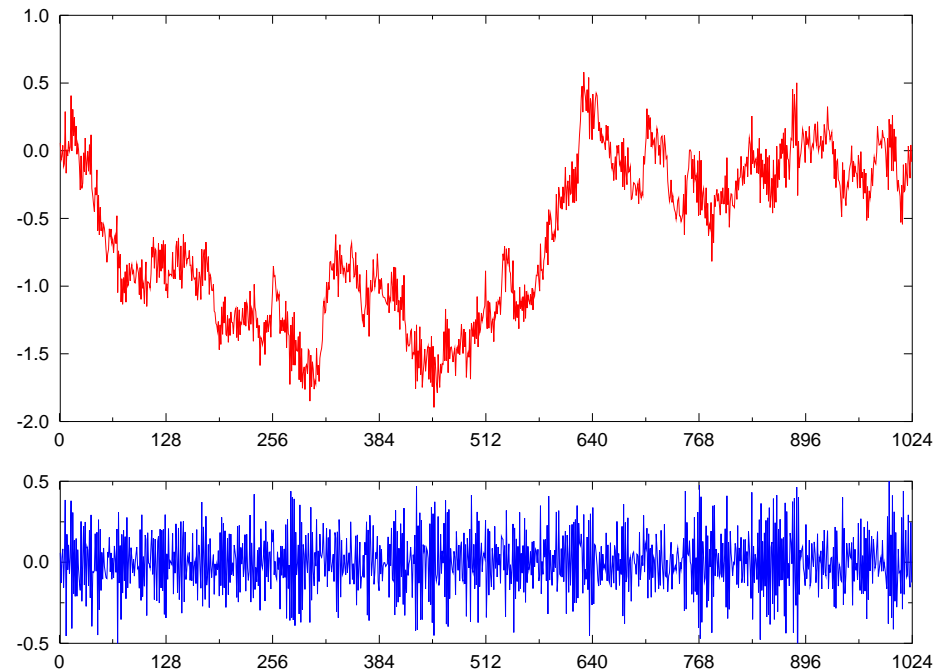
Przykład szeregu z trendem sześciennym (nierealistyczny)



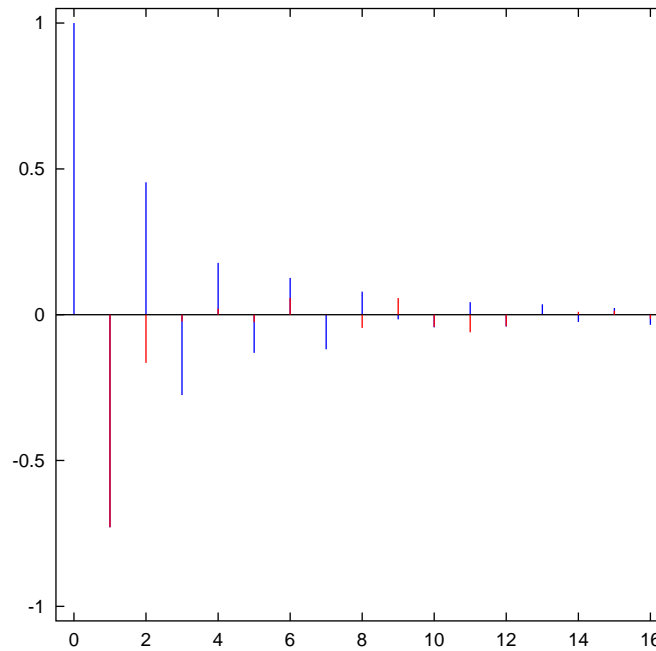
Dopasowywanie modelu

- Z niestacjonarnego szeregu czasowego tworzymy stacjonarny szereg pierwszych, ewentualnie drugich różnic.
- Do stacjonarnego szeregu różnic dopasowujemy model $ARMA(p,q)$.
- Wyjściowy szereg jest wówczas procesem typu $ARIMA(p,d,q)$, gdzie d oznacza rząd różnicy, jaki był potrzebny do uzyskania szeregu stacjonarnego.
- “I” w nazwie ARIMA oznacza “Integrated” — wycałkowany, czyli wysumowany.

