AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
Wydział Zarządzania

Projekt EFiD - Projekt 4B

Autorzy: *Patrycja Piła, Mateusz Strojek, Julia Szutka, Magdalena Wnuk*
Kierunek studiów: *Informatyka i Ekonometria*
Prowadzący: *dr hab. Tomasz Wójtowicz*

Kraków, 2025

Spis treści

Polecenie

Część B Dla podanego modelu ARMA wyprowadzić wzory na wariancje błędów poszczególnych prognoz. Uzupełnić arkusz tak, by obliczał wariancje poszczególnych prognoz i dla zadanego alfa (które może się zmieniać) wyznaczyć przedział ufności dla prognozowanych wartości. Na podstawie wyznaczonych rozkładów warunkowych obliczyć prawdopodobieństwa, że $X_{\mathrm{T+h}} > 0$ dla $h = 1, \dots, 3$. Porównać te wyniki z wynikami dla rozkładu bezwarunkowego.

Model: ARMA(2,0)

Wstęp

Celem niniejszego zadania jest na podstawie przygotowanego wcześniej arkusza kalkulacyjnego w Excelu, który zawiera estymację parametrów ARMA(2,0) metodą największej wiarygodności oraz prognozę na 3 okresów w przód, oszacować wariancje błędów poszczególnych prognoz, wyznaczyć przedział ufności oraz obliczyć podane wcześniej prawdopodobieństwo.

Metoda warunkowej największej wiarygodności polega na maksymalizacji funkcji wiarygodności, przy założeniu znanych początkowych wartości procesu.

Metoda Największej Wiarygodności (reszty z rozkładu normalnego)

Parametry:

- T- liczba obserwacji,
- φ_0 - wyraz wolny,
- φ_1 - parametr dla pierwszego opóźnienia,
- φ_2 - parametr dla drugiego opóźnienia,
- σ^2 – wariancja składnika losowego,
- p – rząd opóźnień.

Formuły w komórkach

Na podstawie szeregu czasowego X_t utworzono dwie kolumny z jego opóźnieniami:

- X_{t-1}
- X_{t-2}

Następnie obliczono wartość teoretyczną zmiennej, oznaczoną jako **x_hat**. Została ona wyznaczona ze wzoru:

- *X_hat* $=\varphi_0 +\varphi_1 X_{t-1} +\varphi_2 X_{t-2}$

Następnie obliczono reszty modelu:

- $\varepsilon t=Xt-X\_hat$

Dla każdej obserwacji obliczono składnik funkcji log-wiarygodności w postaci

- likelihood sig $=\varepsilon t$^2/(2σ^2)

Funkcja celu

Funkcja log-wiarygodności jest podana wzorem:

$$\log f_{Y_T, Y_{T-1}, \ldots, Y_{p+1}|Y_p, \ldots, Y_1}(y_T, y_{T-1}, \ldots, y_{p+1}|y_p, \ldots, y_1; \boldsymbol{\theta})$$
$$= -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) \qquad [5.3.9]$$
$$- \sum_{t=p+1}^{T} \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p})^2}{2\sigma^2}.$$

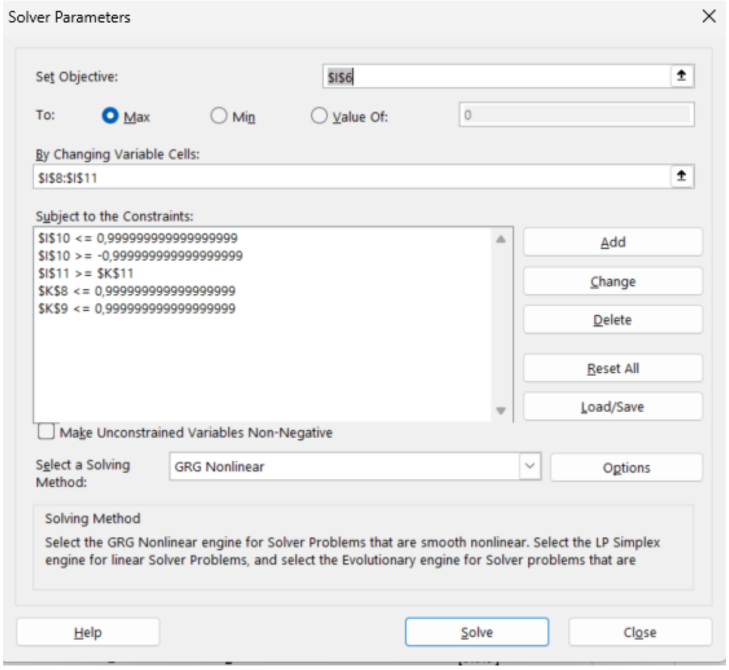
Rysunek 1 Wzór z książki „Time Series Analysis" James D. Hamilton 1994 r.