Przykład



Niestacjonarny **szereg czasowy** i stacjonarny **szereg pierwszych różnic**



Funkcja korelacji ρ_k i funkcja korelacji cząstkowej φ_{kk} dla szeregu pierwszych różnic z poprzedniego rysunku. Uznajemy, iż $\varphi_{33} \simeq 0$.

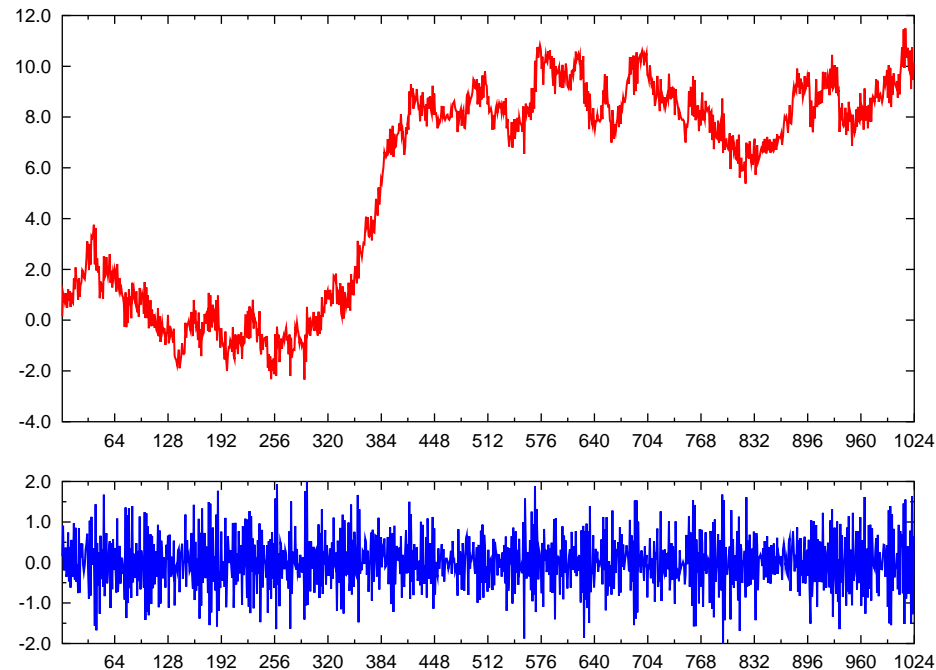
Ponieważ $\varphi_{33} \simeq 0$, przyjmujemy, że szereg pierwszych różnic jest procesem AR($p = 2$). Z równań Yule'a-Walkera obliczamy $\beta_1 = -0.849523$, $\beta_2 = -0.165224$. Wielomian określający stabilność szeregu pierwszych różnic dany jest przez $\lambda^2 - \beta_1\lambda - \beta_2$, a wobec tego wielomian określający stabilność *przecalkowanego* procesu dany jest przez $(\lambda - 1)(\lambda^2 - \beta_1\lambda - \beta_2)$, a zatem badany proces jest procesem ARIMA(2,1,0). Natężenie szumu szacujemy licząc pierwiastek z wariancji szeregu pierwszych różnic. Ostatecznie dla badanego procesu otrzymujemy

$$y_{n+1} = 0.150477 y_n + 0.684299 y_{n-1} + 0.165224 y_{n-2} + 0.189055 \eta_n \quad (6)$$

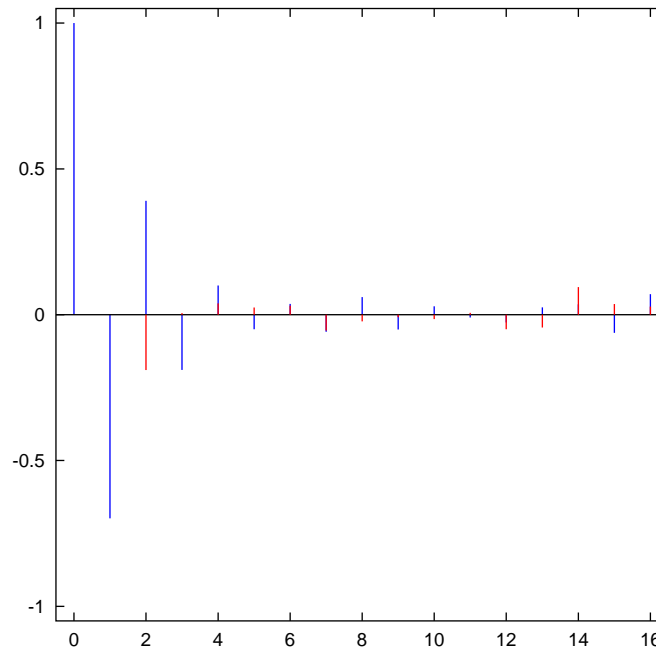
Proces, którego użyto do wygenerowania tego przykładu, miał postać

$$y_{n+1} = 0.166667 y_n + 0.666667 y_{n-1} + 0.166667 y_{n-2} + 0.125000 \eta_n$$

Gdy szum jest silniejszy, sytuacja może się pogorszyć



Niestacjonarny **szereg czasowy** i stacjonarny **szereg pierwszych różnic**



Funkcja korelacji ρ_k i funkcja korelacji cząstkowej φ_{kk} dla szeregu pierwszych różnic z poprzedniego rysunku.

W tym wypadku też moglibyśmy **na oko** uznać, że $\varphi_{33} \simeq 0$, a zatem, że szeregi pierwszych różnic można opisać procesem AR(2). Ale czy istnieje jakieś *kryterium* pozwalające rozstrzygnąć o rzędzie modelu?

Kryterium Akaike

AIC — Akaike Information Criterion

Jest jasne, że jeśli do danych dopasowujemy różne model, dopasowanie będzie tym lepsze, im więcej swobodnych parametrów będziemy mieli do dyspozycji. Z drugiej strony modele ze zbyt dużą liczbą swobodnych parametrów są uważane za niedobre. Powinniśmy więc jednocześnie nagradzać dobre dopasowanie i „karać” za zbyt wiele parametrów. Zakładamy, że błędy pochodzą z rozkładu normalnego i że dopasowanie przeprowadzamy metodą najmniejszych kwadratów.

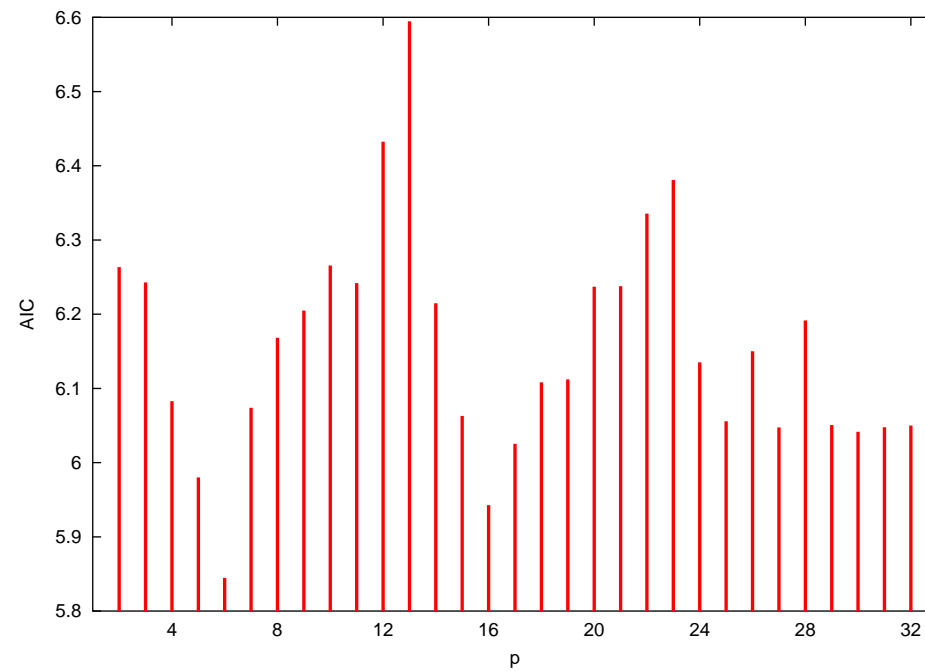
$$AIC = \ln Q + 2\frac{p}{N} \quad (7)$$