Po obliczeniu dwóch pomocniczych składników funkcji log-wiarygodności, zostały one **połączone w jedną formułę**, aby uzyskać końcową wartość logarytmu funkcji wiarygodności dla całego modelu.

-(T-p)/2 * log(2pi)	-96,48854599
-(T-p)/2 * log(sigm^2)	400,7562039
Funkcja celu	251,7676039

Za pomocą dodatku Solver przeprowadzono proces maksymalizacji funkcji celu, odpowiadającej logarytmowi funkcji wiarygodności modelu. Optymalizacja została wykonana względem czterech parametrów modelu: *φ0, φ1, φ2,* σ^2.
Dla parametru *σ^2* wprowadzono ograniczenie dolne równe 0,0001, co zapewnia spełnienie warunku dodatniości wariancji składnika losowego (wariancja musi być większa od zera). Dodatkowo użyto poniższych ograniczeń na parametry *φ1,φ2* które zapewnią stacjonarność modelu ARMA(2,0):

$$|\varphi_2| < 1,$$
$$\varphi_1 + \varphi_2 < 1,$$
$$\varphi_2 - \varphi_1 < 1.$$



Oszacowane parametry ARMA(2,0)

Wyniki uzyskane za pomocą warunkowej metody największej wiarygodności, są zgodne z metodą najmniejszych kwadratów. OLS został obliczony w arkuszu „Regresja".

Parametr	Wartość
ϕ_0	-0,003407035
ϕ_1	-0,140986288
ϕ_2	-0,144095741
σ^2	0,000483987

Tabela 1. Oszacowane parametry modelu AR(2)

Prognozowanie

Prognozowanie w modelach autoregresyjnych polega na wyznaczaniu przyszłych wartości zmiennej na podstawie jej własnych opóźnień oraz oszacowanych parametrów modelu. W przypadku modelu AR(2) przyjmuje się postać:

Dla $h = 1$

$$X_T(1) = \varphi_0 + \varphi_1 X_T + \varphi_2 X_{T-1}$$

Dla $h = 2$

$$X_T(2) = \varphi_0 + \varphi_1 X_T(1) + \varphi_2 X_T$$

Dla $h > 2$

$$X_T(h) = \varphi_0 + \varphi_1 X_T(h-1) + \varphi_2 X_T(h-2)$$

Gdzie:

- $X_T(h)$ to prognoza na h okresów,

- $X_T(h-1)$, $X_T(h-2)$ to prognoza na h-1 oraz h-2 okres,
- φ_0, φ_1, φ_2 to oszacowane parametry modelu,
- T to liczba obserwacji.

Horyzont prognozy	Wartość
1	0,0021
2	-0,0002
3	-0,0037

Tabela 2. Wyniki prognoz dla modelu AR(2)

Prognozy wskazują, że w najbliższym okresie oczekiwany jest niewielki wzrost (0,0021), po którym następuje stabilizacja, a następnie lekki spadek wartości zmiennej. Wraz z wydłużeniem horyzontu prognozy wartości maleją, co oznacza, że wpływ wcześniejszych obserwacji słabnie, a proces dąży do równowagi. Zmiany są niewielkie, więc model sugeruje stabilność badanego zjawiska w krótkim okresie.