gdzie Q jest *rezydualnym błędem*: Przy dopasowywaniu modelu AR(p),

$$Q = \sum_n \left(z_n - \sum_{j=1}^p \beta_j z_{n-j} \right)^2 \quad (8)$$

gdzie $\{\beta_j\}_{j=1}^p$ są dopasowanymi parametrami dla procesu rzędu p . Im wyższe p , tym, potencjalnie, mniejsze Q , ale zarazem tym większy jest człon $2p/N$.
Dobieramy takie p , aby *AIC* było najmniejsze.

Uwaga: gdyby błędy nie pochodziły z rozkładu normalnego, trzeba by przyjąć inną definicję *AIC*.



Kryterium Akaike sugeruje użycie modelu AR(6).

Dla $n = 6$ otrzymalibyśmy $\beta_1 = -0.832051$, $\beta_2 = -0.181079$, $\beta_3 = 0.0407839$, $\beta_4 = 0.0650440$, $\beta_5 = 0.0504210$, $\beta_6 = 0.0310754$, a zatem wyjściowy szereg byłby procesem ARIMA(6,1,0) o parametrach

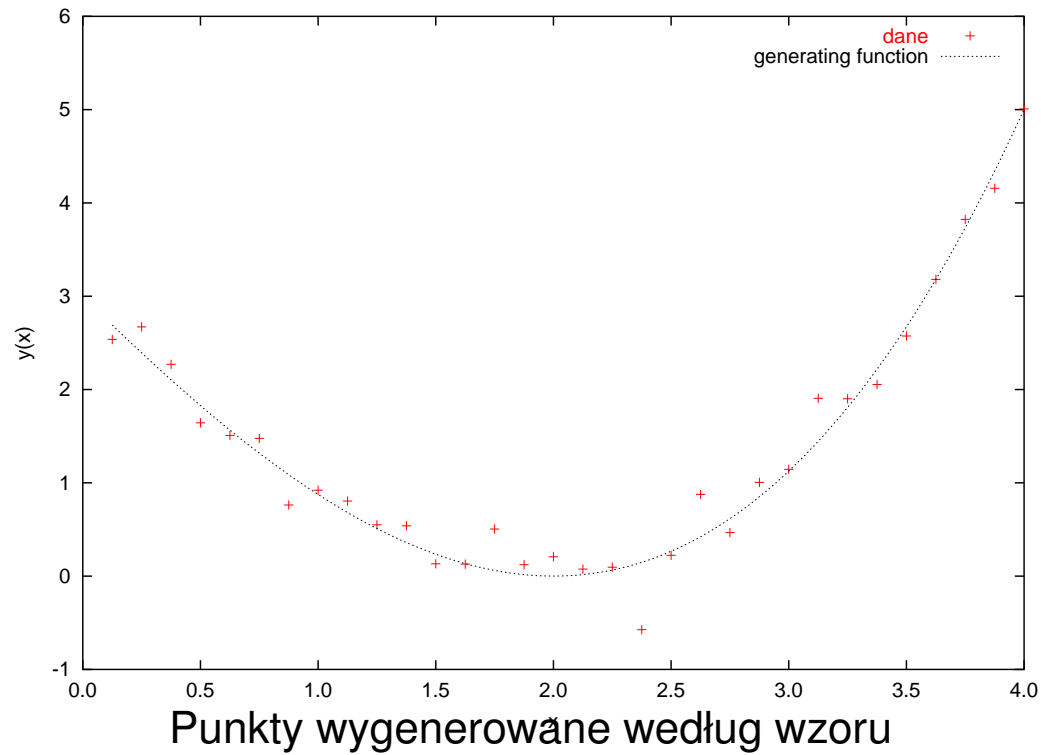
$$\begin{aligned} y_{n+1} = & 0.167949 y_n + 0.650972 y_{n-1} + 0.221863 y_{n-2} \\ & + 0.024260 y_{n-3} - 0.014620 y_{n-4} - 0.081496 y_{n-5} \\ & - 0.031075 y_{n-6} + 0.700224 \eta_n \end{aligned} \quad (9)$$

Dla porównania, przyjęcie $p = 2$ daje $\beta_1 = -0.831124$, $\beta_2 = -0.189943$ i dalej ARIMA(2,1,0):

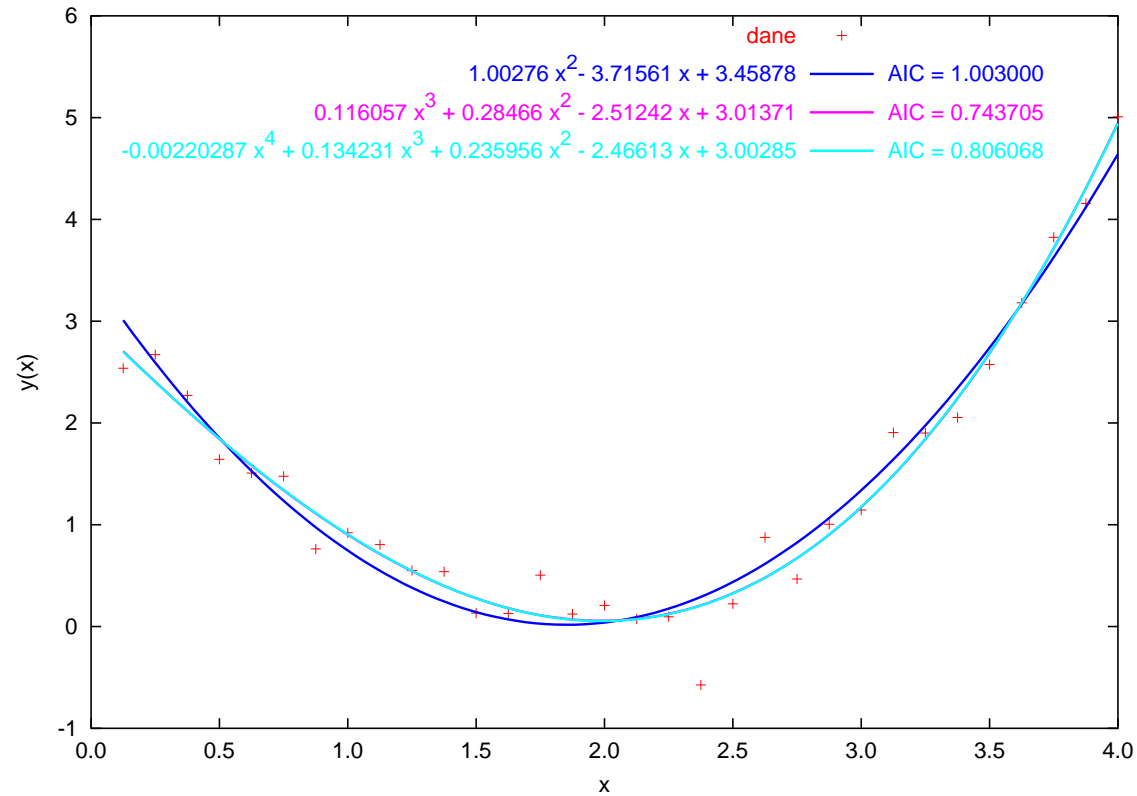
$$y_{n+1} = 0.168876 y_n + 0.641181 y_{n-1} + 0.189943 y_{n-2} + 0.700224 \eta_n \quad (10)$$