Obliczenie błędów

W celu wyprowadzenia błędów prognoz, przy założeniu znajomości obserwacji do chwili T, konieczne jest wcześniejsze dokonanie prognoz.

Dla h = 1:

$$X_T(1) = \varphi_0 + \varphi_1 X_T + \varphi_2 X_{T-1}$$
$$X_{T+1} = \varphi_0 + \varphi_1 X_T + \varphi_2 X_{T-1} + e_{T+1}$$

Odejmując od siebie wartość zaobserwowaną a wartość zaprognozowaną, otrzymujemy błąd prognozy:

$$e_T(1) = X_{T+1} - X_T(1) = e_{T+1}$$

Dla h = 2:

$$X_T(2) = \varphi_0 + \varphi_1 X_T(1) + \varphi_2 X_T$$
$$X_{T+2} = \varphi_0 + \varphi_1 X_{T+1} + \varphi_2 X_T + e_{T+2}$$

Odejmując od siebie wartość zaobserwowaną a wartość zaprognozowaną, otrzymujemy błąd prognozy:

$$e_T(2) = X_{T+2} - X_T(2) = \varphi_1 X_{T+1} - \varphi_1 X_T(1) + e_{T+2} = \varphi_1\big(X_{T+1} - X_T(1)\big) + e_{T+2} = \varphi_1 e_T(1) + e_{T+2}$$

Dla h = 3

$$X_T(3) = \varphi_0 + \varphi_1 X_T(2) + \varphi_2 X_T(1)$$
$$X_{T+3} = \varphi_0 + \varphi_1 X_{T+2} + \varphi_2 X_{T+1} + e_{T+3}$$

Odejmując od siebie wartość zaobserwowaną a wartość zaprognozowaną, otrzymujemy błąd prognozy:

$$\begin{aligned} e_T(3) = X_{T+3} - X(3) &= \varphi_1 X_{T+2} + \varphi_2 X_{T+1} - \varphi_1 X_T(2) - \varphi_2 X_T(1) + e_{T+3} \\ &= \varphi_1\big(X_{T+2} - X_T(2)\big) + \varphi_2\big(X_{T+1} - X_T(1)\big) + e_{T+3} \\ &= \varphi_1 e_T(2) + \varphi_2 e_T(1) + e_{T+3} \end{aligned}$$

Ostatecznie, wzory na błąd prognozy dla ARMA(2, 0) kształtują się następująco:

Dla *h* <= 2:

$$e_r(h) = \varepsilon_{T+h} + \phi_1\big(X_{T+h} - X_r(h-1)\big) + \cdots + \phi_{h-1}\big(X_{T+1} - X_r(1)\big)$$

Dla *h* > 3

$$e_r(h) = \varepsilon_{T+h} + \phi_1\big(X_{T+h} - X_r(h-1)\big) + \cdots + \phi_p\left(X_{T+h-p} - X_T(h-p)\right)$$

Obliczenie wariancji błędów

Mając obliczone błędy prognoz dla *h* = 1,2,3, można z łatwością obliczyć wariancję tych błędów. Zakładając $e_t \sim N(0, \sigma^2)$ oraz że reszty stacjonarne:

Dla *h* = 1:

$$Var(e_T(1)) = Var(e_{T+1}) = \sigma^2$$

Dla *h* = 2:

$$Var\big(e_T(2)\big) = Var(\varphi_1 e_T(1) + e_{T+2}) = \varphi_1^2\sigma^2 + \sigma^2 = \sigma^2(\varphi_1^2 + 1)$$

Dla *h* = 3:

$$Cov\big(e_T(2), e_T(1)\big) = E[e_T(2) * e_T(1)] - E[e_T(2)] * E[e_T(1)] = \varphi_1 * \sigma^2$$

$$\begin{aligned} Var(e_T(3)) &= Var(\varphi_1 e_T(2) + \varphi_2 e_T(1) + e_{T+3}) \\ &= \varphi_1^2\sigma^2(\varphi_1^2 + 1) + \varphi_2^2\sigma^2 + \sigma^2 + 2\varphi_1\varphi_2 Cov(e_T(1), e_T(2)) \\ &= \sigma^2(\varphi_1^4 + 2\varphi_1^2\varphi_2 + \varphi_1^2 + \varphi_2^2 + 1) \end{aligned}$$

Wniosek: Im dalej chcemy zaprognozować, tym większa wariancja błędu prognozy.