co jest *nieco* gorszym przybliżeniem idealnych parametrów użytych do wygenerowania szeregu.

Kryterium Akaike w “zwykłej” metodzie najmniejszych kwadratów



$$y_n = 0.125 x_n^3 + 0.25 x_n^2 - 2.5 x_n + 3.0 + 0.5 \eta_n \quad (11)$$

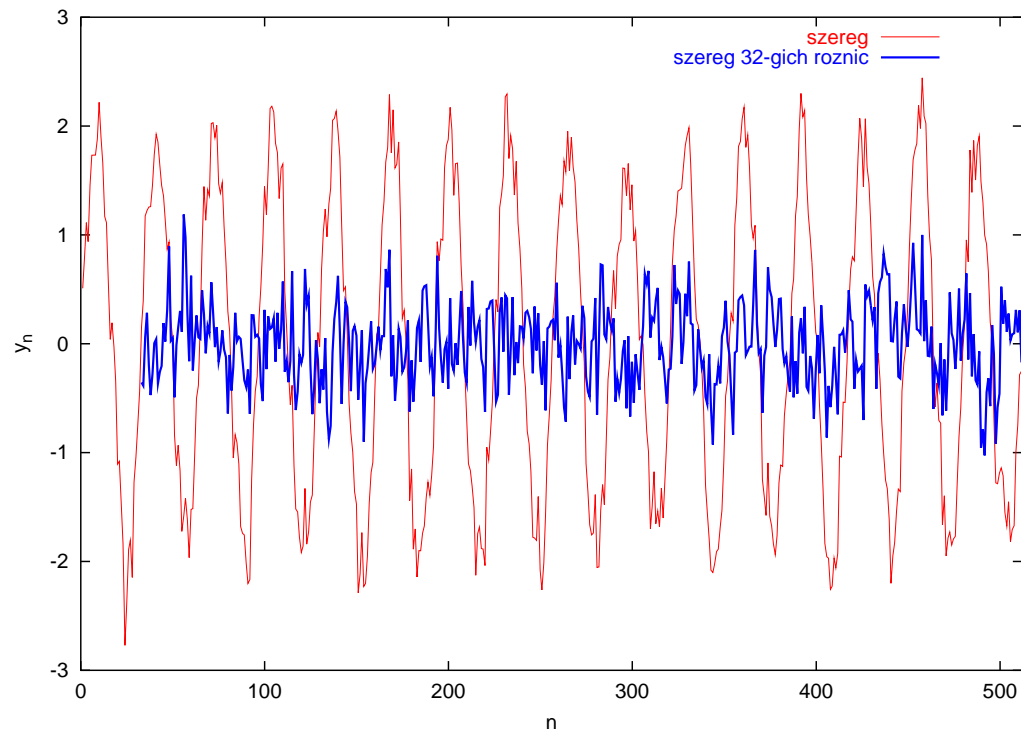


Wyniki dopasowania

Dopasowane krzywe trzeciego i czwartego stopnia — w skali rysunku nierozróżnialne

Szeregi sezonowe

Staramy się zidentyfikować okres, po czym tworzymy szereg odpowiednich różnic



Do stacjonarnego szeregu różnic dopasowujemy jakiś model ARMA