Wyznaczenie przedziału ufności i prawdopodobieństwa

Po obliczeniu wariancji błędów, można przejść do wyznaczenia dolnego oraz górnego przedziału ufności. Dla podanego współczynnika ufności $1 - \alpha$, dolny i górny przedział ufności jest obliczany jest następująco:

$$\left(X_T(h) - u_{\alpha/2} \cdot D\big(e_T(h)\big)\,;\; X_T(h) + u_{\alpha/2} \cdot D\big(e_T(h)\big)\right)$$

Gdzie:

- *$X_T(h)$*- prognoza (wartość oczekiwana) zmiennej X dla okresu *T+h*
- *$e_T(h)$* - błąd prognozy h-okresowej,
- *$D(e_T(h))$* - odchylenie standardowe błędu prognozy dla *h*-kroku naprzód,
- $u_{\alpha/2}$ - kwantyl rzędu 1-α/2 standardowego rozkładu normalnego,
- α - poziom istotności.

Aby obliczyć prawdopodobieństwo, że zaprognozowana wartość będzie większa od zera, potrzebna jest wiedza o wariancji błędu (σ_{T+h}) oraz μ_{T+h}. Wariancję błędu obliczono wcześniej, natomiast średnią w tym przypadku jest prognoza. Korzystając z dystrybuanty standardowego rozkładu normalnego, jesteśmy w stanie obliczyć prawdopodobieństwo w następujący sposób:

$$P(X_{T+h} > 0) = 1 - \Phi\left(\frac{0 - \mu_{T+h}}{\sigma_{T+h}}\right)$$

Gdzie:

- μ_{T+h} – średnia (wartość oczekiwana) prognozowanego rozkładu, czyli po prostu prognoza punktowa *$X_T(h)$*.
- σ_{T+h} – odchylenie standardowe błędu prognozy na okres h
- *Φ(·)* – dystrybuanta standardowego rozkładu normalnego

Otrzymane wyniki

Ostatecznie, dla poziomu ufności równemu 0,95 otrzymano poniższe wyniki:

Horyzont	Prognoza	Wariancja błędów prognoz	Przedział ufności - dół	Przedział ufności - góra	$P(X_{T+h}>0)$	$P(X_t>0)$
1	0,0021	0,000484	-0,04106	0,04518	0,53726	0,45290
2	-0,0002	0,000494	-0,04375	0,04334	0,49632	0,45290
3	-0,0037	0,000501	-0,04755	0,04020	0,43480	0,45290

Tabela 3. Wariancja błędów prognoz, przedziały ufności i prawdopodobieństwa

Wariancja błędów prognoz rośnie wraz ze wzrostem horyzontu prognozy h, co jest zgodne z teorią — im dalszy horyzont, tym większa niepewność prognozy. W granicy dla dużych $h$wariancja błędów prognozy dąży do wariancji zmiennej X_t. Wraz ze wzrostem wariancji prognoz rozszerzają się również przedziały ufności. Dla poziomu ufności 0,95 szerokość przedziału ufności wynosi około 0,087, co odzwierciedla rosnącą niepewność przewidywań.

Zauważalna jest także różnica między prawdopodobieństwem $P(X_{T+h} > 0)$a $P(X_t > 0)$. Dla najbliższego horyzontu ($h = 1$) prognoza przyjmuje wartość dodatnią, co skutkuje wyższym prawdopodobieństwem przekroczenia zera. W miarę wzrostu horyzontu prognozy wartość $P(X_{T+h} > 0)$stopniowo zbliża się do wartości $P(X_t > 0)$, co oznacza, że w dłuższym okresie wpływ bieżących odchyleń maleje, a prognoza zbiega do długookresowej średniej procesu.